Protocolo de preservación digital para el sistema de preservación digital del Archivo Español de Media Art (AEMA)

Proyecto VOREMETUR. Universidad Carlos III de Madrid y Universidad de Castilla La Mancha.

Abril de 2018.

Versión 4.0

Autor: Jesús Robledano Arillo

Tabla de contenidos:

1 Sobre la compatibilidad de estas especificaciones.

Con la intención de mejorar la interoperabilidad del sistema de preservación del Archivo Español de Media Art, se ha buscado intencionadamente la proximidad del diseño del sistema con el del sistema de preservación a largo plazo de las bibliotecas digitales de la Subdirección General de Coordinación Bibliotecaria de la Secretaría de Estado de Cultura (SIPREDI_SGCB)[1], en cuyo diseño también hemos participado. Debido a que la tipología de documentos de los fondos típicos de Media Art difieren de los que se preservan en el sistema mencionado, las especificaciones que aquí presentamos introducen una amplia variedad de nuevos elementos y adaptaciones, por lo que el seguimiento no será nunca total.

Asimismo, se han intentado seguir una amplia variedad de estándares de amplio seguimiento en nuestros días por la comunidad de expertos en preservación digital, de entre los que destacamos la norma ISO 14721:2012[2] y Bag-it[3].

2 Conceptos fundamentales relacionados con el sistema de preservación digital.

2.1 El sistema de preservación digital.

El objetivo fundamental de la preservación digital es mantener en el tiempo —a corto, medio y largo plazo— la capacidad de utilizar los fondos digitales, ya sean producto de la digitalización de fondos no digitales preexistentes o de la acumulación de documentos u obras nacidos en formato digital. Podemos entender la preservación digital como un conjunto de estrategias ideadas para poder ir haciendo frente en el transcurso del tiempo a dos circunstancias: el cambio continuo de la tecnología sobre la que se construye y sustenta el fondo digital y los servicios que se ofrecen a partir de él, y la actuación de agentes de deterioro sobre los soportes informáticos donde se almacenan los datos y objetos digitales.

Mediante la aplicación de las estrategias de preservación digital se persigue mantener la integridad física (entendida como no alteración o destrucción accidental, por deterioro o malintencionada) y la capacidad de acceso y procesado de toda la información en formato digital, esto es, de los contenidos del fondo digital (los ficheros y sus metadatos). Para conseguir este objetivo se precisa

---

[1] Subdirección General de Coordinación Bibliotecaria. Ministerio de Educación, Cultura y Deporte. *Descripción del Sistema de Preservación a Largo Plazo de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria (SIPREDI_SGCB)*. 6 de octubre de 2017. Disponible en: http://travesia.mcu.es/portalnb/jspui/handle/10421/9003

[2] ISO. Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model. ISO 14721:2012. Geneva: ISO, 2012. Equivalente a: CCSDS. Reference Model for an Open Archival Information System (OAIS). Recommended practice CCSDS 650.0-M-2. MAGENTA BOOK. June 2012 [en línea]. Washington, DC: CCSDS, 2012. Disponible en: http://public.ccsds.org/publications/archive/650x0m2.pdf

[3] KUNZE, J., *et al. The BagIt File Packaging Format (V0.97). Draft-kunze-bagit-14*. October 21, 2016. Disponible en: https://tools.ietf.org/html/draft-kunze-bagit-14

conseguir de forma continuada el mantenimiento de la integridad física y no obsolescencia de soportes de almacenamiento y ficheros digitales.

Usamos el término obsolescencia para referir la caída en desuso de una tecnología concreta que provoca un riesgo de falta de soporte por parte de la industria que produce los elementos tecnológicos necesarios para su utilización. En los contenidos en formato digital la obsolescencia puede suponer la imposibilidad de utilizar los soportes donde se han almacenado los ficheros digitales, de abrirlos o reproducirlos desde aplicaciones informáticas (aunque estén contenidos en soportes no obsoletos), o de poder reproducirlos o visualizarlos correctamente sin la pérdida de contenido o atributos formales o elementos funcionales. Cuando los objetos digitales tienen evidencias de validación o de protección, como podría ser una firma digital o una marca de agua, la obsolescencia puede provocar la imposibilidad de validarlos o de activar las medidas de protección incrustadas, aunque estos puedan ser reproducidos correctamente.

Mantener la integridad física y no obsolescencia de soportes y ficheros digitales garantiza la capacidad de acceso a los contenidos originales de los documentos u obras digitales sin alteración alguna y de procesado de los ficheros digitales donde se representan. La información es en todo momento accesible y descodificable con una calidad de fidelidad al original, integridad, validez y fiabilidad suficientes para toda la comunidad de usuarios; y, además, se puede procesar de acuerdo a las necesidades requeridas en el uso y gestión del fondo digital.

Este aspecto incorpora dos perspectivas que hemos de diferenciar muy bien en la preservación digital: una perspectiva física y una perspectiva lógica. No tiene sentido preservar físicamente un fondo en formato digital si no se preserva de forma lógica. Veámoslo con un ejemplo: podríamos tener perfectamente accesible legible y sin errores una secuencia de bytes que representa físicamente toda la información de un fichero digital (preservación desde una perspectiva física), pero si luego no hay forma de descodificarla (preservación desde una perspectiva lógica, o lo que es lo mismo, saber qué representa exactamente cada byte de esa secuencia: una letra determinada, un número o fragmento de número, un píxel…) y poder conseguir el mensaje original que fue codificado mediante ella, no nos servirá para mucho.

El mantenimiento funcional de las aplicaciones software que pueden abrir y trabajar con los ficheros digitales es frecuentemente un prerrequisito para su preservación. Cuando se van sucediendo diferentes generaciones o versiones de una aplicación necesaria para leer determinados tipos de ficheros, normalmente los fabricantes o desarrolladores de software libre incorporan la compatibilidad hacia atrás, es decir, el que los formatos anteriores sean legibles en la nueva aplicación. Pero esto no siempre es así; cuando ha transcurrido mucho tiempo desde que se ha dejado de usar un formato determinado se termina dejándole de dar soporte. Por ello, a partir de determinado momento el formato se queda obsoleto y dependiente únicamente del mantenimiento de la capacidad de poder ejecutar correctamente en un ordenador una versión de la aplicación software que lee ese formato. La alternativa a la preservación digital de ese software, que se irá quedando obsoleto irremediablemente con las siguientes generaciones de sistemas operativos, es convertir los formatos ya obsoletos a otros formatos no obsoletos legibles por las nuevas versiones o aplicaciones que van saliendo en el

mercado; es lo que se denomina comúnmente como migración. En ocasiones este procedimiento es sencillo, pero en otras, cuando los formatos son excesivamente complejos es posible que no haya un formato no obsoleto candidato para migrar que permita mantener toda la funcionalidad y características del formato anterior, haciéndose preciso preservar la aplicación software que lo lee e interpreta con toda su funcionalidad; o programar un emulador, un software nuevo que imite sus funciones y aporte la capacidad de proceso de esas aplicaciones obsoletas.

Es si cabe más complejo preservar software de aplicación informática que un fichero digital de datos. El software tiene muchas características que dificultan su preservación digital; es complejo, a veces opaco para los programadores que no lo desarrollaron en su día; y, en ocasiones, depende de un amplio número de componentes software muy relacionados con el sistema operativo para el que se desarrolló, como, por ejemplo, las DLLs del sistema operativo Windows[4].

Migración y emulación son dos de las estrategias de preservación digital más empleadas hoy día para la lucha contra la obsolescencia tanto de formatos y sistema de codificación de ficheros como de sistemas operativos y aplicaciones software, pero hay otras que deben ser empleadas junto a aquellas de forma conjunta o complementaria.

Repasamos seguidamente las estrategias de preservación digital que de partida hemos contemplado en el proyecto junto con unos principios básicos que aseguren la inocuidad en su utilización:

a) Refresco.

Consiste en copiar el contenido de un soporte de almacenamiento a otro, manteniendo intacto el formato de los ficheros. El nuevo soporte que recibe los ficheros debe estar en perfectas condiciones y sin riesgo de obsolescencia a corto o medio plazo. Los ciclos de refresco deben ser programados por períodos inferiores a las expectativas de vida o de obsolescencia de los soportes.

El refresco de soporte se realiza para evitar la obsolescencia de los soportes, el envejecimiento y deterioro natural o por el uso del soporte, o ante la aparición en el mercado de soportes más eficientes, seguros, baratos y con un alto nivel de estandarización.

b) Migración.

[4] Dynamic Link Librarys. Son ficheros que contienen funciones o recursos que son llamados desde las aplicaciones que corren bajo Windows. Se identifican por la extensión ".dll". Entre otras ventajas, las DLL reducen el tamaño de los ficheros ejecutables, ya que gran parte del código del programa puede estar almacenado en DLLs en lugar de en el propio ejecutable. Pero las DLLs introducen problemas para el mantenimiento del funcionamiento de los programas con el paso del tiempo, pues los programas se hacen dependientes de otros ficheros externos que pueden ser dejados en desuso ante la continua irrupción de nuevas versiones.

Es el proceso de transferir información digital desde una plataforma hardware y software determinada (Sistema operativo, tipo de ordenador, aplicación informática) a otra diferente, o desde una generación de ordenadores a la siguiente. También el proceso de cambio de formato de los ficheros o de sus sistemas de codificación.

Los procesos de migración deben ser aplicados con mucha cautela, pues durante ellos se pueden producir cambios y pérdidas de datos o metadatos en los objetos digitales, o incluso perder la funcionalidad de algunos elementos de contenido, como pueden ser enlaces, scripts, objetos multimedia incrustados, perfiles de color, metadatos, firmas digitales, sistemas de protección de derechos de propiedad intelectual... Los recursos electrónicos deben preservar sus propiedades significativas a través del tiempo y la migración puede ponerlas en riesgo. Las propiedades significativas son características que dan valor de uso y vigencia al contenido de un documento u obra de arte. Por ejemplo: una firma digital, el color y contraste de una imagen, el formato y estructura de encabezamientos de un documento de texto, la forma y tamaño de un gráfico… Antes de proceder a la migración es necesario identificar esas propiedades para garantizar que se mantienen tras la migración de los objetos desde unos formatos de fichero o sistemas de codificación a otros. También se puede poner en peligro la integridad y la autenticidad de los documentos u obras de arte. Antes de migrar el fondo es necesario conseguir un protocolo de trabajo basado en la realización de pruebas previas con una muestra representativa de los objetos digitales a migrar. Se debe incluir, asimismo, un procedimiento de control de calidad de los resultados.

Se debe redactar por cada procedimiento de migración un documento donde se especifique detalladamente el procedimiento. El documento debe preservarse en el tiempo, junto con toda la documentación técnica del proyecto.

Los procesos de migración deben ser registrados dentro de los metadatos de preservación digital.

A la hora de plantear un proceso de migración es de utilidad el siguiente estándar: *UNE-ISO 13008:2013 Información y documentación. Proceso de migración y conversión de documentos electrónicos*[5]. Este estándar desarrolla el concepto de migración, diferenciando entre conversión (entendida como el proceso de cambio de formato de los documentos manteniendo las características de los documentos) y migración (entendida como el proceso de transferencia de los documentos desde una configuración de software o hardware -tal como una aplicación informática, base de datos, sistema operativo, dispositivo de almacenamiento- a otra sin cambiar su formato). El estándar aborda cómo realizar ambos procesos de forma sistemática y planteándose con suficiente cautela los riesgos que para la integridad, fiabilidad, autenticidad y fiabilidad de los documentos pueden suponer éstos.

[5] Esta norma puede ser adquirida desde la Web de AENOR http://www.aenor.es/aenor/normas/normas/fichanorma.asp?tipo=N&codigo=N0052128#.Vopm0fnhCUk

c) Emulación.

Consiste en recrear el entorno técnico requerido para acceder al contenido de un fichero digital o ejecutar una aplicación informática. Este proceso requiere mantener información sobre el software y el hardware empleados para la generación del objeto, para que el sistema pueda ser emulado en otro entorno informático diferente al del original. Requiere también la aplicación de programas informáticos *ad hoc*, los denominados emuladores. Los emuladores intentan recrear la funcionalidad e interfaces de las aplicaciones originales.

La emulación puede evitar la migración de los ficheros a nuevos formatos y puede convertirse una estrategia útil no a muy largo plazo para objetos digitales complejos. Pensemos que el emulador es una aplicación software que también necesita ser preservada digitalmente, pues está sujeta a obsolescencia como cualquier otro componente software.

d) Preservación de la tecnología.

Consiste en mantener tecnología que ya no está en el mercado y que ha sido ampliamente sobrepasada. Es una estrategia costosa porque exige mantener el hardware y software original y en condiciones de uso. El fabricante deja rápidamente de dar soporte técnico a productos discontinuados, por lo que llegará un momento en que sea muy difícil poder recambiar componentes de los equipos. Por eso no es conveniente usar esta estrategia más allá de un período corto tras el reemplazo de la tecnología.

e) Encapsulación.

Se trata de agrupar juntos en un mismo fichero contenedor serializado (como un fichero ZIP o RAR, por ejemplo) o carpeta de sistema de ficheros todos los componentes de información que conforman un documento u obra a preservar o que aportan datos o el contexto necesarios para su gestión, preservación o uso. En los casos más extremos la encapsulación incluye también las especificaciones técnicas de los formatos de fichero y las propias aplicaciones informáticas que los leen.

Es imprescindible encapsular también junto a éstos los metadatos de los objetos digitales a preservar, y en un sistema de representación estandarizado.

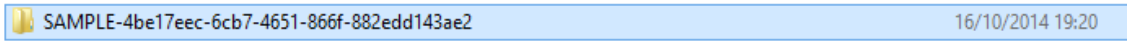
La encapsulación no deja de ser una medida a corto plazo, ya que los formatos de los ficheros, las aplicaciones que los leen, o los propios formatos de los ficheros contenedores o los sistemas de archivos de los soportes informáticos pueden quedarse obsoletos.

La Library of Congress ha desarrollado en los últimos años esta estrategia de preservación, sacando a la luz un estándar denominado Bagit (Bag-it o BagIt). Este estándar es muy sencillo y básico, aunque suficientemente potente como para que esté siendo aplicado de forma masiva en preservación digital. Por ello hemos optado por su utilización en este proyecto. Bagit intenta asentar un estándar al alcance de cualquiera usando herramientas y formatos de uso común: ficheros de texto plano (txt), carpetas, calculadoras hash... El sistema se

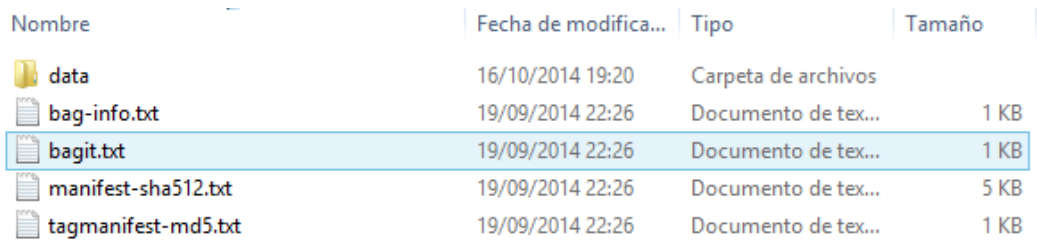
basa en incluir ("empaquetar") en una carpeta de sistema operativo (bag) el objeto u objetos digitales relacionados, organizados en carpetas si necesario, junto a uno o más ficheros en formato de texto plano que incluyen metadatos sobre el propio paquete y una relación de ficheros empaquetados y sus códigos hash respectivos. En un paquete (bag) es conveniente incorporar también un fichero que incluya los metadatos de los objetos en formato estándar, como podría ser METS, por ejemplo.

El empaquetamiento de preservación digital es, en realidad, un concepto más específico que el de encapsulación; refiere a la manera en que se estructuran física y lógicamente los paquetes de preservación digital que contienen los contenidos a preservar, sus metadatos y sus ficheros de control. Estos paquetes suelen ser denominados paquetes AIP en la terminología del estándar OAIS[6] y de los sistemas de preservación digital. Los estándares de encapsulación son muy laxos, no dicen nada sobre cómo unificar la estructura lógica y física de los paquetes de preservación digital AIP. Es la institución o la aplicación de preservación digital quien tiene que definir esta estructura. Veamos un ejemplo de paquete AIP del sistema de preservación digital Archivematica, que estudiaremos más abajo. La estructura de este paquete AIP se basa en los estándares BagIT y METS.

El paquete se denomina *SAMPLE-4be17eec-6cb7-4651-866f-882edd143ae2*. Su carpeta padre es:

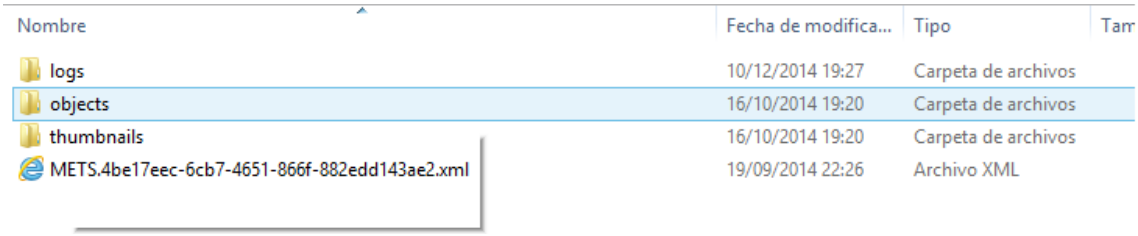


Si accedemos a la carpeta padre nos encontramos con la estructura normativa de Bagit ya explicada:



Si accedemos a la carpeta data encontramos la siguiente estructura:

[6] En los siguientes epígrafes definiremos y explicamos lo que es un OAIS y sus diferentes elementos, entre los cuales se incluye el concepto de paquete AIP.



Esta estructura forma parte ya del modelo de empaquetamiento propio de Archivematica. De acuerdo a este modelo tenemos tres carpetas normativas_

- logs. Almacena datos sobre la transferencia de objetos digitales al repositorio de preservación e información de salida de distintos procesos de preservación digital, como puede ser la identificación de los formatos de los ficheros a preservar o el cambio que pueden haber sufrido los nombres de los ficheros para prepararlos para la preservación digital.
- objects. Contiene los ficheros a preservar.
- thumbnails. Contiene miniaturas de los ficheros a preservar.

Además encontramos el fichero XML en formato METS denominado "METS.4be17eec-6cb7-4651-866f-882edd143ae2.xml" que contiene todos los metadatos de los objetos a preservar en su amplia tipología y las relaciones estructurales entre los ficheros y entre los objetos físicos a que corresponden. Más abajo describimos el estándar METS y cómo lo vamos a utilizar como sistema auxiliar del empaquetamiento de preservación digital.

2.2 Componentes del sistema de preservación digital.

A grandes rasgos, el sistema de preservación digital consiste en una serie de servicios que se aplican a un repositorio de preservación digital que sigue el modelo OAIS. El sistema está formado por cuatro niveles o perspectivas:

1. De contenidos. Incluye los contenidos a preservar y su sistema de organización, pensado para facilitar todas las tareas de preservación digital.
2. Hardware. Son los dispositivos informáticos de procesamiento y almacenamiento que contienen físicamente al depósito de preservación.
3. Servicios de preservación digital. Es la capa de procesos que se aplica a los contenidos en todas las fases de trabajo implicadas en la preservación digital: preparación para la ingesta en el depósito de preservación digital, ingesta y tareas activas de preservación digital.

4. Sistema de búsqueda y acceso a los contenidos preservados. Incluye la aplicación de búsqueda que permite el acceso directo a contenidos concretos o a sus agrupaciones y la extracción automática de conjuntos de contenidos.

En los epígrafes siguientes tratamos de describir las especificaciones de todos estos niveles.

2.2.1 El modelo de sistema de acuerdo al estándar OAIS.

OAIS ha sido ampliamente reconocido como el modelo base para la creación de sistemas de archivo de información digital que garanticen la preservación y acceso a largo plazo. Fue desarrollado por el Consultative Committee for Space Data Systems (CCSDS), pero se ha convertido en estándar ISO desde el año 2003 (ISO:14721:2003), aplicable a cualquier tipo de repositorio de información digital, no sólo a repositorios de datos espaciales. Este estándar ha sido actualizado en el año 2012[7].

El principal propósito del modelo de referencia OAIS es facilitar la comprensión de todos los aspectos que son requeridos para preservar y acceder a la información almacenada en un repositorio a largo plazo en el tiempo. De acuerdo a esta finalidad, el modelo no especifica o diseña ninguna implementación particular, ni siquiera se mencionan sistemas informáticos, o tecnologías concretas para la organización, gestión, proceso, almacenamiento y acceso a los datos; define, exclusivamente, un marco para comprender las relaciones entre todos los elementos de los que depende el acceso y la preservación digital a largo plazo en un archivo de información digital. Por ello, la creación de un sistema de archivo compatible con el modelo OAIS requiere el desarrollo de un nivel intermedio de modelos más concretos y detallados, así como de herramientas software que den soporte a las tareas específicas de preservación digital de estos otros modelos.

El modelo funcional de un OAIS se basa en las siguientes tareas:

- Ingreso o Admisión (Ingest). Son los servicios y funciones para la recepción de los objetos digitales procedentes de los productores y para prepararlos para su almacenamiento digital y gestión en el archivo.
- Almacenamiento (Archival Storage). Es la preservación de los objetos digitales en un archivo. Incluye los servicios y funciones necesarios para el almacenamiento, mantenimiento y recuperación de los documentos y sus metadatos. Incluye todas las estrategias de seguridad y de preservación digital aplicadas a lo largo del tiempo.

---

[7] ISO. Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model. ISO 14721:2012. Geneva: ISO, 2012. Equivalente a: CCSDS. Reference Model for an Open Archival Information System (OAIS). Recommended practice CCSDS 650.0-M-2. MAGENTA BOOK. June 2012 [en línea]. Washington, DC: CCSDS, 2012. Disponible en: http://public.ccsds.org/publications/archive/650x0m2.pdf .

- Gestión de datos (Data Management). Es la gestión, mantenimiento y procuración de acceso a los metadatos descriptivos que identifican y documentan los documentos (catálogos) y a los datos de tipo administrativo usados para la gestión del archivo.
- Gestión (Administration). Son todos los servicios y funciones para la gestión del archivo, incluyen desde la gestión de los nuevos ingresos hasta la gestión del sistema informático o el mantenimiento de las políticas y estándares del archivo.
- Planificación de la Preservación (Preservation Planning). Son las tareas de seguimiento del entorno exterior del archivo para tener en todo momento el estado tecnológico bien identificado y poder asegurar la no obsolescencia de la información y su accesibilidad a los usuarios.
- Acceso (Access). Son los servicios habilitados parar que los usuarios puedan acceder y recuperar los documentos que necesitan.

Estos seis servicios se complementan con otros que son denominados servicios comunes, tales como, servicios de red de datos, funciones de seguridad (control de acceso, autenticación e identificación de usuarios, comprobación de integridad de datos, etc.), servicios de red, etc.

El modelo de información se basa en los denominados Information Package o IP, que agrupan en una misma entidad el objeto digital junto a sus metadatos. Hay varios tipos de IP, el que es preservado en un OAIS se denomina Archival Information Package (AIP):

- SIP. Submission Information Package (Paquete de Información de Envío). El IP utilizado por los proveedores para enviar información al archivo.
- AIP. Archival Information Package (Paquete de Información de Archivo). El utilizado dentro del archivo para el almacenamiento y preservación de los objetos.
- DIP. Dissemination Information Package (Paquete de Información de Difusión). El que es usado para la distribución de los objetos al usuario final.

Veamos en el siguiente gráfico la relación entre tareas e IP[8]:

[8] Obtenido en CCSDS. Reference Model for an Open Archival Information System (OAIS). Recommended Practice. CCSDS 650.0-M-2. Magenta Book. June 2012. Disponible en: http://public.ccsds.org/publications/archive/650x0m2.pdf, pp. 4-1.

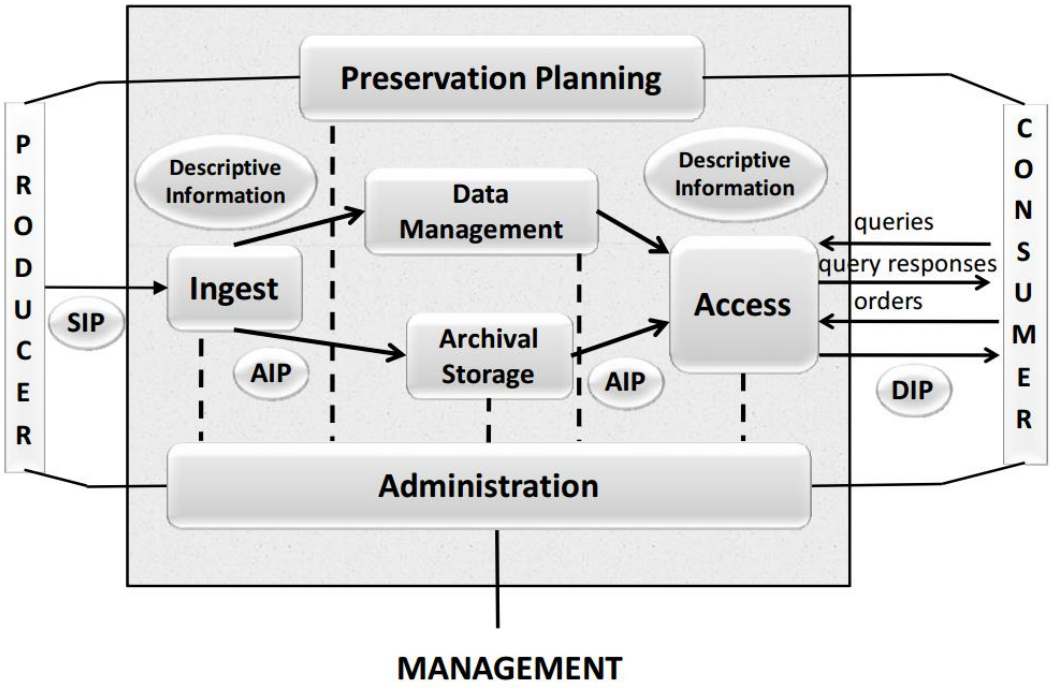


Figura 1. Modelo de componentes de un OAIS. Gráfico original de la publicación de OAIS, Magenta-Book (CCSDS 650.0-M-2)

2.2.2 Paquetes de información para la preservación digital.

Como venimos explicando, en un repositorio de preservación que siga el modelo OAIS los objetos digitales a preservar no se almacenan de manera aislada, sino en conjuntos que podemos denominar como paquetes de información o de preservación. Por ello, hablamos de un sistema o normativa de empaquetamiento sobre cuya base se generan estos paquetes de preservación. La función de los paquetes de preservación es aglutinar los objetos digitales a preservar junto a los otros objetos con los que guardan una relación de pertenencia a un mismo documento, obra o ejemplar y junto a la metainformación que permite identificarlos, controlarlos y contextualizarlos en cualquier momento.

Cada paquete corresponde a una obra física entendida como unidad individual (por ejemplo, un volumen de libro, un vídeo, una fotografía, un audio…) o como unidad a nivel de obra compuesta por varias unidades individualizables (por ejemplo, todos los volúmenes de una obra o todas las fotografías integradas en un portfolio).

De acuerdo a la terminología OAIS, vamos a denominar paquete SIP a cualquier paquete de contenido digital que contiene toda la información de una obra o documento (contenido más sus metadatos de cualquier tipología) remitidos al sistema de preservación por la institución, artista o colección (el remitente) que hace la aportación y que se ajustan a la normativa para los paquetes SIP del repositorio.

Vamos a entender por depósito una unidad de ingreso al sistema de preservación digital, esto es, el conjunto de ficheros que el remitente ingresa en el repositorio al mismo tiempo. El volumen y lo que comprende un depósito es variable, depende de lo que decida ingresar el remitente en una misma operación de ingreso. Un depósito podría ser un solo documento u obra artística, parte de una colección, una colección entera, o incluso todos los fondos digitales de una institución conteniendo múltiples colecciones. El concepto de depósito no se corresponde, por consiguiente, con ninguna unidad de organización de un fondo. Denominamos "Depósito SIP" al conjunto de paquetes SIP integrados en un depósito de remitente. Este concepto es equiparable al concepto "Data Submission Sessions" del modelo OAIS. El remitente puede ser cualquier institución que desee hacer una ingesta, previo acuerdo con los responsables del AEMA, en el sistema de preservación digital del AEMA.

Denominamos "Depósito PreSIP" a los depósitos de un remitente que están a la espera de comprobación del cumplimiento de normativa para SIP y Depósito SIP en la totalidad de todos su ficheros y metadatos, o que no cumplen esa normativa y están a la espera de que el sistema de preservación haga la transformación a SIP para que la cumplan y puedan ser ingresados en el repositorio. Denominamos PreSIP o paquete PreSIP a cada uno de los paquetes que conforman un depósito PreSIP. A cada documento u obra a ingestar en el sistema de preservación digital del AEMA le corresponde un PreSIP, que puede contener uno o muchos ficheros digitales.

El SIP debe estar normativizado, es decir, el sistema de preservación no puede admitir cualquier tipo de paquete, pues debe garantizar la factibilidad del servicio. La transformación del SIP del remitente a AIP de preservación es lo que se lleva a cabo por la aplicación de preservación digital antes de su ingesta.

El paquete de información (IP en terminología OAIS, que puede ser SIP o AIP) debe tener obligatoriamente vinculado física y lógicamente la información normativa de acuerdo a OAIS:

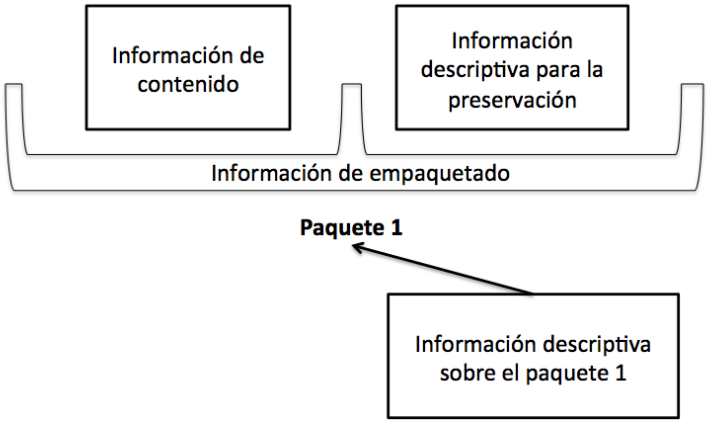


Figura 2. Basado en gráfico original en OAIS, Magenta-Book (CCSDS 650.0-M-2), p. 2-6.

- **Información-Contenido (Content Information o CI).** Es el contenido a preservar (*Content Data Object*, esto es, todos los bits

que conforman un fichero digital a conservar) junto con toda la información que se necesita para que los usuarios puedan comprender el contenido (*Representation Information*).

- **Información de Preservación (Preservation Description Information o PDI).** Son los metadatos necesarios para la preservación y uso en términos legales de la CI. Deben incluir toda la información necesaria para:
  - Documentar la historia de CI. Registrando origen, creador, tecnología y cambios diacrónicos técnicos, de contenido y de custodia del CI (Provenance).
  - Aportar contexto a CI: relaciones del CI con otros CIs y explicar por qué se ha creado CI (Context).
  - Asegurar que es identificable de manera única (Reference).
  - Asegurar su autenticidad e integridad (Fixity).
  - Asegurar la legalidad de cualquier operación de procesado técnico o de contenido requerido para la preservación o difusión, o de uso público o privado del CI (Access Rights).
- **Información de empaquetado (Packaging Information).** Son los datos que agrupan el CI con el PDI en una única entidad. Por ejemplo, información de que ambas entidades de información están vinculadas mediante un fichero METS y archivadas en una determinada carpeta de una unidad de disco.
- **Información Descriptiva (Descriptive Information).** Son metadatos para facilitar la recuperación del objeto y la realización de búsquedas. Por ejemplo, las fichas descriptivas en un catálogo.

En la normativa de empaquetamiento AIP que incorporamos más abajo se detalla con suficiente detenimiento la forma de implementación de estos conceptos OAIS en la arquitectura de modelo de paquete diseñada para el sistema de preservación.

En el siguiente epígrafe detallamos más el concepto de metadatos y su tipología, y lo hacemos de una manera más próxima a la implementación final de paquete AIP que vamos a desarrollar más abajo.

2.3 Metadatos a utilizar en la preservación digital y sus tipos.

Los metadatos son datos estructurados sobre los documentos u obras de arte cuyas versiones digitales van a ser depositadas en el sistema de preservación y sobre los propios objetos digitales que los representan. Los metadatos incluyen también datos sobre el propio proceso y operaciones de custodia y preservación digital u otras informaciones que son útiles para el control de los contenidos a preservar.

Distinguimos dos grupos de metadatos: metadatos que son empaquetados junto con los objetos custodiados y metadatos externos. Los primeros son los metadatos que los estándares y buenas prácticas de preservación digital recomiendan que sean recogidos y almacenados junto a los objetos a preservar (como por ejemplo los metadatos bibliográficos o los técnicos de los ficheros digitales). Estos metadatos se empaquetan junto a los objetos digitales obligatoriamente. Los segundos son metadatos que se van recogiendo durante los procesos de ingreso, salida y gestión de los depósitos de documentos u obras. Estos últimos son datos recogidos en el sistema de preservación digital, en su base de datos interna, para facilitar la eficiencia y eficacia de todas las operaciones del servicio de preservación digital. Algunos metadatos externos también se empaquetan junto a los objetos digitales al ser muy relevantes para su preservación digital, como, por ejemplo, los que registran los eventos que sufre un fichero para su migración a un formato no obsoleto o la forma de organización que presenta un paquete SIP.

Para seguir la mayor conformidad con los estándares, especialmente con METS y PREMIS[9], usaremos esta clasificación, que parte del acoplamiento, en ocasiones algo forzado, del estándar PREMIS en el estándar METS:

1. Descriptivos. Son datos de tipo bibliográfico, que describen física y formalmente los documentos u obras y que permiten identificarlos sin ambigüedad, su contenido, ubicación, etc. Son los datos que se consignan habitualmente de las catalogaciones bibliotecarias, archivísticas o museísticas siguiendo formatos normalizados (MARCXML, MODS, EAD, DC...)
2. Administrativos. Son datos que ayudan al conocimiento técnico y gestión de los objetos digitales.
    a. Técnicos de objeto digital. Datos sobre las características técnicas de los objetos digitales. En esta categoría entran los metadatos de objeto (Object) del estándar Premis.
    b. Procedencia digital. Datos sobre los eventos, y sus agentes, que ha sufrido un objeto digital desde su creación o captura. En esta categoría entran los metadatos de evento (Event) y agente (Agent) del estándar Premis.
    c. Derechos de propiedad intelectual. Información sobre los poseedores de los derechos de explotación de las obras y los permisos para su utilización. En esta categoría entrarían los estándares de metadatos METS-Rights y Premis Rights.
3. Estructurales. Datos que permiten vincular estructuralmente, ya sea a nivel físico o lógico los objetos digitales. En esta categoría entran los elementos

[9] Como explicamos más abajo, METS es el estándar de encapsulación elegido para normalizar la representación y vinculación de metadatos, objetos digitales y documentos de procedencia; y PREMIS el utilizado para normalizar metadatos específicos de preservación digital.

METS para mapas estructurales, directorios de ficheros y agrupaciones de ficheros.

Los metadatos específicos que facilitan la mejor realización de las operaciones técnicas de preservación digital se reparten entre los técnicos y los de procedencia digital.

Además podemos señalar otro tipo de metadatos, que podemos categorizar como de control, y que registran datos que permiten conocer el estado en que llegan los objetos digitales, sus estructuras anteriores de organización y operaciones técnicas de control de ingesta realizadas en las fases de preparación de ingreso y de ingreso.

3 Contenidos que se incluyen en el sistema de preservación digital.

En los paquetes de preservación AIP se incluye cualquier fichero representativo de la obra o documento de fondo de artista que se haya considerado necesario preservar, ya tenga la función de fichero máster o derivado.

La exportación de los metadatos obtenida automáticamente desde el sistema informático de gestión de las obras a un fichero XML será almacenada en la carpeta "metadatos_recibidos" del paquete AIP.

Se preservarán no solamente los ficheros digitales de las obras o documentos, sino todos los ficheros que contemplan las especificaciones de digitalización: másteres, derivados, metadatos descriptivos, ficheros de imagen de control, ficheros de datos de control y ficheros resultantes del control de calidad.

4 Normativa de nomenclatura para ficheros y carpetas.

Seguirá la normativa descrita en los protocolos de digitalización.

La aplicación informática, que hace la evaluación de cumplimiento de normativa SIP de nomenclatura y extensión de rutas y el ajuste automático de los nombres de fichero y carpetas para hacerlos entrar en normativa, aplicará un algoritmo predeterminado por defecto, que se encargará de hacer los cambios en los nombres no normalizados. Este algoritmo controlará: exceso de longitud de rutas de ficheros de acuerdo al límite asentado en la normativa, inclusión de caracteres prohibidos en los nombres de ficheros y carpetas, y exceso de longitud en nombres de fichero y carpetas.

La presencia de puntos u otros caracteres no permitidos en los nombres de fichero implica su sustitución por guiones bajos.

5 Sistema de organización del depósito de preservación digital.

Se ajustará lo más posible a la normativa descrita en los protocolos de digitalización, con las adaptaciones oportunas a las necesidades de la preservación digital.

Describimos a continuación unas especificaciones que se ajustan al requisito anterior. En los discos del sistema de preservación digital se seguirán estas pautas:

- La estructura de carpetas debe ser idéntica en todos los fondos a digitalizar no admitiéndose variaciones salvo en los casos debidamente justificados por los responsables del sistema de preservación digital del AEMA.
- Cada fondo tendrá una carpeta padre/raíz que contendrá todos los ficheros digitales de sus obras. Cuando los ficheros de un mismo fondo no quepan en un único disco, en todos los discos donde se almacenen tendrá que haber una carpeta raíz para el fondo. No se podrán mezclar los ficheros de diferentes fondos en una misma carpeta raíz de fondo. El nombre de la carpeta del fondo será un identificador nemotécnico de éste de pequeña extensión, tal como: “espacio-p” o “pedrogarhel”.
- Cada colección dentro del fondo tendrá una carpeta padre cuyo nombre será un identificador nemotécnico de la colección, tal como “fotografia”, “dibujos”, “acuarelas”, “vídeos”...
- Dentro de cada colección se presentará una carpeta padre por cada documento u obra. El nombre de la carpeta padre deberá llevar un identificador único UUID, tal y como explicamos más abajo en la normativa de empaquetamiento. La estructura dentro de la carpeta padre seguirá sin

variación alguna la estructura descrita más abajo en la normativa de empaquetamiento.

6 Normativa para los paquetes PreSIP.

Es la normativa a seguirse para aceptar discos con contenido para ser preservados en el sistema de preservación del AEMA. Al menos se exigirá:

- La entrega en discos duros externos conectables a un PC a través de puerto USB. Los discos duros deben estar en buen estado y no obsoletos.
- Una estructura de carpetas que agrupe todos los ficheros de la obra en una carpeta padre, con subcarpetas diferenciadas para los másteres, los derivados, los ficheros de control y los metadatos. Se admitirán otras estructuras siempre y cuando permitan hacer la agrupación con facilidad y poco tiempo, ya sea manual o automática.
- Cada carpeta de obra debe contener una carpeta con los metadatos descriptivos en el formato establecido en esta normativa. O al menos un formato estandarizado de uso común.
- Se seguirá la normativa de nomenclatura de ficheros y carpetas dadas en las especificaciones de digitalización, o al menos, se evitará el uso de aberraciones en nombre de fichero y carpeta y longitudes de ruta. Entendemos por aberraciones el uso de caracteres o longitudes de nombre prohibidos por el sistema operativo Windows en sus últimas versiones.
- Se aportará en una carpeta así especificada un fichero con los metadatos de custodia acordes a la normativa dada más abajo, en el apartado de metadatos, a este respecto. El fichero se denominará “datos_custodia.xml”.
- Se aportará en una carpeta así especificada, un fichero con los datos de propiedad intelectual aplicables al fondo a preservar. Los datos se ajustarán a lo que se describe más abajo a este respecto, en el apartado de metadatos.
- No se aceptarán carpetas vacías o carpetas que contengan otras carpetas vacías.
- No se aceptarán envíos de depósito en carpetas comprimidas.
- No se aceptarán envíos de depósito como imágenes de disco.
- No se aceptará la inclusión de ficheros del sistema operativo.
- No se aceptará la inclusión de ficheros o directorios ocultos.
- En el caso de ficheros que por necesidad tengan que estar vinculados a otros para evitar la pérdida de características intrínsecas relevantes o de contenidos vinculados, éstos tendrán que estar en el lugar que les corresponde para que no se pierda su vinculación con el fichero principal. Es el caso de ficheros de perfiles de color ICC, de hojas de estilo para

documentos en formato HTML o XML, o de objetos vinculados a páginas HTML, por poner algunos ejemplos.

7 Normativa para los paquetes SIP.

Es la normativa que hay que seguir para dejar la ingesta en un modo que se pueda automatizar la ingesta por el sistema de preservación digital haciéndose automáticamente los AIP y la estructura de organización del disco de preservación.

No se seguirán los estándares PAIMAS, PAIS ni XFDU[10], pese a que han sido diseñados para la codificación de este tipo de paquetes de información. La razón es que estamos ante modelos muy complejos, que han sido ideados para las necesidades de transferencia de objetos e información espacial al archivo de preservación en el contexto de centros de investigación astronómica. Fuera de este contexto pueden ser modelos extremadamente complejos que compliquen innecesariamente el trabajo con los datos y objetos digitales a preservar.

La estructura normativa para los paquetes PreSIP dada anteriormente debería ser suficiente para la automatización de la ingesta. Si así no fuera, se harán las adaptaciones necesarias para este proceso, de acuerdo a las indicaciones del técnico de preservación digital.

8 Plan de preservación digital para la normalización de la ingesta.

Los formatos de preservación y codecs para los distintos tipos de objetos digitales deberán seguir los cinco criterios recomendados por la Digital Preservation Coalition [11]:

1. Adopción. Refiere el grado de generalización en el uso del formato. Se deberá optar por formatos contenedores y codecs ampliamente usados a nivel mundial.

---

[10] PAIMAS está descrito en: *Producer-Archive Interface Methodology Abstract Standard. Issue 1. Recommendation for Space Data System Practices (Magenta Book), CCSDS 651.0-M-1*. Washington, D.C.: CCSDS, May 2004. [Equivalent to ISO 20652:2006.]. PAIS está descrito en: ISO 20104:2015. XFDU está descrito en: *XML Formatted Data Unit (XFDU) Structure and Construction Rules. Issue 1. Recommendation for Space Data System Standards (Blue Book)*, CCSDS 661.0-B-1. Washington, D.C.: CCSDS, September 2008. [Equivalent to ISO 13527:2010.]

[11] Todd, Malcolm (2009). Technology Watch Report File formats for preservation. Digital Preservation Coalition. 43 p. Disponible en: http://www.dpconline.org/component/docman/doc_download/375-file- formats-for-preservation

2. Independencia de la plataforma. Refiere el grado de independencia del formato o códec de plataformas hardware y software específicas. Se deberán usar formatos y codecs sin ningún problema de reconocimiento en todos los sistemas operativos de amplio uso en la actualidad en sus diferentes versiones o distribuciones, y muy especialmente Mac OS, Linux y Windows.
3. Divulgación. Refiere si la especificación del formato es de dominio público y es fácilmente accesible. Sólo se deberá optar por formatos y códecs cuyas especificaciones sean abiertas y fácilmente conseguibles desde la Web, o, en su defecto, que estén descritas en estándares que operen internacionalmente.
4. Transparencia. Indica la facilidad con la que el formato de archivo puede ser identificado y sus datos de contenido descodificados de acuerdo a su especificación técnica.
5. Soporte de metadatos. Indica si el formato contenedor cuenta con la posibilidad de incrustar metadatos en los ficheros y de que estos sean extraíbles con facilidad, así como el nivel de compatibilidad de esos metadatos con estándares de metadatos que operen a nivel internacional. Sólo se admitirán formatos contenedores con la posibilidad de incrustar metadatos técnicos descriptivos y administrativos de diverso tipo que sean necesarios para la contextualización de los objetos digitales.

El sistema de preservación deberá desarrollar un mecanismo de registro de los datos de formatos en su base de datos, que incluya los formatos y versiones de éstos, agrupados por tipo de medio, que se encuentren efectivamente en el repositorio. Este registro debería incluir los siguientes datos:

- Identificador de formato PUID (PRONOM Unique Identifier).
- Identificador propio en el sistema de preservación.
- Nombre del formato.
- Versión del formato.
- Extensión.
- Tipo MIME.
- Riesgo del formato.
- Formatos recomendados de migración.
- Formatos relacionados.
- Descripción.
- Código (binario, texto).
- orden de los bytes (big-endian, little-endian).
- Firmas digitales, internas y externas.
- Sistemas de compresión.
- Codificación de caracteres.
- Dependencias tecnológicas.

- Derechos.
- Propiedades inherentes, propiedades de instancia.

Además de identificar y registrar la información sobre los formatos presentes en el repositorio, será preciso definir las estrategias de migración, a qué formatos y qué versiones, y con qué herramientas puede llevarse a cabo esta migración.

Tipo de medio	Formatos de fichero	Sistema de compresión admitido	Codificación de caracteres en textos	Formatos de preservación	Herramienta para la normalización
Imágenes raster	TIFF	Ninguno	N.A	Para imágenes en color y escala de grises: TIFF sin compresión o TIFF con compresión sin pérdida mediante algoritmos sin riesgo cercano de obsolescencia, como LZW. Para imágenes en blanco y negro puro (bitonal) TIFF con compresión sin pérdida en algoritmos de amplio uso (como CCITT g4 o g3, o LZW).	ImageMagick (http://www.imagemagick.org/) Adobe Photoshop

Tipo de medio	Formatos de fichero	Sistema de compresión admitido	Codificación de caracteres en textos	Formatos de preservación	Herramienta para la normalización
				No se admitirá TIFF con compresión JPEG o ZIP. No se admitirán imágenes en paleta de colores.	
	JPG			Se mantiene como formato de preservación para los derivados y miniaturas, pero no para los másteres[12]. Se mantiene como formato de preservación en su versión JFIF para los derivados de previsualización y también	JPGtoTIFF (GraphicsMagick) JPGtoTIFF (ImageMagick) ImageIO Adobe Photoshop

[12] No obstante, será un aspecto a determinar con el remitente si desea mantener el formato JPEG como formato de los másteres en los casos en que sus versiones másteres estén en este formato, en lugar de su transformación a TIFF. Si el remitente acepta mantener sus másteres en JPEG se hará constar esto en el plan de preservación específico para sus fondos.

Tipo de medio	Formatos de fichero	Sistema de compresión admitido	Codificación de caracteres en textos	Formatos de preservación	Herramienta para la normalización
				para los másteres 13 de las imágenes para las que no haya formato TIFF alternativo. No se recomienda su conversión a TIFF, ya que no se gana en calidad y puede llevar a error en un futuro tener un TIFF resultado de la transformación de un JPEG. Recordemos que JPEG es estándar también.	
	JP2			No se admite JPEG2000 como formato de preservación	

[13] La normativa del sistema de preservación indica que, no obstante el plan de preservación general, será un aspecto a determinar con el remitente si desea mantener el formato JPEG como formato de los másteres en los casos en que sus versiones másteres estén en este formato, en lugar de su transformación a TIFF. Si el remitente acepta mantener sus másteres en JPEG se hará constar esto en el plan de preservación específico para sus fondos.

Tipo de medio	Formatos de fichero	Sistema de compresión admitido	Codificación de caracteres en textos	Formatos de preservación	Herramienta para la normalización
				.	
	RAW nativo de cámaras digitales (CR2, NEF, IIQ…)	El del propio RAW	-	DNG con compresión sin pérdida y fichero RAW nativo incrustado. Si por el aumento de espacio de almacenamiento se estima inviable incrustar el RAW original, se admite DNG con compresión sin pérdida. Los metadatos sidecar en formato XMP que genera el revelador Adobe Camera RAW se deberán incrustar en el propio formato RAW.	Adobe DNG Converter o Adobe Camera RAW en sus últimas versiones.
Portable Document Format	PDF			PDF/A PDF 1.4 o superior.	Ghostscript (http://www.ghostscript.com/) PDF/A Converter

Tipo de medio	Formatos de fichero	Sistema de compresión admitido	Codificación de caracteres en textos	Formatos de preservación	Herramienta para la normalización
					Adobe preflight Adobe Acrobat
Texto plano	TXT			Formato original	None
Audio	AC3			WAVE (LPCM)[14]	FFmpeg (http://ffmpeg.org/)
	AIFF				
	MP3, MP2				
	WAV				
	WMA				
Video	AVI			Decisión a tomar por los responsables del sistema de preservación.	FFmpeg (http://ffmpeg.org)
	FLV				
	MOV				

[14] De la misma manera que en el caso de másteres que llegan en JPEG, cuando lleguen versiones másteres de ficheros de audio en formatos que no sean de preservación pero considerados como estándares o de amplio uso, como por ejemplo MP3 o MP2, será un aspecto a determinar con el remitente si desea mantener esos formatos como formatos de los másteres, en lugar de su transformación a WAVE. Si el remitente acepta mantener sus másteres en esos formatos se hará constar esto en el plan de preservación específico para sus fondos.

Tipo de medio	Formatos de fichero	Sistema de compresión admitido	Codificación de caracteres en textos	Formatos de preservación	Herramienta para la normalización
	MPEG-1				
	MPEG-2				
	MPEG-4				
	SWF				
	WMV				
Interactivos multimedia	DIR (Macromedia o Adobe Director)			DIR	Se dejarán los formatos tal y como están sin migrar o recrear en nuevos formatos, ya que se plantea como mejor estrategia la virtualización de sistemas operativos donde corren las diferentes versiones de Director que pueden abrir las obras interactivas.
Imagen de disco	Cualquier formato de imagen de disco, de soportes tipo CD, DVD, Blu-			Formato ISO, extensión .iso[15] para imágenes de discos	

[15] Se pueden utilizar múltiples aplicaciones para hacer la conversión desde un formato de imagen de disco no obsoleto y de amplio uso al formato de imagen ISO, tal como Magic ISO Maker.

Tipo de medio	Formatos de fichero	Sistema de compresión admitido	Codificación de caracteres en textos	Formatos de preservación	Herramienta para la normalización
	Ray o disquete, no obsoleto en el momento del ingreso.			ópticos. Formato .img o .ima[16] para imágenes de disquetes. No se admitirá en ningún caso la compresión de imágenes de disco.	

Algunos medios, como es patente en el caso del vídeo, no cuentan todavía con un formato de preservación digital unánimemente admitido por la comunidad de expertos en vídeo digital y en preservación digital ni en lo que respecta a los contenedores ni a los formatos. Algunas instituciones se decantan por MXF como formato contenedor, y Motion JPEG2000 con compresión sin pérdida como formato (Library of Congress), debido al amplio uso que se les está dando en los últimos años en las televisiones y archivos de televisión. Otras optan por Matroska (Mkv), con FFV1 como formato, debido a su carácter abierto y mayor sencillez de codificación. Otras han venido optando por AVI como contenedor con formatos no comprimidos o con compresión sin pérdida. La decisión final para el formato de preservación digital en los vídeos deberá ser tomada por los responsables del sistema de preservación.

[16] El formato .img y el .uma fueron ideados inicialmente para contener imágenes de disco de disquetes, aunque se han usado por algunas aplicaciones, usualmente junto con ficheros de metadatos adicionales, también para contener imágenes de otros tipos de soportes de almacenamiento, como discos duros o soportes ópticos. También sea venido usando para contener imágenes en formato ISO. La extensión .img, genera, consiguientemente mucha confusión. Para contenidos distribuidos en disquetes, sólo admitiremos estos formatos como imágenes de disquetes, no de otros tipos de soportes, pues de otra forma las aplicaciones de virtualización no podrán trabajar con ellos. El depositario o el servicio de preservación digital, en caso de tener que crear imágenes de disco de disquetes o de sus contenidos, deberá usar una aplicación de confianza para su generación y siempre en modo Imagen de Disco de Disquete, para el tipo de disquete originario donde estaban contenidos los ficheros, y le dará la extensión .img o .uma.

En las especificaciones de digitalización de Espacio P, para los formatos de peor calidad técnica y que pueden presentar materiales resultado de grabación y recodificación desde las cintas máster U-Matic se ha recomendado un códec profesional de tipo intermedio, pues, a pesar de incluir compresión con pérdida, se adaptan bien a los posibles procesos posteriores de edición digital y subtitulado, y representan un equilibrio entre calidad de imagen y sonido y optimización de recursos de almacenamiento, procesado y ancho de banda. En concreto, se ha recomendado ProRes SD 422 en formato contenedor Quicktime (MOV). Para las cintas máster U-Matic se ha recomendado un códec que no implique compresión alguna, o, en su defecto, compresión con pérdida, tal como v210 YUV o YUV sin compresión, sobre el contenedor AVI o Quicktime. Todas estas opciones derivan en formatos de fichero altamente legibles en entornos Windows, Mac y Linux.

Para las versiones derivadas, en caso de generarse durante la fase de ingesta para ser custodiadas en el sistema de preservación se recomienda las especificaciones dadas para las digitalizaciones del fondo de archivo Espacio P, que son formato contenedor MP4 con códec H.264 y resolución SD.

De la misma manera que en el caso de másteres de audio que llegan en formatos que no sean admitidos como de preservación, cuando lleguen versiones másteres de ficheros de vídeo en formatos que no sean de preservación pero considerados como estándares o de amplio uso, como, por ejemplo, MPEG, será un aspecto a determinar con el remitente si desea mantener esos formatos como formatos de los másteres, en lugar de su transformación al formato contenedor y al sistema de codificación admitido como de preservación. Si el remitente acepta mantener sus másteres en esos formatos o sistemas de codificación no admitidos como de preservación, se hará constar esto en el plan de preservación específico para sus fondos.

Es muy importante garantizar que no se pierden las propiedades intrínsecas significativas en los procesos de migración. Si no se configura correctamente el proceso es fácil perderlas, por ejemplo, se podría perder la profundidad de bits de entrada en un archivo de imagen ráster o su espacio de color. Por lo que se recomienda que los responsable del sistema de preservación se asesoren previamente por expertos antes de definir los procesos de migración y normalización.

Se admitirá además de la versión 6 de TIFF la versión 5.0, pues es compatible al 100% con la anterior para imágenes que no contengan ningún elemento introducido por la versión 6. Las imágenes que conforman perfectamente una especificación TIFF anterior son también conformes a la TIFF 6.0 (Compatibilidad hacia atrás). Por tanto, se admite también TIFF en versión 5.0 como formato de preservación para las imágenes compatibles con TIFF 5.0 y con TIFF 6.0, es decir, que no añadan alguna característica no contemplada por la 5.0 y para la que sólo valdría la 6.0.

El plan de preservación inicial que aquí proponemos deberá ser ajustado a los requerimientos y preferencias del remitente. Una vez aceptado por ambas partes será codificado en la aplicación de proceso AIP del sistema de preservación digital.

9 Normativa general de empaquetamiento para los AIP.

9.1 Objetivos y necesidad de la normativa para empaquetamiento y representación de los AIP.

La finalidad de esta normativa es asentar un sistema de empaquetamiento acorde con el estándar del modelo de sistema de archivo de preservación OAIS[17] y de certificación TRAC (*Trustworthy Repositories Audit and Certification Criteria*)[18]. OAIS es el modelo tomado por TRAC, aunque TRAC describe una serie de requisitos a considerar en un proceso de certificación que no podemos pasar por alto. OAIS no asienta un formato de empaquetamiento, sino una serie de requisitos que tienen que cumplir los paquetes AIP, que son los que hemos tratado de seguir. Hemos seleccionado un sistema relativamente sencillo de implementar y de entender a nivel humano y a nivel máquina, de forma que se simplifiquen los procesos humanos y automatizados, no se requiera de un programa específico para su descodificación ni codificación, y cualquier usuario pueda tanto empaquetar como desempaquetar los contenidos con las aplicaciones típicas de escritorio al alcance de cualquier usuario de un ordenador. La lectura de los paquetes y de sus ficheros de control y metadatos se podrá hacer desde cualquier explorador de archivos de cualquier sistema operativo y un Bloc de Notas o aplicación de lectura de ficheros de texto simple (TXT).

Esta normativa puede quedar desactualizado en uno o dos años, debido a la rápida evolución de la tecnología informática y de los estándares bibliotecarios y archivísticos. Por tanto, será el propio equipo encargado del sistema de preservación el que actualice y mantenga en perfecto estado de actualización la normativa.

9.2 Alcance y consideraciones sobre el registro de las diferentes versiones de la normativa para empaquetamiento y representación de los AIP.

La normativa de empaquetamiento y representación de los AIP cobra mucha relevancia en los estándares de preservación digital citados, de tal manera que se exige que quede registrada por escrito y que en cualquier momento pueda establecerse la vinculación entre la normativa aplicada a un AIP y el AIP que la sigue. Es requisito, por tanto, para el sistema de preservación que las diferentes versiones de la normativa AIP se conserven convenientemente y quede registro de

---

[17] ISO. Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model. ISO 14721:2012. Geneva: ISO, 2012. Equivalente a: CCSDS. Reference Model for an Open Archival Information System (OAIS). Recommended practice CCSDS 650.0-M-2. MAGENTA BOOK. June 2012 [en línea]. Washington, DC: CCSDS, 2012. Disponible en: http://public.ccsds.org/publications/archive/650x0m2.pdf .

[18] ISO 16363: 2012. Space data and information transfer systems -- Audit and certification of trustworthy digital repositories. (Basado en el Libro magenta, 2011. Disponible en: http://public.ccsds.org/publications/archive/652x0m1.pdf)

la versión que se aplica a cada paquete AIP. En el modelo de empaquetamiento descrito más abajo se indica cómo consignar este dato en los ficheros de control del paquete y en el sistema de base de datos para la gestión del repositorio.

La normativa para la construcción de los AIP debe figurar en documento aparte, teniendo cada documento un nombre único de manera que pueda ser identificado en cualquier momento sin ambigüedad y no se corra el riesgo de borrado accidental o de pérdida de integridad referencial con respecto a cada paquete AIP al que se ha aplicado.

9.3 Formatos de fichero y sistemas de compresión y codificación permitidos en los AIP.

Los formatos de fichero y sistemas de codificación o compresión admitidos serán los que figuren en el Plan de Preservación Digital específico para este proyecto identificados como de archivo de preservación digital. Este plan tendrá que ser actualizado de forma continua en el tiempo, por lo que tanto el listado de formatos de fichero como el de los sistemas de codificación y compresión admitidos variarán de forma diacrónica.

9.4 Sistema de empaquetamiento y estructura interna del paquete AIP.

9.4.1 Seguimiento de estándares.

Además del estándar OAIS y TRAC, hemos considerado el estándar de la Library of Congress de los Estados Unidos denominado BagIt[19], debido a su sencillez y al seguimiento que se le ha dado en diferentes sistemas de preservación digital de código abierto, como es el caso del conocido sistema Archivematica de la empresa Artefactual[20].

Bagit asienta un sistema de empaquetamiento consistente en reunir en una carpeta/directorio de sistema operativo (Bag) el objeto u objetos digitales relacionados con un documento, obra u otra unidad documental. Dentro de esta carpeta padre se pueden crear carpetas hijo para poder organizar de una forma más estructurada los ficheros. Además de estas carpetas y los objetos digitales, el

---

[19] Podemos acceder a información y herramientas sobre este sistema en:

http://www.digitalpreservation.gov/documents/bagitspec.pdf

http://www.digitalpreservation.gov/multimedia/videos/bagit0609.html

http://www.youtube.com/watch?v=14ZPtYLtUYA

http://en.wikipedia.org/wiki/BagIt

http://tools.ietf.org/html/draft-kunze-bagit-14

[20] Puede accederse a una explicación extensa de este sistema en https://www.archivematica.org/wiki/Main_Page

paquete (Bag) incluye uno o más ficheros TXT que contienen metadatos sobre el propio paquete y una relación de ficheros empaquetados junto a sus códigos hash respectivos. BagIt recomienda incorporar también un fichero que incluya los metadatos de los objetos en formato estándar. La única imposición para la codificación de los ficheros TXT es que se haga siguiendo el formato de codificación de caracteres UTF-8 (8-bit Unicode Transformation Format).

La carpeta base que contiene el Bag puede ser serializada junto con su contenido, incluyéndolas en un fichero único que permita incrustar ficheros y carpetas, tal como ZIP o TAR. Es lo más recomendable para evitar el riesgo de pérdida de contenido de un paquete en un proceso de transferencia entre diferentes sistemas de almacenamiento o dentro de una misma estructura de disco, aunque supone un riesgo añadido al superponer un nuevo sistema de codificación que puede quedarse obsoleto en un plazo no muy alto. Bajo la consideración de este riesgo, no hemos contemplado la serialización para estas normas de empaquetamiento. En este caso, el fichero serializado debe tener como nombre el nombre de la carpeta base. Veamos con más detenimiento los posibles contenidos de un paquete BagIt:

- Carpeta “data”. Contiene los ficheros del objeto digital que se empaqueta, puede contener subcarpetas que estructuren de una forma comprensible esos ficheros. Pero no puede contener carpetas vacías.
- Fichero(s) en formato texto (“manifest-xxxxxx.txt”). Contiene una relación de los nombres de fichero incluidos en el Bag antecedidos por la ruta de carpetas desde la carpeta “data” y por los códigos hash generados. En el nombre del fichero manifest se sustituyen las xxxxx por la abreviatura del algoritmo hash que se haya usado. Por ejemplo: “tagmanifest-md5.txt”, si se ha usado MD5.
- Fichero “bagit.txt”, que identifica la carpeta como un Bag y contiene la versión de la especificación BagIt que se ha usado y la codificación de caracteres utilizado para los ficheros TXT.
- Opcional. Un fichero “bag-info.txt”, que detalla los metadatos del Bag, bajo la forma de pares campo/valor separados por dos puntos.
- Opcional. Un fichero “tagmanifest-xxxxx.txt” que enumera los archivos TXT y sus códigos hash, por ejemplo “tagmanifest-md5.txt”
- Opcional. Un fichero “fetch.txt” que contiene URLs de objetos no almacenados con el paquete.

Un ejemplo de paquete BagIt a primer nivel es:

En el sistema de empaquetamiento AIP se fuerza al uso de códigos UUID, como explicaremos más abajo.

9.4.2 Sistema de empaquetamiento general a nivel de documento.

En este epígrafe se describe el sistema de empaquetamiento general. Este sistema es una base sobre la que partir en la creación de normativas de empaquetamiento ajustadas a los diferentes tipos de objetos que puedan ser ingestados en el sistema de preservación, si se estimara así necesario a lo largo del período de vida del sistema de preservación digital. Todos los aspectos particulares del empaquetamiento que impliquen cambios con respecto a este sistema son descritos más abajo en los epígrafes de normas específicas para el tratamiento de cada tipo de objeto.

El AIP contiene obligatoriamente y con este sistema de estructuración:

- Una **carpeta padre** para cada documento u obra física digitalizada, cuya denominación será el nombre de fichero que tienen los ficheros máster correspondientes en el fondo digital actual, sin incluir la extensión de fichero, más un código UUID separado del anterior por un guion medio. En el caso de que existan ficheros para ambas caras de la misma obra o para las distintas páginas u hojas de un documento u obra multipágina, la carpeta padre no incluirá los caracteres identificativos del tipo de cara (Anverso/Reverso) o del número de página u hoja, aunque, evidentemente, todos estos ficheros serán empaquetados juntos. El código UUID se calculará automáticamente sobre la marcha en el momento de la conversión del SIP al AIP. El nombre aportado por el remitente sufrirá un proceso de ajuste en el caso de haberse usado caracteres no permitidos para los nombres de carpeta en los sistemas operativos más habituales o sobrepasar un número determinado de caracteres. Esta normativa de nomenclatura se debe mantener actualizada y se incluye en el apartado específico de nomenclatura que hemos incluido más arriba. La carpeta padre contendrá todos los ficheros de una misma obra junto a sus metadatos. Insistimos en que para las obras multipágina o compuestos de varios objetos digitales, se deberá generar obligatoriamente sólo una carpeta por obra (libro, número de revista, álbum de fotografías, portfolio…), que integre los ficheros que la componen.

    - Una **carpeta hija** denominada “data” que contendrá los ficheros que se empaquetan, sus metadatos y los ficheros de control que indique la normativa de empaquetamiento más actualizada, de acuerdo a la siguiente estructura:

- Una **carpeta hija** denominada “logs_datos_sip” que contendrá los ficheros que incluyen los datos del paquete SIP original y sus ficheros. La finalidad de esta información es permitir un proceso de reconstrucción íntegra y casi exacta del SIP a partir del AIP, así como documentar perfectamente el SIP en el AIP. Esos ficheros y sus contenidos son:

    - Fichero denominado “listado.txt” que contenga un listado de carpetas y sus ficheros contenidos en el paquete SIP correspondiente al AIP con los datos de nombre, tamaño, fecha y hora de última modificación, tamaño en bytes y la identificación de si es carpeta o fichero (sería el resultado de aplicar el comando DIR /A /Q, del sistema operativo Windows, a cada una de las carpetas contenidas en el SIP incluyendo la carpeta padre contenedora). Al comienzo del fichero se deberá describir a modo de comentario mediante un texto muy breve, de no más de una línea, que es lo que contiene este fichero, precedido del carácter #. La línea de comentario debe finalizar obligatoriamente con un carácter Intro (Retorno de carro y salto de línea). En el caso de que los nombres de ficheros y carpetas del paquete SIP no coincidan con los del PreSIP[21], se usarán los nombres del PreSIP para garantizar que se registran los nombres de fichero y carpeta tal y como son aportados por los remitentes, sin la normalización de nomenclatura de la fase SIP. En el caso de que el paquete SIP contuviera ficheros que durante su proceso de limpieza hayan sido borrados, estos ficheros borrados serán incluidos en el listado pero se pondrá a su derecha la frase “NO PRESERVADO. ELIMINADO del SIP.”

    - Fichero denominado “tab_corp.txt” que contenga una tabla de correspondencia entre los nombres de ficheros y carpetas del SIP correspondiente con los nombres de ficheros y carpetas del AIP, con el siguiente formato: cada fila tendrá los datos de una

[21] Un PreSIP es un paquete SIP previo a la normalización para cumplimiento de normativa para los paquetes SIP o que aún no ha sido validado como SIP

correspondencia de ficheros o carpetas en la forma *nombre en SIP, nombre AIP*. Las filas se separarán por un carácter Intro. Se pondrán las rutas completas de los ficheros. Si a una sola carpeta de origen en SIP corresponden varias en el AIP se repetirá la fila tantas veces como carpetas correspondan en el AIP, teniendo la columna para el SIP el mismo valor de nombre de carpeta SIP. Si ocurre a la inversa, se repetirán también la fila pero ahora el valor común será para la carpeta AIP. Al comienzo de este fichero se abrirá una línea extra que contendrá dos elementos, de izquierda a derecha: normativa_AIP, seguido del nombre de fichero identificador único de la normativa de empaquetamiento y representación AIP aplicada. Este dato deberá ser registrado asimismo en el registro de datos correspondiente para el AIP en el sistema de gestión del repositorio. Al comienzo del fichero se deberá describir a modo de comentario mediante un texto muy breve, de no más de una línea, que es lo que contiene este fichero, precedido del carácter #. La línea de comentario debe finalizar obligatoriamente con un carácter Intro (Retorno de carro y salto de línea). En el caso de que los nombres de ficheros y carpetas del paquete SIP no coincidan con los del PreSIP, se usarán los nombres del PreSIP para garantizar que se registran los nombres de fichero y carpeta tal y como son aportados por los remitentes, sin la normalización de nomenclatura de la fase SIP.

- Fichero denominado “sip_estr_crp.txt” que contenga la estructura original de carpetas y ficheros del paquete SIP. Con el formato de presentación de datos, no necesariamente de estructura, que vemos representado en el siguiente ejemplo:

```
.Tratado_de_botanica
|_.derivados
| |_P1050152.jpg
| |_P1050154.jpg
| \_P1050155.jpg
|_.masteres
| |_P1050152.tif
```

```
| |_P1050154.tif
| \_P1050155.tif
|_.miniaturas
| |_P1050152.jpg
| |_P1050154.jpg
| \_P1050155.jpg
\_Tratado_de_botanica_METS.xml
```

En este ejemplo apreciamos como la diferencia entre carpeta y fichero se establece por el carácter "." que llevan las carpetas como identificador al comienzo de su nombre. La representación de la estructura se basa en la combinación de los caracteres siguientes: *espacio en blanco* | _ \ , que corresponden respectivamente a los códigos ASCII: 32, 124, 95 y 92. Al comienzo del fichero se deberá describir a modo de comentario mediante un texto muy breve, de no más de una línea, que es lo que contiene este fichero, precedido del carácter #. La línea de comentario debe finalizar obligatoriamente con un carácter Intro (Retorno de carro y salto de línea).

Se añadirá al final de cada nombre de fichero el código hash que le corresponde.

En el caso de que los nombres de ficheros y carpetas del paquete SIP no coincidan con los del PreSIP, se usarán los nombres del PreSIP para garantizar que se registran los nombres de fichero y carpeta tal y como son aportados por los remitentes, sin la normalización de nomenclatura de la fase SIP.

En el caso de que el paquete SIP contenga ficheros que serán eliminados durante el proceso de limpieza del paquete, esos ficheros eliminados deberán constar en su lugar correspondiente junto a sus códigos hash. Hemos de pensar que este fichero de control se usará para documentar con la mayor exactitud posible el paquete SIP, por lo que no se puede perder esta información.

- Fichero denominado "Id_form_fich.txt" que contenga el resultado de la identificación del formato de los ficheros del SIP correspondiente, con el formato: nombre de fichero, nombre del formato, nombre de versión del formato, Format Registry Name (nombre del formato aportado por el sistema de registro utilizado) y Format Registry Key (código identificador único del formato de acuerdo al sistema de registro utilizado). Correspondiendo cada fila de datos a un fichero del SIP. Se usará como separador de fila el carácter Intro. Al comienzo del fichero se deberá describir a modo de comentario mediante un texto muy breve, de no más de una línea, que es lo que contiene este fichero, precedido del carácter #. La

línea de comentario debe finalizar obligatoriamente con un carácter Intro (Retorno de carro y salto de línea). En el caso de que los nombres de ficheros y carpetas del paquete SIP no coincidan con los del PreSIP, se usarán los nombres del PreSIP para garantizar que se registran los nombres de fichero y carpeta tal y como son aportados por los remitentes, sin la normalización de nomenclatura de la fase SIP.

La utilidad de este fichero es que si la extensión de los ficheros no es suficiente para identificar el formato y su versión (por ejemplo, la diferencia entre PDF y PDF/A o la versión exacta de PDF), estos datos lo permitan sin ambigüedad. La identificación deberá hacerse a través de alguno de los sistemas de registro de formato que la comunidad de expertos de la preservación digital ha ido implementando en los últimos años. Se recomienda la utilización del sistema de registro PRONOM debido a su amplio uso en la comunidad del patrimonio cultural[22]. Hay aplicaciones software libre, como DROID[23] que permiten analizar automáticamente los ficheros organizados en grupos para la detección del formato y la extracción de estos datos de identificación mediante un listado en formato CSV, por lo que este proceso se puede automatizar en su totalidad. Otra herramienta software libre más completa de este tipo, que también permite analizar la validez del fichero en cuanto a conformidad con su especificación técnica, es JHOVE[24].

Todos los ficheros en formato texto y con extensión TXT anteriores deberán llevar al comienzo una línea que indique el sistema de codificación de caracteres aplicado.

- Una **carpeta hija** denominada "metadatos_recibidos" que contendrá los ficheros de metadatos correspondientes al documento u obra a la que se dedica el paquete AIP por parte del remitente, en el formato estipulado por la normativa de ingreso de depósitos PreSIP. El motivo de introducir esta carpeta es por seguridad, en previsión de que durante las tareas de preservación digital o en la fase de ingesta pueda detectarse algún problema con los metadatos que requiera acudir a los ficheros tal y como los entregó el remitente, y no se haya conservado el depósito PreSIP. Estos ficheros de

[22] Se pude acceder al registro PRONOM desde http://apps.nationalarchives.gov.uk/PRONOM/Default.aspx

[23] Se puede descargar DROID desde http://www.nationalarchives.gov.uk/aboutapps/PRONOM/tools.htm

[24] Se puede descargar JHOVE desde http://www.nationalarchives.gov.uk/aboutapps/PRONOM/tools.htm

metadatos no se preservan digitalmente, por lo que no sufrirán procesos de migración ni otras estrategias de preservación digital. Esto es así porque todos los metadatos a preservar son los que se incluyen dentro del fichero METS del AIP, y el contenido de estos otros ficheros ya ha sido utilizado para crear los metadatos de ese fichero METS, estando pues solapados con ellos. Para evitar la pérdida de metadatos ante la probabilidad de que algunos ficheros tengan en mismo nombre y extensión se creará la estructura de carpetas necesaria para identificar cada formato de metadatos recibido, almacenándose cada tipo de formato en su carpeta correspondiente.

- Una **carpeta hija** denominada "objetos" que contendrá los ficheros del documento u obra empaquetada con la estructura que se indica en los siguientes párrafos, considerando que si no existe algún tipo de objeto no se creará la capeta correspondiente, pues no se permite la creación de carpetas vacías bajo ningún concepto. La carpeta objetos contiene todos los ficheros a preservar digitalmente del documento u obra a la que se dedica el paquete AIP, ningún fichero a preservar digitalmente puede estar fuera de esta carpeta o sus carpetas hijas salvo del fichero METS del paquete AIP que se crea durante la ingesta. El contenido de esta carpeta es:

    - Carpeta "masteres". Esta estructura se debe adecuar a la descrita en el apartado "Sistema de organización de los ficheros de originales y de control en los discos de almacenamiento" de las especificaciones de digitalización.

        - Ficheros másteres. Se respetará el nombre de los másteres aportados por el remitente pero ajustados a la normativa de nomenclatura de los AIP. Cada fichero máster llevará un código UUID a continuación del nombre del máster, separado de éste por un guion medio.

En la normativa de digitalización, en la parte dedicada a estructura de los discos de almacenamiento, se indica que en la carpeta donde se almacena el fichero o ficheros máster de una obra se incluya:

- Un fichero de texto plano con el mismo nombre que el del fichero máster pero finalizado con la cadena de caracteres "_hash" y codificado en UTF-8, que contendrá dos datos separados por un carácter retorno de carro: el código hash de fichero máster y el algoritmo usado para obtener ese código. Esta información podrá ser usada posteriormente para validar que el fichero máster no ha sufrido algún tipo de modificación desde la fecha de su almacenamiento definitivo en disco.
- Un fichero de texto plano con el mismo nombre que el del fichero máster pero finalizado con la cadena de caracteres "_mdcapt" y codificado en UTF-8, que contendrá tres datos separados por un carácter retorno de carro: el modelo de dispositivo usado para la captura expresado de la forma más completa posible pero sin usar ningún carácter de puntuación (ni coma, ni punto, ni punto y coma ni dos puntos o cualquier otro) para separar las palabras usadas; el nombre y versión de la aplicación o aplicaciones informáticas utilizadas para obtener, comprimir (en su caso) y dar formato de fichero al fichero máster, separadas éstas por coma, cuando haya dos o más (no usarán otros signos de puntuación entre las palabras de un nombre de aplicación, salvo el punto para expresar la versión exacta de ésta), aunque si ésta última coincide únicamente con el controlador del escáner o la aplicación usada para hacer el revelado RAW, se hará constar sólo el nombre y versión de esta aplicación; el nombre y apellidos de la persona que realizó su captura; la empresa o institución responsable del trabajo de digitalización. Esta información podrá ser usada posteriormente obtener con facilidad los datos de evento de captura que se requieren en preservación digital y que no siempre aparecen incrustados en las cabeceras de los ficheros máster.

Se respetarán estos ficheros en el repositorio de preservación e irán en esta carpeta junto a sus correspondientes ficheros másteres, debiéndosele añadir al nombre un código UUID. En el caso de que una obra tenga varios ficheros másteres (como puede

ocurrir en un documento multipágina) el nombre de estos ficheros de control será el de la carpeta padre del máster. Su extensión será “txt”.

- Carpeta “derivados”. Con las subcarpetas que se han utilizado en los discos de digitalización y de acuerdo a la norma de estructura aportada en las especificaciones de digitalización. En la carpeta derivados, o, en su caso, en sus subcarpetas habrá:

    - Ficheros derivados. Se respetará el nombre de los derivados aportados por el remitente pero ajustados a la normativa de nomenclatura para derivados de los SIP. Los derivados proporcionados por el propio remitente llevarán su propio código UUID. Recordemos que su nombre, a excepción del código UUID, debe ser igual al de los másteres a que corresponden.

No se admitirá que la carpeta derivados se denomine con el nombre de la extensión de los derivados, tal como JPG, JPEG, JPEGs, MP2, MP3, MP4…. ya que con el tiempo puede cambiar el formato en que se custodian este tipo de ficheros. Si el remitente denomina a la carpeta de derivados de estas formas se hará una correspondencia entre su carpeta a la carpeta “derivados” del AIP, pero sin admitirse el cambio de nombre en la última.

En la normativa de digitalización, en la parte dedicada a estructura de los discos de almacenamiento, se indica que en la carpeta donde se almacena el fichero o ficheros derivados de una obra se incluya el mismo sistema de control referido para la versión máster. Pero en lo que respecta al fichero “_mdcapt” se usarán los datos de la persona o personas que han preparado y ejecutado el procesado para obtener los derivados desde los másteres y la aplicación informática usada para ello, sin consignarse el dato del dispositivo de captura, ya que entendemos que los derivados se generan directamente desde los ficheros másteres.. Se seguirán manteniendo en el paquete de preservación estos ficheros con los mismos requisitos que los que se han definido para el caso de los másteres.

- Carpeta “raws”. En su caso.
    - Ficheros en formato RAW de cámaras digitales, normalizados de acuerdo al plan de preservación digital. Se respetará el nombre de los ficheros RAW aportados por el remitente pero ajustados a la normativa de nomenclatura para derivados de los SIP. Llevarán su propio código UUID. Recordemos que su nombre debe ser igual al de los másteres a que corresponden salvo que llevan su propio código UUID.
- Carpeta “miniaturas”. En su caso.
    - Ficheros de miniatura que se hayan decidido preservar digitalmente.
- Una carpeta para contener el fichero o ficheros de subtítulos, en su caso, si se trata de contenidos audiovisuales.
- Capeta “control”. En la normativa de digitalización se exige la creación de esta carpeta para contener todos los ficheros de control de calidad de las capturas. Se preservará dentro de esta carpeta el contenido de la carpeta control creada durante los trabajos de digitalización. Se deberá asignar a cada fichero de control un código UUID, ya que son ficheros a preservar digitalmente. El contenido de esta carpeta será el marcado por la normativa de digitalización. Como estos ficheros de control deben ser preservados, se les aplicará su correspondiente código UUID.

En ningún caso debe admitirse una carpeta vacía de contenido o que contenga una carpeta hija vacía de contenido.

Será requisito imprescindible que los nombres de los derivados, miniaturas, subtítulos y ficheros RAW sean exactamente iguales que los de los másteres a que corresponden, a excepción de la extensión, que será la adecuada para la identificación del formato de fichero y en minúsculas, y el código UUID. En el caso de que haya derivados diferentes que compartan el mismo nombre y extensión, salvo el código UUID, deberán presentarse en el SIP obligatoriamente en carpetas separadas y con nombre diferente, no pudiéndose mezclar en una misma carpeta.

- Fichero de metadatos en formato METS. Su contenido es explicado con detalle más adelante en el epígrafe *Metadatos del AIP, sus esquemas y formas de codificación admitidas*. Este fichero debe contener codificados en METS todos los

metadatos en toda la tipología exigida por la normativa. Sólo contendrá metadatos relativos al documento u obra a la que corresponde el paquete AIP y a los ficheros digitales que lo conforman y están, por tanto, almacenados en la carpeta "objetos". Su extensión será "xml" (no admitiéndose la extensión "mets" que puede verse usualmente en este tipo de ficheros), y su nombre coincidirá con el nombre de la carpeta padre del AIP con la cadena "mets-" justo delante. El nombre deberá incluir también obligatoriamente el código UUID de la carpeta padre. Este fichero METS sólo debe referir en sus secciones fileSec, structMap y de metadatos al contenido de la carpeta objetos, que es el contenido digital que debe ser preservado digitalmente en el tiempo. Evidentemente, también deben ser objeto de preservación digital el propio fichero METS y los ficheros adjuntos en formato TXT de control de BagIT y del sistema de preservación, pues éstos aportan a los objetos a preservar su contexto e información de proceso e integridad que puede ser requerida en cualquier momento dentro del flujo de trabajo de la preservación digital activa. La finalidad del fichero METS es: codificar los metadatos, enlazar los objetos a preservar con sus metadatos, y expresar las relaciones físicas y lógicas entre los objetos dentro del paquete AIP.

- Fichero bag-info.txt. Contendrá la información normativa del estándar BagIT para este fichero.
- Fichero bagit.txt. Contendrá la información normativa del estándar BagIT para este fichero.
- Fichero manifest-xxxx. Contendrá la información normativa del estándar BagIT para este fichero. Los x serán sustituidos por el identificador del método hash usado para obtener los códigos hash, usándose tantos caracteres como se necesite.
- Fichero tagmanifest-xxxx. Contendrá la información normativa del estándar BagIT para este fichero. Los x serán sustituidos por el por el identificador del método hash usado para obtener los códigos hash, usándose tantos caracteres como se necesite.

La normativa de preservación digital que hemos mencionado más arriba es muy rigurosa, pues exige que cualquier tipo de procesado aplicado a un AIP o a cualquiera de sus objetos de información pueda ser conocido a lo largo de todo el período de vida del repositorio y en cualquier momento, independientemente de los cambios de sistemas informáticos y de almacenamiento que haya podido sufrir.

Por ello, es imprescindible para el cumplimiento íntegro de los estándares de preservación que cualquier cambio realizado sobre los contenidos aportados por el remitente queden registrados como metadatos de preservación, por lo que los AIP deberán heredar como eventos las transformaciones que pueden recibir los PreSIP para convertirlos a SIP normativos. La única forma de poder automatizar esta herencia es que el sistema de gestión del repositorio registre todos los eventos que sufren los PreSIP para su conversión a AIP normativos en su sistema de gestión de datos.

Las normas de nomenclatura a seguir para el AIP son:

- Se respetará el nombre de carpeta de remitente siempre y cuando siga las normas mínimas asentadas en la normativa para SIP en el epígrafe respectivo más arriba. Si durante la fase de ingesta del SIP para su conversión al AIP se detectará una nomenclatura incorrecta, no se ingestará el paquete SIP, generándose una entrada al fichero log de errores que deberá revisar el encargado de la ingesta de los SIP al repositorio de los AIP.
- Se respetará el nombre original aportado por el remitente para cada fichero, siempre y cuando siga las normas mínimas asentadas en la normativa para depósitos SIP ya citada, en los mismos términos explicados en el párrafo anterior. Al nombre original del fichero máster o derivado aportado por el remitente, se le adjuntará el código UUID entre la finalización del nombre y la extensión, separado por un guion medio, tal como vemos en el siguiente ejemplo:

"006308.tif" pasará a tener en el AIP el nombre "006308-58fe8e40-c452-455a-a623-9aefdfbb3335.tif" en el hipotético caso de que "58fe8e40-c452-455a-a623-9aefdfbb3335" sea el código UUID generado para este paquete por el sistema de preservación digital.

9.4.3 Empaquetamiento de los ficheros de control a nivel de lote de captura o colección.

Este contenido deberá coordinarse con el de las especificaciones de organización de ficheros que aparecen en el epígrafe "Sistema de organización de los ficheros de originales y de control en los discos de almacenamiento" de las especificaciones de digitalización, en lo que atañe con los ficheros de control independientes que tengan contenidos aplicables a un lote de documentos. Todos los ficheros de control deben ser objeto de preservación digital, por lo que llevarán asignado su código UUID correspondiente.

Reproducimos justo a continuación esta normativa:

- Los ficheros de control independientes se ubicará en una carpeta hija de la carpeta padre de colección denominada "control_*NombreColeccion*", en la que *NombreColeccion* se sustituirá por el nombre de carpeta asignado a la colección. Por ejemplo, si al conjunto de fotografías de un fondo concreto

se le da el nombre de carpeta padre “fotografía” la carpeta de control tendrá como nombre “control_fotografia”. Los ficheros de control son las capturas de las cartas de control usadas para el control de calidad o la calibración y caracterización de los equipos y sus ficheros de referencia, así como otros ficheros que aporten datos sobre el control de calidad u otros procesos de control realizados a los lotes de documentos.

- Dentro de esta carpeta habrá tantas subcarpetas como capturas de control se hayan hecho para esa colección, cuyo nombre será “control_lote*N*”, siendo sustituida la “*N*” por el número de lote, que será asignado por el operador de dispositivo de captura para que nunca haya coincidencia de número entre un lote y otro. Los números de lote tendrán los mismos caracteres de extensión, usándose ceros delante hasta completar la extensión máxima necesaria para contener a todos los números. Un lote contiene las capturas de control de todos los ficheros que se han capturado con condiciones idénticas de captura y almacenamiento. En esta carpeta se almacenarán los ficheros de las capturas de control de ese lote. Además se deberán guardar, por cada carta de control usada para la que corresponda, un fichero TXT, con codificación UTF-8, con los datos de referencia de cada parche de densidad o de color de la carta. Si los datos se expresan en sistema de color CIELAB o CIE XYZ, se referirá el iluminante utilizado y el observador estándar. Cuando se trate de cartas que sirvan para la creación de un perfil ICC, como la carta colorchecker® o IT8, se guardará también el fichero de referencia de formato CGATS. Los ficheros de texto y CGATS serán guardados con codificación UTF-8. Los nombres de estos ficheros identificarán el modelo de carta al que refieren de la forma más precisa posible.
- Dentro de la carpeta correspondiente a cada lote se almacenará un fichero de texto con extensión TXT y codificación UTF-8 que contendrá el listado de ficheros máster de ese lote identificados por su nombre de archivo con extensión de archivo incluida. Su nombre será el de la carpeta donde se ubica más la cadena de caracteres “_listfich”.
- Dentro de la carpeta correspondiente a cada lote se almacenará un fichero en formato PDF/A que contendrá un informe de calidad que contemple la aplicación y el resultado de los parámetros de control de calidad establecidos en el epígrafe de este trabajo denominado como *Control de aseguramiento de calidad previo a la captura*. Las características que debe cumplir este informe aparecen descritas en el subapartado del anterior epígrafe denominado *Registro de almacenamiento de metadatos y otra documentación sobre control de calidad*. El nombre de este fichero será el de la carpeta padre del lote donde se ubica seguido por la cadena de caracteres “inf_calidad”. Su extensión será “.pdf”.

- Si se opta por guardar los ficheros de datos resultados de las pruebas de calidad que suelen dar como salida las aplicaciones de control de calidad, estos ficheros serán guardados dentro de la carpeta padre del lote y renombrados para poder cumplir la normativa de nomenclatura de estas especificaciones, incluyendo en el nombre el nombre de la carpeta padre correspondiente al lote.
- El nombre de los ficheros de las cartas de control será el indicado en su epígrafe correspondiente más abajo, ya que puede variar según el tipo de documento.

En el caso de los vídeos se exige:

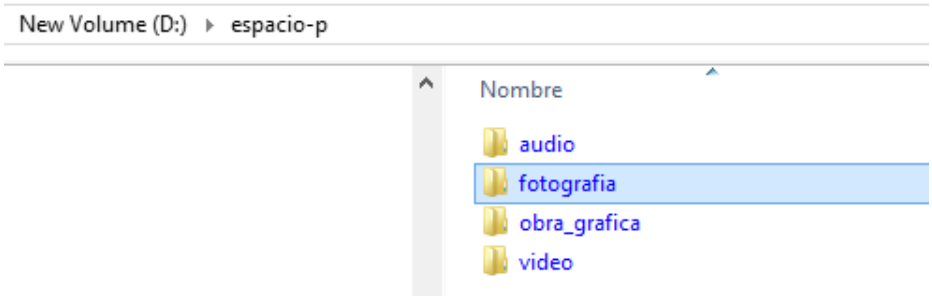
- Los ficheros de control independientes, si los hay, se ubicarán en una carpeta hija de la carpeta padre de colección denominada "control".
- Dentro de esta carpeta habrá tantas subcarpetas como capturas de control se hayan hecho para esa colección, cuyo nombre será "control_lote*N*", siendo sustituida la "*N*" por el número de lote. Un lote contiene las capturas de control de todos los ficheros que se han capturado con condiciones idénticas de captura y almacenamiento.
- Dentro de la carpeta correspondiente a cada lote se almacenará un fichero de texto con extensión TXT y codificación UTF-8 que contendrá el listado de ficheros máster de ese lote identificados por su nombre de archivo con extensión de archivo incluida.
- Se almacenará también en esta carpeta, cuando sea de aplicación, un fichero en formato PDF, que contendrá un Informe de procesamiento de la señal digital, que tendrá asimismo el nombre del fichero máster finalizado con la cadena "_proces". Su extensión será la correspondiente al formato PDF. El contenido de este informe se describe en el apartado dedicado a Informe de procesamiento de la señal digital aplicado a las capturas.

La carpeta "control" será tratada, además, como un paquete Bag-it, por lo que tendrá todas sus subcarpetas de control dentro de una carpeta padre denominada como "data", y los ficheros de texto correspondientes de Bag-it.

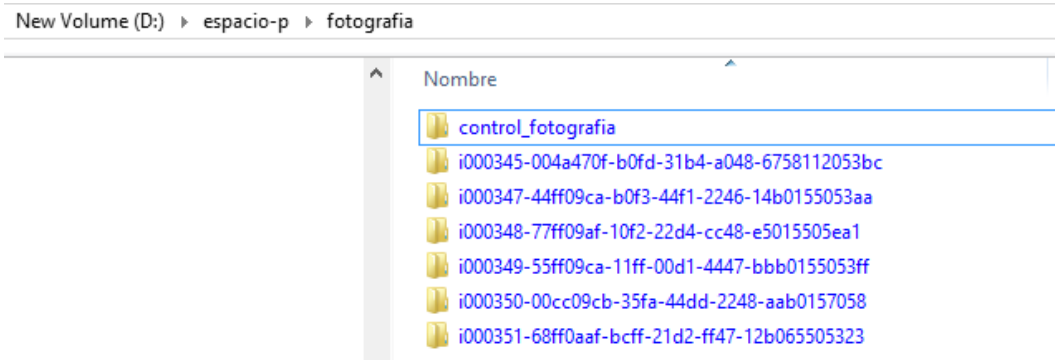
9.4.4 Ejemplo de estructura organizativa del depósito de preservación.

A continuación, ejemplificamos cómo quedaría en el sistema de almacenamiento la estructura organizativa del depósito de preservación.

En un primer nivel tendríamos la carpeta padre de un fondo concreto a preservar, que en este caso es el fondo de Espacio-P.

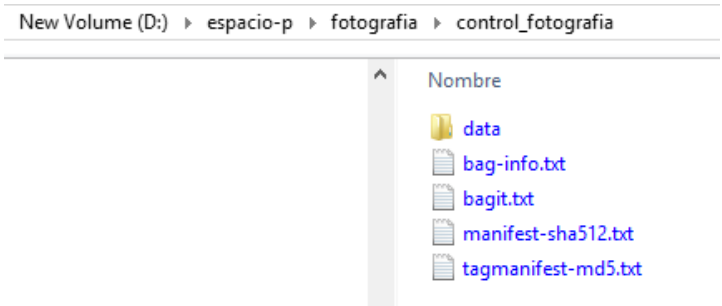


Pensemos que en este fondo tenemos cuatro colecciones: audio, vídeo, obra gráfica y fotografía. Si vamos a la colección de fotografía podríamos encontrar una carpeta padre por cada obra a preservar, más la carpeta control.

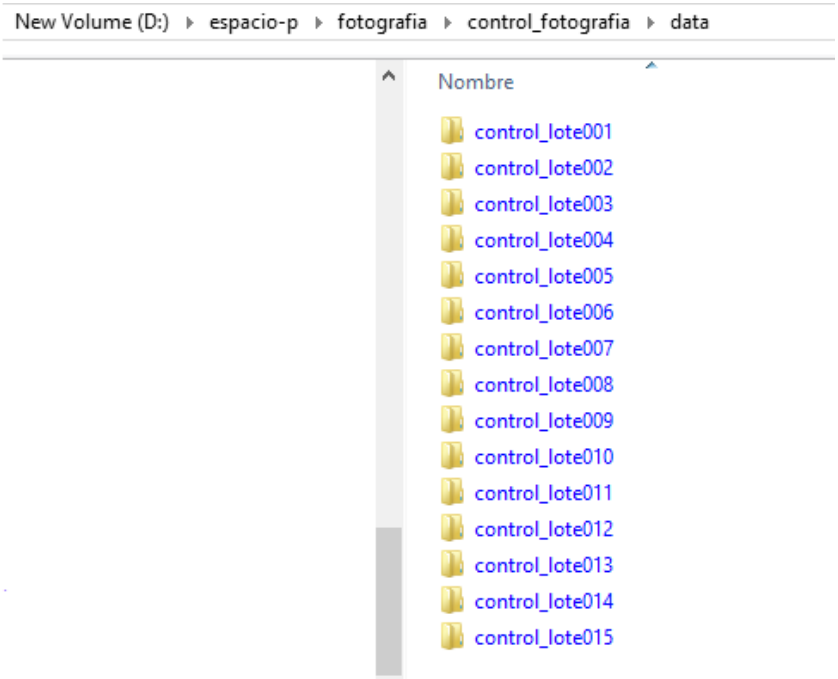


Las carpetas padre tienen en su nombre, un identificador de obra, que sería su número de inventario más el correspondiente código UUID.

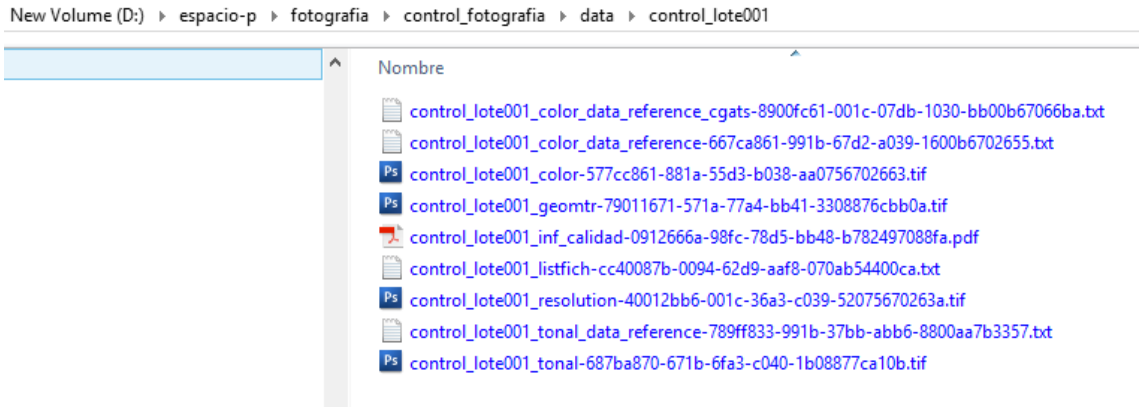
Si accedemos a la carpeta control_fotografía, veremos los ficheros preceptivos de Bag-it, pues esta carpeta es un paquete Bag-it.



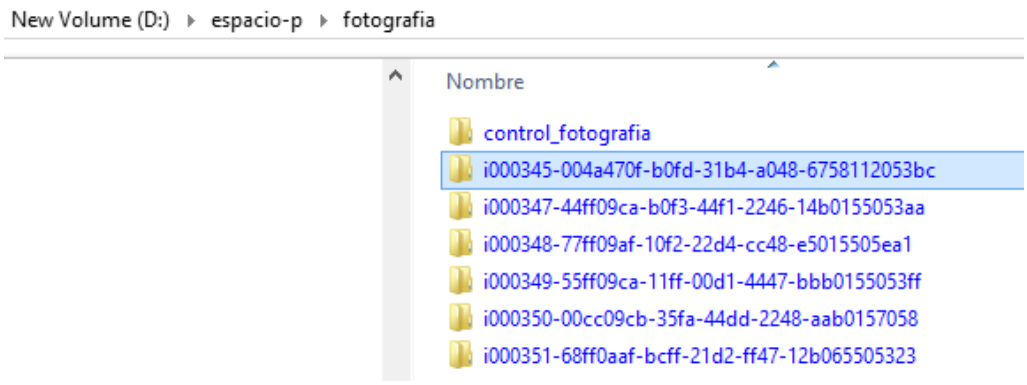
Si accedemos a la carpeta data, encontraremos las carpetas de cada lote de control de la colección de fotografía. Pensemos que hay 15 lotes de control.

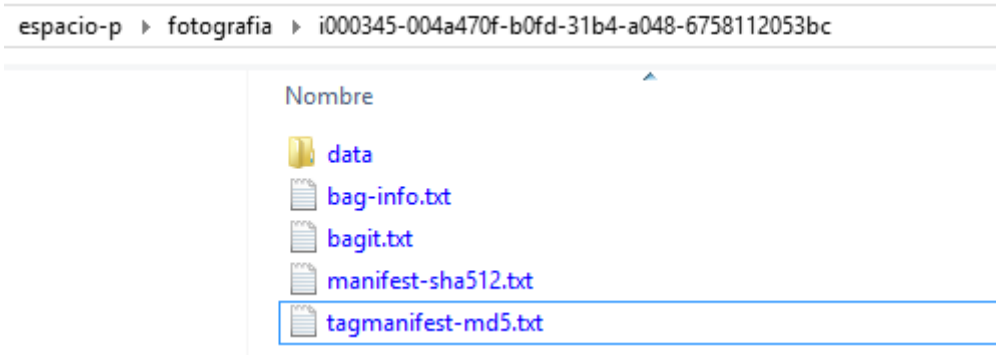


Si vamos a una carpeta de control de lote concreta, veremos los ficheros de control, de acuerdo a la normativa explicada más arriba. Todos los ficheros llevan su UUID, pues deben ser preservados digitalmente.



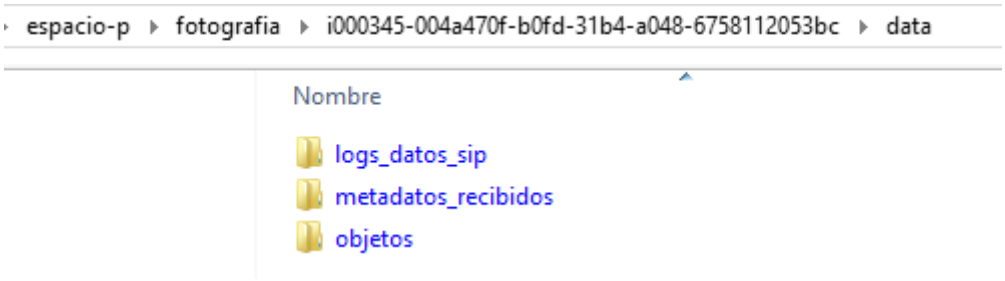
Si ahora vamos a la carpeta de una obra concreta, por ejemplo la i000345-004a470f-b0fd-31b4-a048-6758112053bc, veremos su contenido.



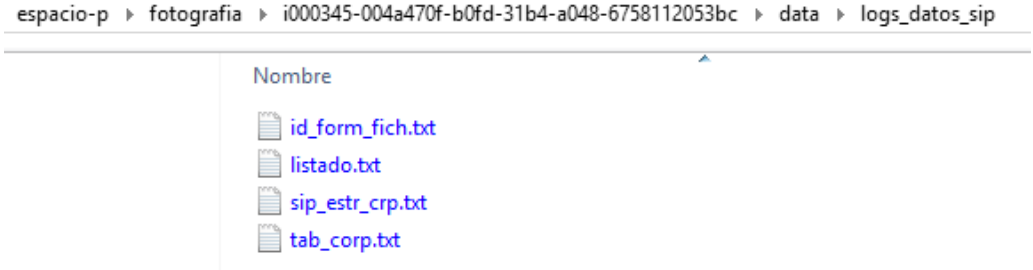


Estamos también en un paquete Bag-it, que tiene la carpeta data, con todo el contenido a preservar, y los ficheros propios de este estándar de la LC.

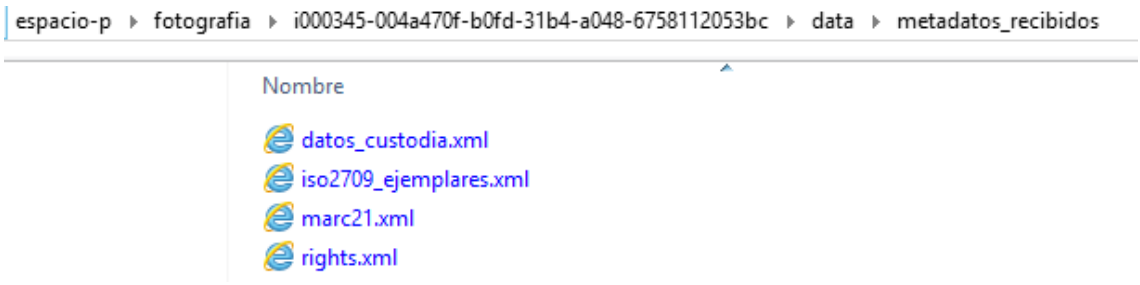
Si accedemos a la carpeta data, veremos su contenido.



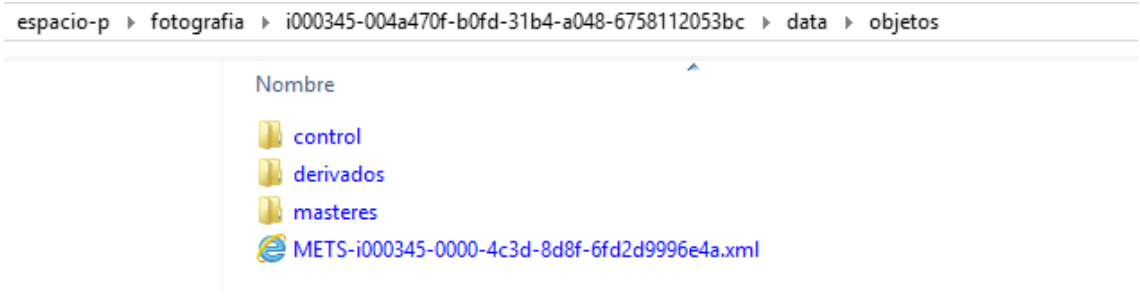
Si vamos a logs_datos_sip, vemos los ficheros de control tipo log.



Si vamos a metadatos_recibidos, veremos los metadatos con los que ha ingresado la obra digitalizada al depósito de preservación.



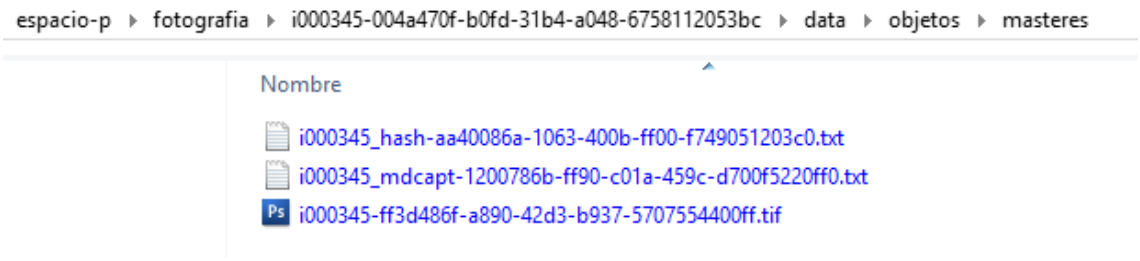
Si vamos a objetos, veremos las carpetas que estructuran los ficheros digitales de esta obra.



Si accedemos a la carpeta control, veremos ficheros de control específicos de la digitalización concreta de esta obra. En este ejemplo, contamos con estos ficheros, pero puede haber obras que no tengan ficheros de control específicos.

Si vamos a derivados.

Si vamos a másteres.



En todos los ficheros de este ejemplo está controlada la longitud máxima de ruta, de manera que no se sobrepase. La longitud de ruta más larga tendría 138 caracteres.

9.5 Creación o mantenimiento de identificadores persistentes y únicos para los contenidos a preservar en los AIP.

La exigencia de los estándares de preservación digital obliga que el sistema de preservación obtenga, o genere, y aplique a todos los paquetes AIP un identificador único, así como a los objetos digitales a preservar que integra. Por ello se exige en esta normativa la asignación de identificadores de acuerdo a sistema UUID para los másteres y derivados en su diferente tipología proporcionados por el remitente con la finalidad de su preservación digital.

Es más, la normativa TRAC obliga a que se documente cualquier cambio en el sistema de identificadores únicos que se aplique durante toda la vida del repositorio. Los cambios de este tipo deberán quedar reflejados en los metadatos PREMIS que se integran en los ficheros METS de los AIP y también en los ficheros que establecen las tablas de correspondencia de los nombres de fichero de los paquetes AIP antes de la migración y los de después de la migración.

Si los objetos publicados en la Web tienen asignados identificadores únicos tales como PURL, Handle o similares, se deberán registrar en los metadatos de manera que nunca pueda perderse su vinculación con los objetos a que corresponden.

9.6 Metadatos del AIP, sus esquemas y formas de codificación admitidas.

Los metadatos se registrarán en el fichero METS indicado en la normativa de empaquetamiento. Ese fichero debe ser un fichero XML acorde con el estándar de la Library of Congress METS (*Metadata Encoding and Transmission Standard*). Se usará la última versión de este estándar, obtenible en: http://www.loc.gov/standards/mets/ . Se considerará asimismo el documento que a modo de guía establece los requisitos para la compatibilidad de los metadatos PREMIS con el modelo METS, que puede ser accedido en:

> *Using PREMIS with METS*. Disponible en: http://www.loc.gov/standards/premis/premis-mets.html

Más abajo dedicamos un epígrafe completo a la descripción del perfil METS que debe ser aplicado para la generación del fichero METS de cada AIP.

Recordamos que sólo se generarán metadatos administrativos o de cualquier tipo dentro del fichero METS del AIP para los objetos digitales a preservar, que son únicamente los contenidos dentro de la carpeta "objetos", hija de la carpeta "data". Y que en las secciones fileSec y structMap sólo pueden referirse estos objetos.

En los epígrafes siguientes describimos los tipos de metadatos reconocidos por METS cuya consignación es obligatoria de acuerdo a la terminología METS.

9.6.1 Metadatos descriptivos (bibliográficos) (dmdSec METS).

Estos metadatos describen aspectos de autoría, formales y de contenido de los documentos u obras a los que corresponden los objetos digitales resultado de la digitalización o de las unidades intelectuales nacidas digitales.

Será obligatoria la presencia de metadatos descriptivos para cada documento u obra, y que se aporten en alguno de los esquemas de metadatos y sistemas de codificación admitidos para el sistema de preservación. Estos metadatos descriptivos se insertarán en una o varias secciones "dmdSec METS" siguiendo la normativa METS. En ningún caso se podrá insertar una sección de este tipo de metadatos en un sistema de representación considerado obsoleto por la comunidad de expertos en preservación digital. Este aspecto deberá estar controlado en el plan general de preservación digital y en los planes específicos que correspondan.

Se usará como sistema de representación de los metadatos bibliográficos el formato Dublin Core (DC) cualificado. Estos metadatos se obtendrán desde las bases de datos bibliográficas que se usen para la búsqueda y gestión de los fondos. La codificación XML obedecerá al plan de mapeo entre los campos de las bases de datos y los esquemas DC y DCTERMS (DC cualificado) que deberá ser establecido previamente.

9.6.2 Metadatos administrativos (amdSec METS).

Incluyen una amplia tipología de metadatos:

- Técnicos (techMD METS). Describen parámetros técnicos de los ficheros digitales. Se admite cualquier esquema de metadatos técnicos de uso común o propios de los formatos de fichero en que ingresen los objetos digitales, tales como NISO Technical Metadata en su sistema de representación XML con el esquema MIX[25], Exif, Metadatos de cabecera TIFF, XMP, PBCore, etc. Pero será obligatorio obtener una representación de metadatos técnicos en XML de acuerdo al modelo PREMIS en la Entidad Object. Se representarán todas las unidades semánticas PREMIS Object de las que se disponga información técnica y se pueda extraer automáticamente su contenido, garantizando que estén todas las unidades que son obligatorias en PREMIS cuando corresponde su aplicación al tipo de objeto procesado o son requeridas para el modelo de preservación digital descrito, que son: 1.1 objectIdentifier (1.1.1 objectIdentifierType, 1.1.2 objectIdentifierValue), 1.2 objectCategory, 1.5 objectCharacteristics (1.5.1 compositionLevel, 1.5.2 fixity, 1.5.3 size, 1.5.4 format [1.5.4.1 formatDesignation, 1.5.4.2 formatRegistry]), 1.6 originalName, 1.9 signatureInformation, 1.10 relationship (1.10.1 relationshipType, relationshipSubType , 1.10.2 relationshipSubType, 1.10.3 relatedObjectIdentification, 1.10.4 relatedEventIdentification).

- Procedencia digital (digiprovMD METS). Contiene metadatos sobre la procedencia digital (información sobre la relación entre el documento original y su representación digital, incluyendo la relación entre copias maestras y derivadas, migraciones y transformaciones realizadas sobre los archivos desde su digitalización inicial). Incluye las entidades PREMIS Evento y Agente y sus respectivas unidades semánticas. Será obligatorio representar las siguientes unidades en XML de acuerdo al modelo PREMIS.
    - De evento: 2.1 eventIdentifier (2.1.1 eventIdentifierType , 2.1.2 eventIdentifierValue), 2.2 eventType, 2.3 eventDateTime, 2.4 eventDetail, 2.5 eventOutcomeInformation (2.5.1 eventOutcome, 2.5.2 eventOutcomeDetail [2.5.2.1 eventOutcomeDetailNote]), 2.6 linkingAgentIdentifier (2.6.1 linkingAgentIdentifierType, 2.6.2 linkingAgentIdentifierValue).

[25] NISO Technical Metadata for Digital Still Images Standards in XML. Disponible en http://www.loc.gov/standards/mix/

- De agente: 3.1 agentIdentifier (3.1.1 agentIdentifierType, 3.1.2 agentIdentifierValue), 3.2 agentName, 3.3 agentType).

- Derechos de propiedad intelectual (rightsMD METS). Derechos y permisos de uso o transformación de objetos digitales. Contienen las autorizaciones y límites de uso por derecho de imagen y cláusulas de confidencialidad. Incluye la entidad PREMIS Derechos y sus unidades semánticas. La forma de la declaración de estos metadatos se ajusta a una plantilla que se realice por los responsables del proyecto de acuerdo al esquema METS-Rights, tal como la que aparece a continuación, sin datos o con unos valores por defecto.

```
<rts:RightsDeclarationMD xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:rts="http://cosimo.stanford.edu/sdr/metsrights/" xsi:schemaLocation="http://cosimo.stanford.edu/sdr/metsrights/ http://www.loc.gov/standards/rights/METSRights.xsd" RIGHTSCATEGORY="COPYRIGHTED">
<rts:RightsDeclaration>.</rts:RightsDeclaration>
  <rts:RightsHolder RIGHTSHOLDERID="HOLDER001">
<rts:RightsHolderName> </rts:RightsHolderName>
<rts:RightsHolderContact>
<rts:RightsHolderContactAddress> </rts:RightsHolderContactAddress>
<rts:RightsHolderContactPhone PHONETYPE="BUSINESS"></rts:RightsHolderContactPhone>
<rts:RightsHolderContactEmail> </rts:RightsHolderContactEmail>
</rts:RightsHolderContact>
</rts:RightsHolder>
<rts:Context CONTEXTCLASS="GENERAL PUBLIC">
<rts:Permissions DISCOVER="true" DISPLAY="true" COPY="true" DUPLICATE="false" MODIFY="false" DELETE="false" PRINT="true" OTHER="false"/>
</rts:Context>
<rts:Context CONTEXTCLASS="REPOSITORY MGR">
<rts:Permissions DISCOVER="true" DISPLAY="true" COPY="true" DUPLICATE="true" MODIFY="true" DELETE="true" PRINT="true" OTHER="false"/>
</rts:Context>
</rts:RightsDeclarationMD>
```

El idioma de los valores predeterminados de los metadatos Premis es el inglés. Por ejemplo, el nombre de evento para la captura digital en lugar de “captura” será “capture”.

Los metadatos de preservación digital se reparten entre los técnicos y los de procedencia digital.

9.6.3 Metadatos estructurales (fileSec y structMap METS) .

Permiten vincular física y lógicamente los ficheros que conforman parte de un mismo objeto digital. Se seguirá a rajatabla el estándar METS, siendo obligatorio incluir las secciones fileSec y structMap, de manera que todos los objetos incluidos en la carpeta “objetos” del paquete AIP queden vinculados correctamente y en la secuencia que les corresponde en el documento original.

9.6.4 Codificación de la estructura lógica del fondo de las instituciones o archivos que depositen contenidos en el sistema de preservación del AEMA en los metadatos METS.

En los metadatos de cada AIP se incluyen los datos de la estructura organizativa de la que depende el documento correspondiente al AIP. Allí consta la estructura jerárquica completa de acuerdo al sistema bibliotecario de la organización cliente.

Para la simplificación de la representación codificada de este tipo de datos se ha realizado un esquema XML propio, muy sencillo, denominado “datos_custodia” de tal manera que estos datos se representen en lenguaje XML empaquetados dentro del fichero METS que describe el contenido del paquete AIP. Se incluirán como una instancia nueva de metadatos descriptivos, esto es, como una nueva sección “mets:dmdsec”. Los elementos de este esquema son:

Para la simplificación de la representación codificada de este tipo de datos se ha realizado un esquema XML propio, muy sencillo, denominado “datos_custodia” de tal manera que estos datos se representen en lenguaje XML empaquetados dentro del fichero METS que describe el contenido del paquete AIP. Se incluirán como una instancia nueva de metadatos descriptivos, esto es, como una nueva sección “mets:dmdsec”. Los elementos de este esquema son:

<datos_custodia>

<institucion valor=”” /> //El atributo valor tiene el nombre de la institución que alberga la biblioteca o archivo.

<unidad nivel=”n” valor=”” /> // Puede llevar como valor del atributo nivel del 1 al infinito para indicar el nivel jerárquico de las unidades departamentales de las que dependen las bibliotecas. Se repiten tantos elementos <institucion> como sean necesarios según los niveles que haya en la descripción. El atributo valor tiene el nombre de la unidad.

<fondo nivel=”n” valor=””/> // Idem al elemento unidad pero para la entidad fondo y sus sub niveles de localización.

<codigoBiblioteca valor=" "/> // Contiene el código de institución que ha digitalizado el documento

<digitalizacion valor=""/> // El atributo valor representa el año de la digitalización.

<soporte valor=""/> // El atributo valor representa los códigos de soportes de almacenamiento digital (discos duros externos, DVDs, Blu-ray u otros que siga manteniendo activos la institución depositaria y quiera referenciar en el sistema de preservación digital) separados por comas.

</datos_custodia>

En ejemplo de cómo quedan estos datos para un documento concreto es:

```
<datos_custodia:datos_custodia xmlns="http://galan.uc3m.es/~jroble/datos_custodia/”
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="
http://galan.uc3m.es/~jroble/datos_custodia/
http://galan.uc3m.es/~jroble/datos_custodia/datos_custodia.xsd" >

        <institucion valor="Archivo Histórico Provincial de  Santa Cruz de Tenerife"/>
        <fondo valor="Archivos Privados" nivel="1"/>
        <fondo valor=" Archivos Personales y Familiares " nivel="2"/>
        <fondo valor=" Pedro Garhel" nivel="3"/>
        <codigoBiblioteca valor="Universidad Carlos III de Madrid"/>
        <digitalizacion valor="2018"/>
        <soporte valor="i003450 _2018 "/>
</datos_custodia:datos_custodia>
```

9.7 Fichero de comprobación general de los contenidos del depósito de preservación.

Se deberá crear automáticamente un fichero con el nombre *check_aip.txt* que contendrá la información de todos los paquetes AIP que conforman los contenidos a preservar de todas las colecciones. Este fichero permitirá garantizar que la estructura completa del depósito de preservación pueda ser verificada por cualquier aplicación externa. Este fichero contendrá una línea por cada paquete del depósito con la estructura:

Código-Hash Dirección del Paquete BagIt

El código hash será el correspondiente al archivo manifest-md5.txt de cada paquete, que es el archivo que contiene los códigos de integridad de cada fichero del paquete. Para obtenerlo se leerá desde el correspondiente archivo tagmanifest-md5.txt.

9.8 Perfil METS para todos los tipos de obras.

9.8.1 Objetivos y motivación.

En este epígrafe se describen las características principales del sistema de codificación de los ficheros METS que se usan en el sistema de empaquetamiento del AIP, detallándose la forma de uso de los atributos y elementos METS que deben ser utilizados para la generación de estos ficheros, así como los elementos o atributos opcionales cuyo uso no se permite. Se han considerado pautas de codificación comunes para todos los tipos de documentos u obras y también las específicas para cada tipo de medio.

El esquema METS asienta un formato de representación muy flexible para poder representar toda la tipología de metadatos necesarios para la perfecta documentación y contextualización de los objetos digitales. De esta flexibilidad deriva en que la misma información se pueda codificar de múltiples formas, e incluso en que se puedan aplicar diferentes niveles de precisión en su representación. Se propicia la misma flexibilidad para la representación de los valores de muchos de sus atributos, no habiendo vocabularios normativos que los cubran en su totalidad, y ni siquiera un idioma de representación común. Muchos de los elementos y atributos de METS son además opcionales, y los mismos atributos pueden ser aplicados a diferentes elementos con la misma función, debiendo el implementador de este estándar decidir cómo se aplican éstos.

Esta ventaja se convierte en una desventaja a la hora de crear codificaciones consistentes en el seno de una misma organización, que no varíen su formato de representación para los mismos casos de unos objetos digitales del mismo tipo a otros. Para conseguir la consistencia necesaria es preciso la creación de perfiles

METS, o el uso de perfiles ya creados por otras organizaciones y hechos públicos a través de la Web de la Library of Congress[26] u otros medios.

Aunque empleamos la denominación de perfil, no se trata de un perfil METS elaborado de acuerdo a la normativa para la creación de perfiles METS de la Library of Congress[27], pues no es nuestra intención crear un perfil público, sino asentar y describir la manera en que se van a codificar estos ficheros METS, con la doble finalidad de conseguir consistencia en las codificaciones que se hagan en el sistema de preservación, y de informar a los remitentes de la sistemática seguida para este procedimiento. No se va a aplicar ningún perfil METS ya creado y publicado, porque la misión del METS de empaquetamiento AIP no es la difusión ni el intercambio de datos sino la de representar toda la información que se necesita para la preservación digital y las relaciones entre los metadatos y los objetos digitales del paquete AIP.

Por tanto, se ha hecho una aplicación de METS específica y no ajustada a requerimientos de otras organizaciones, sino a los propios de la gestión de la preservación digital del sistema de preservación. En base a las necesidades de este servicio se ha elaborado un perfil METS personalizado, cuyas directrices son las que describimos a continuación.

9.8.2 Especificaciones comunes a todos los tipos de medios, documentos y obras.

9.8.2.1 Codificación de caracteres.

Cada fichero METS XML debe seguir el Sistema de codificación de caracteres de Unicode UTF-8, por lo que el documento XML debe comenzar con la siguiente declaración del estándar XML:

<?xml version="1.0" encoding="UTF-8"?>

9.8.2.2 Valores de fecha.

A no ser que se especifique lo contrario (o quede delimitado por el esquema XML que se aplica a este dato), todos los valores de fecha deben ajustarse al formato de W3C-DTF, llegando a especificar al menos hasta el día, de la forma AAAA-MM-DD. Se admitirán datos de fecha más específicos, tales como AAAA-MM-DDTHH:MM:SS, pero no fechas incompletas, como AAAA-MM o AAAA. La representación de un indicador de zona horaria es opcional.

---

[26] Accesibles desde la página *Registered Profiles*, http://www.loc.gov/standards/mets/mets-registered-profiles.html

[27] Accesible en la página METS Profiles, http://www.loc.gov/standards/mets/mets-profiles.html

9.8.2.3 Expresión de rutas de fichero en atributos de tipo dirección.

La expresión de todas las rutas de fichero debe ser relativa a la localización del propio documento METS. Esta norma se aplicará a todos los casos de expresión de rutas, incluyendo el atributo xlink:href. Esta práctica permitirá que las rutas de fichero sean independientes de las unidades y carpetas contenedoras de los AIP en los dispositivos de almacenamiento usados, sin que pierdan su sentido cuando se procede a traspasar los AIP a otros dispositivos o configuraciones de almacenamiento donde pueden cambiar las denominaciones de las unidades o carpetas contenedoras, o su estructura.

9.8.2.4 El elemento raíz (mets) y la referencia a esquemas XML en otros elementos padre.

El elemento raíz se codificará de la siguiente manera, incorporando necesariamente las declaraciones de xmlns de los esquemas XML XSI, XLINK Y METS, de la forma que muestra el siguiente ejemplo:

<mets:mets xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/version111/mets.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:mets="http://www.loc.gov/METS/">

No se permitirá que el atributo xsi:schemaLocation refiera a la URL del esquema que representa la versión actual del METS, tal como http://www.loc.gov/standards/mets/mets.xsd, pues de esta forma no queda registrada la versión del esquema que se está usando, sino la versión vigente de cada momento, que no tiene que coincidir necesariamente con la versión a que corresponde el fichero METS. En su lugar se usará la URL del esquema concreto que se esté usando para la codificación del fichero METS, que en el caso de la versión 11.1 (desde mayo de 2015) es http://www.loc.gov/standards/mets/version111/mets.xsd

No se usarán más atributos de este elemento. Llamamos la atención sobre el hecho de que no debe usarse el atributo profile, pues este perfil METS no es público, por lo que no aparece publicado en el registro de perfiles METS del sitio Web de METS de la Library of Congress.

Todos los elementos del documento METS deben llevar el prefijo correspondiente a su espacio de nombres, incluyendo los elementos del propio METS, por lo que el esquema de METS se referirá tal como aparece en el ejemplo de arriba: xmlns:mets="http://www.loc.gov/METS/", sin admitir espacios de nombres por defecto, salvo para los metadatos incrustados en la sección dmdSec, si así se decide por los responsables del sistema de preservación en acuerdo con el remitente.

Los esquemas del resto de elementos, como pueden ser los de PREMIS o MIX, MARC, MODS, DC, etc. se referirán en su etiqueta padre correspondiente, tal y como vemos en el siguiente ejemplo para los metadatos PREMIS:

<premis:object xsi:schemaLocation="info:lc/xmlns/premis-v2 http://www.loc.gov/standards/premis/v2/premis-v2-2.xsd" xsi:type="premis:file" version="2.2" xmlns:premis="info:lc/xmlns/premis-v2">

En el siguiente ejemplo vemos una declaración de metadatos MARC, a la que se aplica el espacio de nombres por defecto del esquema MARC21 XML:

<mets:dmdSec ID="dmdSec-UUID-fijo-prueba">

<mets:mdWrap MDTYPE="MARC">

<mets:xmlData>

<collection xsi:schemaLocation="http://www.loc.gov/MARC21/slim http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd" xmlns="http://www.loc.gov/MARC21/slim">

<record>

<leader>00000cam 82200000 i 4500</leader>

<controlfield tag="001">ES-MAAEC20130000824</controlfield>

<controlfield tag="003">ES-MAAEC</controlfield>

<controlfield tag="005">20140116101816.0</controlfield>

<controlfield tag="008">921125s1607 mex 000 0 nah c</controlfield>

9.8.2.5 Cabecera (metsHdr).

El elemento de cabecera se utiliza para registrar información sobre el propio documento METS. Se hará uso del atributo CREATEDATE con el formato de fecha reflejado en el siguiente ejemplo:

<mets:metsHdr CREATEDATE="2015-03-20T14:46:58">.

No se usarán más atributos para este elemento.

Dentro de la cabecera se incluirán los elementos que permiten registrar los datos de creador y editor del fichero METS, tal y como vemos en el siguiente ejemplo:

<mets:agent TYPE="ORGANIZATION" ROLE="CREATOR”>

<mets:name>Universidad Carlos III de Madrid</mets:name>

</mets:agent>

<mets:agent TYPE="ORGANIZATION" ROLE="EDITOR”>

<mets:name> Empresa H</mets:name>

</mets:agent>

En el caso de que la institución remitente de contenido a preservar desee incluir más metadatos de cabecera, deberá comunicarlo a los responsables del AEMA, proporcionando al mismo tiempo los datos necesarios para su codificación.

9.8.2.6 Sección Descriptiva (dmdSEC).

En esta sección se incluirán, de forma interna (sin referencias a ficheros externos), los metadatos descriptivos del documento u obra a la que corresponde el paquete AIP (número de publicación periódica, libro, documento manuscrito, fotografía…), en los formatos y esquemas admitidos en la normativa de repositorio y, en todo caso, codificados en XML. Recordamos que no se admitirá la preservación digital de objetos para los que no se disponga de un conjunto de metadatos descriptivos que permitan conocer al menos sus características bibliográficas básicas.

Sólo se hará uso del atributo requerido ID en el elemento padre y se procurará no usar espacios de nombres por defecto, tal y como vemos en el siguiente ejemplo.

<mets:dmdSec ID="DM1">

<mets:mdWrap MDTYPE="MARC">

<mets:xmlData>

<marc:collection xsi:schemaLocation=http://www.loc.gov/MARC21/slim http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmln:marc="http://www.loc.gov/MARC21/slim">

<marc:record>

<marc:leader>00000caa 822000004b 4500</leader>

[…]

En el caso en que haya varias secciones de metadatos descriptivos aplicables al documento u obra se usará el atributo “DMDID” de cada uno de los elementos FILE para contener los valores de los atributos ID de esas secciones. De acuerdo al estándar METS los valores se deberán separar por un espacio en blanco y dentro de un mismo juego de comillas, como vemos en el siguiente ejemplo:

<mets:fileSec>

<mets:fileGrp ID="fileGrp-UUID-fijo-prueba">

<mets:file ID="_00d00001-0001-46cc-826a-bf2b95a0ff0c" MIMETYPE="application/xml" GROUPID="fileGrp-UUID-fijo-prueba" **DMDID="dmd-1 dmd-2"** ADMID="amd001 amd005">

<mets:FLocat xlink:type="simple" xlink:href=" objetos\derivados\PDF\optimizado\fc_012772-00d00001-0005-469c-a978-ec5089a46b5d.pdf" OTHERLOCTYPE="SYSTEM" LOCTYPE="OTHER"/>

</mets:file>

Además se referirán los metadatos descriptivos dentro de la sección o secciones StructMap a través del elemento <DIV> padre de esa sección, también mediante el atributo DMDID. Está redundancia se mantiene por considerarla positiva de cara a la preservación digital futura de los objetos digitales referenciados en el fichero METS, pudiéndose eliminar de acuerdo con el criterio del remitente.

Los metadatos descriptivos de carácter bibliográfico a incluir en el METS del AIP serán siempre los metadatos aportados por el sistema de gestión museístico en formato Dublin Core (DC) en XML. Estos metadatos se obtendrán automáticamente del sistema usado para la catalogación y gestión de las obras.

En el siguiente ejemplo (ficticio) vemos cómo quedaría una declaración de metadatos en DC dentro de una sección dmdSec:

<mets:dmdSec ID="dublin_core-00203fb4-0000-442e-aa70-e7369f31bdd2">

<mets:mdWrap MDTYPE="DC">

<mets:xmlData>

<metadata xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/" xsi:schemaLocation="http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2008/02/11/dc.xsd http://purl.org/dc/terms/ http://dublincore.org/schemas/xmls/qdc/dcterms.xsd">

<dc:source>Archivo Espacio P</dc:source>

<dc:type>Fotografía</dc:type>

<dc:title>Yack el Performador</dc:title>

<dc:creator>Fernando Suárez Cabeza</dc:creator>

<dcterms:medium>Papel fotográfico en blanco y negro</dcterms:medium>

<dcterms:extent>Positivo</dcterms:extent>

<dcterms:extent>27x36 cm</dcterms:extent>

<dcterms:spatial>España</dcterms:spatial>

<dcterms:spatial>Madrid</dcterms:spatial>

<dcterms:spatial>Madrid</dcterms:spatial>

<dc:description>Fundido a negro sobre fondo gris en muro lateral</dc:description>

<dc:subject>Fotografías </dc:subject>

<dcterms:isPartOf>Fondo Archivo Fotográfico Espacio P</dcterms:isPartOf>

<dcterms:provenance>Colección F</dcterms:provenance>

<dc:language xsi:type="dcterms:ISO639-3">spa</dc:language>

<dc:rights>©Fondo Archivo Fotográfico Espacio P. No duplicar.</dc:rights>

<dc:publisher>UC3M, UCLM</dc:publisher>

<dc:identifier xsi:type="dcterms:URI">http://www.uc3m.es/archivo_ma/show_ficha.do?archivo=AMA&record=ANA-12201</dc:identifier>

</metadata>

</mets:xmlData>

</mets:mdWrap>

</mets:dmdSec>

9.8.2.7 Sección Administrativa (amdSec).

Se incluirá, al menos, una sección administrativa para cada uno de los objetos digitales del AIP contenidos en la carpeta "objetos", ya sean másteres, derivados, ficheros en formato RAW de cámaras digitales, miniaturas u objetos con otros tipos de función. Se usarán obligatoriamente las siguientes subsecciones, y para cada uno de los objetos digitales a preservar digitalmente incluidos en el AIP:

- Metadatos técnicos (techMD). Se usará obligatoriamente el estándar PREMIS, que será complementado con:
    - Un elemento <premis:objectCharacteristicsExtension> con los metadatos técnicos en estándar MIX, en el caso de imágenes digitales raster, o de otros estándares específicos de otros medios (vídeo o audio) codificados en XML[28]. Se aplicará a todos los ficheros almacenados en la carpeta "objetos" que sean preservados finalmente: tiff, AVI, wav, mp4, wma, mp3, jpeg, raws…

[28] Existen varios formatos para vídeo y audio sobre XML y estándares para considerar de cara a la decisión final, aunque de entrada sólo se admitirán las últimas versiones de los estándares MPEG 7, PBCore o EBUCore.

- Los metadatos de objeto, además, serán los que aparecen en el siguiente ejemplo:

```
<premis:object xsi:type="premis:file">
        <premis:objectIdentifier>
         <premis:objectIdentifierType>UUID</premis:objectIdentifierType>
         <premis:objectIdentifierValue>0010000e-0004-4d43-a6f7-
928718b4b061</premis:objectIdentifierValue>
        </premis:objectIdentifier>
        <premis:objectCharacteristics>
         <premis:compositionLevel>0</premis:compositionLevel>
         <premis:fixity>
          <premis:messageDigestAlgorithm>SHA-
1</premis:messageDigestAlgorithm>

<premis:messageDigest>88f971b9a6973e872879b5f443f079b750c186e6</premi
s:messageDigest>
         </premis:fixity>
         <premis:size>443768</premis:size>
         <premis:format>
          <premis:formatDesignation>
           <premis:formatName>image/jpeg</premis:formatName>
           <premis:formatVersion>1.01</premis:formatVersion>
          </premis:formatDesignation>
          <premis:formatRegistry>
           <premis:formatRegistryName>PRONOM</premis:formatRegistryName>
           <premis:formatRegistryKey>fmt/43</premis:formatRegistryKey>
          </premis:formatRegistry>
         </premis:format>
         <premis:objectCharacteristicsExtension>
          <mix:mix>

   [….Aquí se incluyen los metadatos MIX o en otros estándares en XML aptos para
metadatos técnicos de vídeo o audio, tales como MPEG 7 o PBCore…]
        </mix:mix>
```

```
        </premis:objectCharacteristicsExtension>
      </premis:objectCharacteristics>
      <premis:originalName>objects\derivados\00000290700_0001-0010000e-0004-4d43-a6f7-928718b4b061.jpg</premis:originalName>
      <premis:relationship>
        <premis:relationshipType>derivation</premis:relationshipType>
        <premis:relationshipSubType>has source</premis:relationshipSubType>
        <premis:relatedObjectIdentification>

<premis:relatedObjectIdentifierType>UUID</premis:relatedObjectIdentifierType>
          <premis:relatedObjectIdentifierValue>0010000e-001a-4ca0-b977-2deedc10ca0c</premis:relatedObjectIdentifierValue>
        </premis:relatedObjectIdentification>
      </premis:relationship>
    </premis:object>
```

Se admitirán sólo como formatos XML para los metadatos técnicos de vídeo y audio digital los siguientes esquemas[29]:

- MPEG-7.
- PBCore 2.
- EBUCore 1.6.

No obstante lo anterior, el sistema de preservación queda abierto a otros posibles esquemas, previa valoración por los responsables del AEMA.

Las relaciones de derivación solo se incluirán si se dan entre diferentes ficheros a preservar correspondientes al mismo documento u obra.

- Metadatos de Procedencia Digital (digiprovMD). Se usará PREMIS, en concreto los elementos PREMIS Event y Agent que sean necesarios para indicar todos los procesos sufridos por los objetos y los agentes responsables de su ejecución. Para los ficheros derivados los eventos serán: metadata embedding, virus check, fixity check SIP, format identification, validation, replication y fixity check AIP. Para los ficheros másteres (TIFF, Wave, AVI, MOV…) y los RAWs además se incluirá el evento "capture". Los

[29] Estos metadatos pueden ser extraídos automáticamente con herramientas de uso libre como MediaInfo. Disponible en: https://mediaarea.net/

datos de agente de captura coincidirán con los nombres de las empresas digitalizadoras.

- Metadatos de derechos. Es obligatorio incluir al menos una sección de metadatos administrativos con metadatos de derechos (rightsMD). La declaración de derechos ha de ser acordada con los responsables del proyecto o de los fondos a depositar. Un ejemplo de declaración acorde a METS-Rights sería:

<rts:RightsDeclarationMD xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:rts="http://cosimo.stanford.edu/sdr/metsrights/" xsi:schemaLocation="http://cosimo.stanford.edu/sdr/metsrights/ http://www.loc.gov/standards/rights/METSRights.xsd" RIGHTSCATEGORY="COPYRIGHTED">

<rts:RightsDeclaration>El *[Sustituir por el nombre real del titular]* es el único titular de la propiedad intelectual de la presente copia digital y de sus metadatos, por lo que no podrá realizarse comercialización ni distribución alguna de los mismos por personas o entidades distintas del titular.</rts:RightsDeclaration>

<rts:RightsHolder RIGHTSHOLDERID="HOLDER001">

<rts:RightsHolderName>*[Sustituir por el nombre real del titular]* </rts:RightsHolderName>

<rts:RightsHolderContact>

<rts:RightsHolderContactAddress>*[Sustituir por los datos reales del titular]* </rts:RightsHolderContactAddress>

<rts:RightsHolderContactPhone PHONETYPE="BUSINESS">*[Sustituir por los datos reales del titular]* </rts:RightsHolderContactPhone>

<rts:RightsHolderContactEmail> *[Sustituir por los datos reales del titular]* </rts:RightsHolderContactEmail>

</rts:RightsHolderContact>

</rts:RightsHolder>

```
<rts:Context CONTEXTCLASS="GENERAL PUBLIC">

<rts:Permissions DISCOVER="true" DISPLAY="true" COPY="true"
DUPLICATE="false" MODIFY="false" DELETE="false" PRINT="true"
OTHER="false"/>
</rts:Context>

<rts:Context CONTEXTCLASS="REPOSITORY MGR">

<rts:Permissions DISCOVER="true" DISPLAY="true" COPY="true"
DUPLICATE="true" MODIFY="true" DELETE="true" PRINT="true"
OTHER="false"/>

</rts:Context>

</rts:RightsDeclarationMD>
```

La conexión de los metadatos de derechos con cada uno de los elementos file de la fileSec se hará a través del identificador (atributo ID) de esta nueva amdSec. Para ello se incluirá el valor del atributo ID dentro del atributo AMDID de cada uno de los elementos file, tal como vemos en el siguiente ejemplo:

```
<mets:amdSec ID="amdSec-00d00001-ffff-469c-a978-ec5089a46b5d">
<mets:rightsMD ID="rights-00d00001-ffff-469c-a978-ec5089a46b5d">
<mets:mdWrap MDTYPE="PREMIS:RIGHTS">
<mets:xmlData>
<rights>
<rightsStatement>
<rightsStatementIdentifier>
<rightsStatementIdentifierType>URI</rightsStatementIdentifierType>
<rightsStatementIdentifierValue>info:nyu-dl/x-v1/pr/xpnvx139/rmd/v0001
</rightsStatementIdentifierValue>
</rightsStatementIdentifier>
<rightsBasis>license</rightsBasis>
<licenseInformation>
<licenseIdentifier>
```

<licenseIdentifierType>NYU-Legal </licenseIdentifierType>

<licenseIdentifierValue>0123456789</licenseIdentifierValue>

</licenseIdentifier>

<licenseNote> Contact Information: New York University Office of Legal Counsel 70 Washington Square South 11h Floor New York, NY 10012 U.S.A. legal@nyu.edu </licenseNote>

</licenseInformation>

</rightsStatement>

</rights>

</mets:xmlData>

</mets:mdWrap>

</mets:rightsMD>

</mets:amdSec>

<mets:fileSec>

<mets:fileGrp ID="fileGrp-UUID-fijo-prueba">

<mets:file ID="_00d00001-0001-46cc-826a-bf2b95a0ff0c" MIMETYPE="application/xml" GROUPID="fileGrp-UUID-fijo-prueba" DMDID="dmd001 " ADMID="**amdSec-00d00001-ffff-469c-a978-ec5089a46b5d** amdSec-00d00001-0001-46cc-826a-bf2b95a0ff0c">

<mets:FLocat xlink:type="simple" xlink:href=" objetos\derivados\PDF\optimizado\fc_012772-00d00001-0005-469c-a978-ec5089a46b5d.pdf" OTHERLOCTYPE="SYSTEM" LOCTYPE="OTHER"/>

</mets:file>

<mets:file ID="_00d00001-0005-469c-a978-ec5089a46b5d" MIMETYPE="application/pdf" GROUPID="fileGrp-UUID-fijo-prueba" DMDID="dmdSec-UUID-fijo-prueba" ADMID="**amdSec-00d00001-ffff-469c-a978-ec5089a46b5d** amdSec-00d00001-0005-469c-a978-ec5089a46b5d">

<mets:FLocat xlink:type="simple" xlink:href=" objetos\derivados\TIFF_editado\fc_012772_0001-00d00001-0008-4fab-9cf4-601978f8f4d5.tif " OTHERLOCTYPE="SYSTEM" LOCTYPE="OTHER"/>

</mets:file>

[etc….]

No se creará redundancia incluyendo los metadatos de derechos dentro de todas las amdSec de cada uno de los ficheros incluidos en la fileSec. Como acabamos de explicar, los metadatos de derechos sólo se representarán una única vez dentro del documento METS en su propio elemento de metadatos administrativos.

Puede ocurrir que dentro de un mismo documento u obra se dé el caso de varios ficheros que no comparten la misma declaración de metadatos de derechos. En este caso se crearán tantas declaraciones de metadatos de derechos como se necesite en secciones amdSec independientes, procediéndose a la conexión de cada declaración de metadatos de derechos con sus ficheros correspondientes mediante la inclusión del valor del atributo ID de la amdSec en el atributo AMDID del elemento file que le corresponda. Este caso puede darse en algunas instituciones que no confieren las mismas autorizaciones de propiedad a las imágenes, según sean másteres, másteres editados, derivados en JPEG o derivados en PDF.

No se usará el elemento SourceMD (que se dedica en METS a incorporar metadatos del documento fuente original que ha sido digitalizado y al que corresponden los ficheros digitales que se empaquetan en el AIP), salvo que el remitente proporcione ya este tipo de metadatos codificados en XML (aptos y sin errores) para ser introducido en el fichero METS del AIP en la subsección sourceMD dentro de la sección amdSec.

Tal y como obliga el propio estándar METS, los elementos techMD, rightsMD y digiprovMD, llevarán obligatoriamente un atributo ID, cuyo valor será un identificador único. Aunque es opcional en este estándar, el elemento amdSec también llevará obligatoriamente un atributo ID con valor único.

Cuando los metadatos descriptivos sean objeto de procesos de migración, se deberán crear eventos PREMIS que los registren adecuadamente[30], usándose los siguientes posibles valores, según corresponda al tipo de evento, ya sea en su versión inglesa o española:

- METADATA_TRANSFORMATION. Transformación de un formato de metadatos en otro.
- METADATA_CREATION. Creación de un nuevo registro de metadatos.
- METADATA_MODIFICATION. Modificación de un registro de metadatos que no cambia el formato.
- METADATA_DELETION. Eliminación de un registro de metadatos.

Cuando el mapa estructural sea objeto de alguna transformación en el curso de los trabajos de preservación digital, se deberán crear eventos PREMIS que los registren adecuadamente[31], usándose los siguientes posibles valores, según corresponda al tipo de evento, ya sea en su versión inglesa o española:

[30] El vocabulario que sigue ha sido tomado del perfil , sacado de perfil METS público *ECHO Dep Generic METS Profile for Preservation and Digital Repository Interoperability*, disponible en http://www.loc.gov/standards/mets/profiles/00000015.html

[31] El vocabulario que sigue ha sido tomado del perfil , sacado de perfil METS público *ECHO Dep Generic METS Profile for Preservation and Digital Repository Interoperability*, disponible en http://www.loc.gov/standards/mets/profiles/00000015.htm l

- STRUCTMAP_TRANSFORMATION. Cambio de un mapa estructural que afecta a la compatibilidad con los sistemas de procesamiento actuales.
- STRUCTMAP_CREATION. Creación de un nuevo mapa estructural.
- STRUCTMAP_MODIFICATION. Cambio de un mapa estructural que no afecta a la compatibilidad con los sistemas de procesamiento actuales.
- STRUCTMAP_DELETION. Eliminación de un mapa estructural.

9.8.2.8 Sección de datos de ficheros (fileSec) con indicaciones específicas para los mapas estructurales.

Sólo se incluirán en esta sección los ficheros contenidos dentro de la carpeta objetos y sus carpetas hijas. No se admitirá, en consecuencia, la inclusión de los ficheros contenidos directamente en la carpeta padre del paquete AIP ni en sus hijas "data", "logs_datos_sip" y "metadatos_recibidos", salvo requerimiento del remitente, que deberá comunicarlo al sistema de preservación. No se admitirá la recursividad que implica referenciar al propio fichero METS desde su propia sección fileSec.

Los ficheros serán agrupados por su función, tal y como se expresa más abajo y en el siguiente ejemplo para una imagen máster, con el elemento fileGrp y su atributo:

<mets:fileGrp ID="fgr001" USE="master image">

Por consiguiente, la sección fileSec debe contener diversos elementos fileGrp dedicados a cada tipo de función de los ficheros del AIP. Por ejemplo, el fileSec de un documento METS típico de AIP podría contener un elemento fileGrp para agrupar las imágenes TIFF máster, otro elemento fileGrp para agrupar las imágenes derivadas de alta resolución en formato JPEG y otro fileGrp para contener las imágenes JPEG a baja resolución con la función de miniatura. En el caso de un video resultado de la digitalización de una cinta analógica, habrá un fileGrp para la versión máster (por ejemplo, un formato AVI con un códec sin compresión) y tantos fileGrp como versiones derivadas o editadas se hayan podido hacer (por ejemplo, un formato MP4 para difusión Web, y un formato MOV con códec de tipo intermedio con compresión con pérdida, como ProRes SD 422, para una versión editada a alta calidad pensada para su retransmisión televisiva) y se quieran preservar junto a su máster. Cada uno de estos elementos fileGrp correspondientes a imágenes derivadas podrían asimismo contener otros elementos fileGrp hijos para contener subgrupos que representen a otros tipos de derivados, como podrían ser ficheros multipágina en PDF, o ficheros derivados TIFF editados.

No se usará el atributo USE dentro del elemento file, por lo que incluso para agrupaciones de un solo elemento, como podría suceder en el caso de ficheros de audio o vídeo que cuentan normalmente con un solo ítem con la función de máster y con otro con la función de derivado, se usará obligatoriamente un elemento fileGrp.

Los ficheros además deben apuntar a las secciones informativas que les correspondan, tal y como vemos en el siguiente ejemplo, con el uso de los atributos DMDID y AMDID, que serán de uso obligatorio:

<mets:file ID="FID0" DMDID="DM1" AMDID="ADM1" SEQ="1" MIMETYPE="image/jpeg">

<mets:FLocat xlink:href=" objetos\derivados\fc_012772_0001-00d00001-0008-4fab-9cf4-601978f8f4d5.tif " OTHERLOCTYPE="SYSTEM" LOCTYPE="OTHER"/>

</mets:file>

<mets:file ID="FID1" DMDID="DM1" AMDID="ADM2" SEQ="2" MIMETYPE="image/jpeg">

<mets:FLocat xlink:href=" objetos\derivados\TIFF_editado\fc_012772_0001-00d00001-0009-f134-0be5-990878f8f4b3.tif " OTHERLOCTYPE="SYSTEM" LOCTYPE="OTHER"/>

</mets:file>

Los atributos DMDID y AMDID podrán llevar cuantos valores se requiera para poder conectar los ficheros con todas las declaraciones de metadatos que les correspondan, como podemos apreciar en el siguiente ejemplo:

<mets:file ID="App4FID1" MIMETYPE="image/tiff" SEQ="1" ADMID="App4ADM1 App4ADM2" DMDID="DM1 DM14" GROUPID="GID1">

Todos los metadatos descriptivos y administrativos insertos en el fichero METS deberán ser obligatoriamente conectados con los ficheros a que correspondan.

Todos los elementos file y fileGrp deberán llevar un atributo ID cuyo valor debe ser único.

En el caso de documentos u obras de arte compuestos por páginas o ítems en secuencia, los elementos file deben llevar obligatoriamente un atributo SEQ cuyo valor será el número correspondiente de la secuencia de objeto en el orden lógico del original. Por ejemplo, SEQ="1" en un libro podría corresponder a la captura de la cubierta, SEQ="2" a la de la primera página de guarda, SEQ="3" a la página 1, etc. El orden se entiende dentro de una misma agrupación de documentos, esto es, dentro del elemento fileGrp que contiene a todos los ficheros que le corresponden.

El atributo MIMETYPE debe ser incluido para todos los elementos file obligatoriamente. Su valor debe coincidir necesariamente con alguno de los valores admitidos por el estándar *IANA MIME Media Type* en su última actualización[32]. En el caso de documentos de texto y si está disponible el dato se

[32] Accesible desde http://www.iana.org/assignments/media-types/media-types.xhtml

incluirá el charset (sistema de codificación de caracteres usado). Si se desconociera, sin posibilidad de averiguarlo, entonces se usará el valor "application/octet-stream".

El atributo GROUPID es obligatorio siempre que haya necesidad de agrupar ficheros correspondientes a un mismo ítem, como explicamos más abajo en el epígrafe dedicado a describir cómo relacionar cada fichero máster con su correspondiente o correspondientes ficheros derivados.

El elemento file debe contener obligatoriamente un elemento FLocat que incluya la dirección física al fichero. El elemento FLocat debe tener un atributo LOCTYPE con un valor de "OTHER" acompañado de un atributo OTHERLOCTYPE="SYSTEM". Este elemento debe también tener un atributo xlink:href que contenga la ruta de acceso al fichero. La ruta debe ser siempre relativa a la localización del propio documento METS dentro del paquete AIP.

Los metadatos administrativos de propiedad intelectual (METSRights), previamente declarados en una sección de metadatos administrativos específica, se podrán vincular únicamente a los elementos fileGrp que organizan los elementos file en la sección de ficheros (fileSec). De esta forma se evita incluir el valor del atributo ID de la <admSec> de derechos en el atributo ADMID de los elementos file. Mediante este procedimiento se usará el atributo ADMID en los elementos fileGrp, cuyo valor será el valor del atributo ID de la <admSec> dedicada a los metadatos METSRights, como podemos apreciar en el siguiente ejemplo:

<mets:fileSec>

<mets:fileGrp **ADMID="amdSec-rights-0010000e-0000-41d8-869e-30d3dba619a2"** ID="fileGrp-Master_Image-0010000e-0000-41d8-869e-30d3dba619a2" USE="master image">

<mets:file ADMID="amdSec-0010000e-001a-4ca0-b977-2deedc10ca0c" DMDID="dmdSec-0010000e-0000-41d8-869e-30d3dba619a2 dmdSec-0010000e-0000-4122-060f-0012dba609b3" GROUPID="page_1" ID="_0010000e-001a-4ca0-b977-2deedc10ca0c" MIMETYPE="image/tiff" SEQ="1">

<mets:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="objects\masteres\00000290700_0001-0010000e-001a-4ca0-b977-2deedc10ca0c.tif" xlink:type="simple"/>

</mets:file>

<mets:file ADMID="amdSec-0010000e-001b-4cb2-9e7d-d90fc17f20ce" DMDID="dmdSec-0010000e-0000-41d8-869e-30d3dba619a2 dmdSec -0010000e-0000-4009-ffc1-32e0d1a71903" GROUPID="page_2" ID="_0010000e-001b-4cb2-9e7d-d90fc17f20ce" MIMETYPE="image/tiff" SEQ="2">

<mets:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="objects\masteres\00000290700_0002-0010000e-001b-4cb2-9e7d-d90fc17f20ce.tif" xlink:type="simple"/>

</mets:file>

[…]

Los documentos simples gráficos con ficheros sólo para una de sus caras llevarán un solo elemento file dentro de la fileSec para cada uno de los ficheros, dentro de un elemento fileGroup específico, tal y como vemos en el siguiente ejemplo, para el caso de una fotografía para la que hay versión máster en TIFF, derivado en JPEG y RAW en DNG:

<mets:fileSec>

<mets:fileGrp ID="fileGrp-Master_Image-0010000e-0000-41d8-869e-30d3dba619a2" USE="master image" ADMID="amdSec-rights-0010000e-0000-41d8-869e-30d3dba619a2">

<mets:file ADMID="amdSec-0010000e-001a-4ca0-b977-2deedc10ca0c" DMDID="dmdSec-0010000e-0000-41d8-869e-30d3dba619a2 dmdSec -0010000e-0000-00d8-869e-30d3dba619a3" GROUPID="cara_1" ID="_0010000e-001a-4ca0-b977-2deedc10ca0c" MIMETYPE="image/tiff" >

<mets:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="objects\masteres\006309-00100001-0000-563f-7890-3452d9960444.tif" xlink:type="simple"/>

</mets:file>

</mets:fileGrp>

<mets:fileGrp ID="fileGrp-Reference_Image-0010000e-0000-41d8-869e-30d3dba619a2" USE="reference image" ADMID="amdSec-rights-0010000e-0000-41d8-869e-30d3dba619a2">

<mets:file ADMID="amdSec-0010000e-0004-4d43-a6f7-928718b4b061" DMDID="dmdSec-0010000e-0000-41d8-869e-30d3dba619a2 dmdSec -0010000e-0000-00f8-345a-30d3dba887b3" GROUPID="cara_1" ID="_0010000e-0004-4d43-a6f7-928718b4b061" MIMETYPE="image/jpeg" >

<mets:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="objects\derivados\006309-00100001-0000-563f-1840-45d2d9960f55.jpg" xlink:type="simple"/>

</mets:file>

</mets:fileGrp>

<mets:fileGrp ID="fileGrp-Raw_Image-0010000e-0000-41d8-869e-30d3dba619a2" USE="raw image" ADMID="amdSec-rights-0010000e-0000-41d8-869e-30d3dba619a2">

<mets:file ADMID="amdSec-0010000e-0004-4d43-a6f7-928718b4b061" DMDID="dmdSec-0010000e-0000-41d8-869e-30d3dba619a2 dmdSec -0010000e-

0000-567c-908e-30d3dba77a3" GROUPID="cara_1" ID="_0010000e-0004-4d43-a6f7-928718b4b061" MIMETYPE="image/x-raw">

<mets:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="objects\derivados\006309-00100001-0000-234a-258b-5552d9960f10.dng" xlink:type="simple"/>

</mets:file>

</mets:fileGrp>

</mets:fileSec>

En este caso no se usará el elemento SEQ, puesto que no tenemos ninguna secuencia de ficheros. Y se usa el elemento GROUPID con valor "cara_1", al entenderse que se ha capturado la cara principal del soporte fotográfico, la que lleva la imagen.

El mapa estructural será único, y sus elementos DIV de fichero no llevarán el atributo ORDER, ya que no tenemos una secuencia de ficheros. A continuación vemos el mapa estructural correspondiente a la fileSec anterior:

<mets:structMap ID="structMapPhysical-0010000e-0000-41d8-869e-30d3dba619a2" LABEL="PRIMARY_STRUCTMAP" TYPE="physical">

<mets:div DMDID="dmdSec-0010000e-0000-41d8-869e-30d3dba619a2 dmdSec-0010000e-0000-56fa-967b-400adba892f4" ID="_0010000e-0000-41d8-869e-30d3dba619a2_DataDirectory" LABEL="Data Directory: 006309-0010000e-0000-41d8-869e-30d3dba619a2\data\objetos" TYPE="Directory">

<mets:div ID="div-0010000e-0003-472b-adb6-8573a40e7374" LABEL="derivados" TYPE="Directory">

<mets:div ID="div-0010000e-0004-4d43-a6f7-928718b4b061" LABEL="006309-00100001-0000-563f-1840-45d2d9960f55.jpg" TYPE="Item">

<mets:fptr FILEID="_0010000e-0004-4d43-a6f7-928718b4b061"/>

</mets:div>

</mets:div>

<mets:div ID="div-0010000e-0019-40b5-ad6a-fd9c566d4459" LABEL="masteres" TYPE="Directory">

<mets:div ID="div-0010000e-001a-4ca0-b977-2deedc10ca0c" LABEL="006309-00100001-0000-563f-7890-3452d9960444.tif" TYPE="Item">

<mets:fptr FILEID="_0010000e-001a-4ca0-b977-2deedc10ca0c"/>

</mets:div>

</mets:div>

<mets:div ID="div-0010000e-0019-40b5-ad6a-fd9c566d4459" LABEL="raws" TYPE="Directory">

<mets:div ID="div-0010000e-001a-4ca0-b977-2deedc10ca0c" LABEL="006309-00100001-0000-234a-258b-5552d9960f10.dng" TYPE="Item">

<mets:fptr FILEID="_0010000e-001a-4ca0-b977-2deedc10ca0c"/>

</mets:div>

</mets:div>

</mets:div>

</mets:structMap>

Para los documentos simples gráficos con ficheros para sus dos caras, el modelo será igual al anterior, salvo que:

- En la fileSec habrá dos elementos file (uno para cada cara de la fotografía) y los elementos file llevarán el atributo SEQ, con valor 1 o 2, para identificar el orden de caras en la secuencia. El GROUPID de cada elemento file llevará como valor "cara_1" o "cara_2", según corresponda: “cara_1”, para el anverso, que suele estar identificado en el nombre de fichero con la secuencia de caracteres “AN” justo al final del nombre y antes del punto de separación de la extensión; y “cara_2”, para el reverso, que suele estar identificado en el nombre de fichero con la secuencia de caracteres “RV” justo al final del nombre y antes del punto de separación de la extensión.
- En el mapa estructural habrá dos elementos DIV de fichero por cada elemento DIV de directorio (carpeta). Cada elemento DIV de fichero deberá llevar el atributo ORDER con el valor 1 o 2, de acuerdo a su secuencia.

En el caso de los documentos u obras multipágina:

- La fileSec tendrá tantos elementos file como páginas, que llevarán el atributo SEQ, con el valor numérico correspondiente a su orden dentro de la secuencia de páginas.
- En el mapa estructural habrá tantos elementos DIV de fichero dentro de cada elemento DIV de directorio (carpeta), como ficheros hayan sido declarados dentro de la fileSec. Cada elemento DIV de fichero deberá llevar el atributo ORDER con el valor numérico que le corresponde dentro de la secuencia de páginas.

9.8.2.9 Sección de mapa estructural (structMap). Normas generales.

Se creará en principio un único mapa, que reflejará la estructura física de carpetas del AIP con objetos digitales del documento u obra a preservar digitalmente, esto es, partiendo de la carpeta padre “objetos”. Hemos de pensar que la finalidad de este mapa no es la de aportar un medio de navegación alternativo por los objetos digitales del documento a preservar ni la de vehicular metadatos estructurales de los documentos u obras originales que fueron digitalizados. Su finalidad es servir como sistema de documentación a nivel de estructura del paquete de preservación. Su función principal es ofrecer metadatos estructurales que registran la disposición de los objetos a preservar dentro del paquete de preservación AIP (ubicación, secuencia y estructura) y que facilitan la navegación dentro de este paquete y la realización de operaciones de procesado, transferencia y preservación digital.

Se admitirá en el futuro la inclusión de otros mapas estructurales si así se decide por parte de los responsables de preservación del AEMA, por ello el mapa estructural referido en el primer párrafo será considerado como el mapa estructural primario, y será identificado como tal mediante su atributo LABEL que deberá llevar como valor “PRIMARY_STRUCTMAP”.

Cualquier cambio en el mapa estructural creado originalmente en el AIP deberá ser reflejado en los metadatos de evento.

Se podrán incluir tantas divisiones estructurales se necesiten en el mapa estructural primario, pero siempre siguiendo la normativa de estructura de los AIP asentada por la normativa del sistema de preservación.

En el momento actual, de acuerdo a esta normativa, el mapa estructural primario deberá ser de tipo físico referido a la estructura del paquete AIP, aunque para su mejor identificación en procedimientos automatizados deberá llevar como valor del atributo TYPE la cadena "PHYSICAL” y como valor de su atributo LABEL la cadena “PRIMARY_STRUCTMAP”. Deberá llevar también un atributo ID con valor único, como podemos apreciar en el siguiente ejemplo:

<mets:structMap LABEL="PRIMARY_STRUCTMAP" TYPE=”PHYSICAL” ID="structMap001">

A continuación del elemento padre structMap se incluirá la estructura de carpetas del AIP mediante elementos DIV anidados, como se muestra en el siguiente ejemplo:

```
<mets:structMap ID="structMap-00100001-0000-4c3d-8d8f-6fd2d9996e4a" TYPE="physical" LABEL="PRIMARY_STRUCTMAP">
  <mets:div ID="_00100001-0000-4c3d-8d8f-6fd2d9996e4a-DataDirectory" LABEL="Data Directory: 00000111600_Los_buenos_mozos-00100001-0000-4c3d-8d8f-6fd2d9996e4a\data\objetos"
  TYPE="Directory" DMID="dmdSec-00100001-0000-4c3d-8d8f-6fd2d9996e4a">
    <mets:div ID="div-00100001-0003-4433-b715-895058073fc4" LABEL="derivados" TYPE="Directory">
      <mets:div ID="div-00100001-0004-4be6-b897-6290287c171e" LABEL="00000111600_0001-00100001-0004-4be6-b897-6290287c171e.jpg" ORDER="1" TYPE="Item">
        <mets:fptr FILEID="_00100001-0004-4be6-b897-6290287c171e"/>
      </mets:div>
      <mets:div ID="div-00100001-0005-4b3f-ac64-817dc247e072" LABEL="00000111600_0002-00100001-0005-4b3f-ac64-817dc247e072.jpg" ORDER="2" TYPE="Item">
        <mets:fptr FILEID="_00100001-0005-4b3f-ac64-817dc247e072"/>
      </mets:div>
      <mets:div ID="div-00100001-0006-488b-8a91-9c95b819f752" LABEL="00000111600_0003-00100001-0006-488b-8a91-9c95b819f752.jpg" ORDER="3" TYPE="Item">
        <mets:fptr FILEID="_00100001-0006-488b-8a91-9c95b819f752"/>
      </mets:div>
      <mets:div ID="div-00100001-0007-4873-818e-c788ccf568e0" LABEL="00000111600_0004-00100001-0007-4873-818e-c788ccf568e0.jpg" ORDER="4" TYPE="Item">
        <mets:fptr FILEID="_00100001-0007-4873-818e-c788ccf568e0"/>
      </mets:div>
```

Como podemos apreciar en el anterior ejemplo, debe haber elementos DIV para indicar los directorios (carpetas, con atributo TYPE=”directory”) y ficheros (con atributo TYPE=”Item”).

Se usará el atributo LABEL para representar el nombre exacto de la carpeta o fichero. En el caso de la carpeta padre “objetos”, el valor del atributo LABEL será el nombre de la carpeta antecedida por su ruta, considerada desde el nombre de la carpeta padre del paquete AIP, tal y como vemos en el siguiente ejemplo:

<mets:structMap LABEL="PRIMARY_STRUCTMAP" TYPE=”PHYSICAL” ID="structMap-00100001-0000-4991-824b-eb56cb3a79a2">

<mets:div TYPE="Directory" ID="_00100001-0000-4991-824b-eb56cb3a79a2-DataDirectory" DMDID="dmdSec-00100001-0000-4c3d-8d8f-6fd2d9996e4a” **LABEL="Data Directory: 7020_Colleccion de viajes.REDUX-00100001-0000-4991-824b-eb56cb3a79a2\data\objetos**">

<mets:div TYPE="Directory" ID="div-00100001-0003-462f-9df5-dc798e92bc21" LABEL="derivados">

<mets:div TYPE="Item" ID="div-00100001-0004-48d0-9f6f-f263635df253" LABEL="0001-00100001-0004-48d0-9f6f-f263635df253.jpg" ORDER="1">

<mets:fptr FILEID="_00100001-0004-48d0-9f6f-f263635df253"/>

</mets:div>

<mets:div TYPE="Item" ID="div-00100001-0005-47d9-8a3f-6af8b38e4736" LABEL="0002-00100001-0005-47d9-8a3f-6af8b38e4736.jpg" ORDER="2">

<mets:fptr FILEID="_00100001-0005-47d9-8a3f-6af8b38e4736"/>

[…]

Este primer DIV debe incluir un atributo DMDID que enlace con una sección de metadatos descriptivos (<dmdSec>). El valor de este atributo ha de ser idéntico al valor del atributo ID de la referida sección de metadatos descriptivos.

Un elemento DIV puede o no puede contener directamente elementos fptr. En el caso de que sea un DIV dedicado a una carpeta que no tenga ficheros hijo, sino sólo carpetas hija no contendrá dicho elemento.

Se usará un atributo ID para ambos tipos de DIV, con un valor único. Los elementos DIV tipo Item llevarán siempre y obligatoriamente como elemento hijo un elemento fptr cuyo valor de atributo FILEID será el identificador único de fichero declarado en el elemento file correspondiente mediante el atributo ID, tal y como vemos en el siguiente ejemplo:

<mets:file **ID="_00d00001-0005-469c-a978-ec5089a46b5d"** MIMETYPE="application/pdf" GROUPID="fileGrp-UUID-fijo-prueba" DMDID="dmdSec-UUID-fijo-prueba" ADMID="amdSec-00d00001-0005-469c-a978-ec5089a46b5d">

<mets:FLocat xlink:type="simple" xlink:href="
objetos\derivados\PDF\optimizado\fc_012772-00d00001-0005-469c-a978-
ec5089a46b5d.pdf" OTHERLOCTYPE="SYSTEM" LOCTYPE="OTHER"/>

</mets:file>

[…]

<mets:div TYPE="Item" ID="div-00d00001-0005-469c-a978-ec5089a46b5d"
LABEL="fc_012772-00d00001-0005-469c-a978-ec5089a46b5d.pdf">

<mets:fptr FILEID="_00d00001-0005-469c-a978-ec5089a46b5d"/>

</mets:div>

[…]

Los elementos DIV no deben contener en ningún caso elementos mptr. Estos elementos son punteros a contenido representado por un documento METS externo.

9.8.2.10 Cómo especificar las derivaciones de unos ficheros desde otros ficheros y viceversa.

Se hará con el elemento Premis RELATIONSHIP y sus subtipos de relación "is source" y "has source" dentro del elemento OBJECT.

Veamos un ejemplo, en el caso de la representación de los metadatos en el caso de ficheros derivados desde una versión máster:

<premis:object xsi:schemaLocation="info:lc/xmlns/premis-v2
http://www.loc.gov/standards/premis/v2/premis-v2-2.xsd" version="2.2"
xsi:type="premis:file" xmlns:premis="info:lc/xmlns/premis-v2">

<premis:objectIdentifier><premis:objectIdentifierType>UUID</premis:objectIden
tifierType><premis:objectIdentifierValue>8f13c43c-38cb-47c4-98dd-
417668d57b40</premis:objectIdentifierValue></premis:objectIdentifier>

<premis:relationship>

<premis:relationshipType>derivation</premis:relationshipType>

<premis:relationshipSubType>has source</premis:relationshipSubType>

<premis:relatedObjectIdentification>

<premis:relatedObjectIdentifierType>UUID</premis:relatedObjectIdentifierType>

<premis:relatedObjectIdentifierValue>fb4adc6e-f9a4-4c45-ae35-f49506c7432d</premis:relatedObjectIdentifierValue>

</premis:relatedObjectIdentification>

<premis:relatedEventIdentification>

<premis:relatedEventIdentifierType>UUID</premis:relatedEventIdentifierType>

<premis:relatedEventIdentifierValue>955db4fc-ed3f-430b-b477-e9fbabd082c0</premis:relatedEventIdentifierValue>

</premis:relatedEventIdentification>

</premis:relationship>

</premis:object>

Veamos un ejemplo en el caso de los másteres de los cuales se derivan los ficheros derivados, en el que mostramos sólo el elemento relationship y su contenido:

<premis:relationship>

<premis:relationshipType>derivation</premis:relationshipType>

<premis:relationshipSubType>is source of</premis:relationshipSubType>

<premis:relatedObjectIdentification>

<premis:relatedObjectIdentifierType>UUID</premis:relatedObjectIdentifierType>

<premis:relatedObjectIdentifierValue>dfca8d45-8433-4e16-9fc9-fb28cc9418fd</premis:relatedObjectIdentifierValue>

</premis:relatedObjectIdentification>

<premis:relatedEventIdentification>

<premis:relatedEventIdentifierType>UUID</premis:relatedEventIdentifierType>

<premis:relatedEventIdentifierValue/>

</premis:relatedEventIdentification>

</premis:relationship>

9.8.2.11 Cómo especificar la función de los ficheros: si máster, derivado o miniatura.

Se usará el elemento FILEGROUP de la sección fileSec, indicando la función mediante el atributo USE, como vemos en el siguiente ejemplo:

<mets:fileGrp USE="master image">

Los posibles valores para el atributo USE serán, en el caso de que se decida aplicar un vocabulario en inglés:

- master image. Para un fichero de imagen raster o vectorial con la función de máster.
- raw image. Para un fichero en formato RAW de cámara digital (DNG, CR2, NEF…)
- master audio. Para un fichero de audio con la función de máster.
- master video. Para un fichero de audio con la función de máster.
- master text. Para un fichero de texto con la función de máster.
- reference image. Para un fichero de imagen raster o vectorial con la función de derivado de visualización o uso (derivado no miniatura).
- reference audio. Para un fichero de audio con la función de derivado.
- reference video. Para un fichero de vídeo con la función de derivado.
- reference text. Para un fichero de texto con la función de derivado.
- thumbnail. Para un fichero de miniatura.

Si hay otros tipos de derivados como “TIFF Editado” o “PDF_OCR” se deben generar subgrupos dentro del elemento fileGrp para estos derivados, de manera que quede reflejada esta tipología de forma jerárquica. Así lo recomienda el propio estándar METS que se proceda, como en el siguiente ejemplo sacado de la propia normativa METS:

<mets:**fileSec**>

<mets:**fileGrp** ID=”TIFF_GRP01” USE=”master image”>

. . .

</mets:fileSec>

Al ser derivados sus valores para el atributo USE serán del tipo reference, según corresponda a su tipo de medio.

9.8.2.12 Cómo relacionar cada fichero máster con su correspondiente o correspondientes ficheros derivados.

Cada versión máster debe quedar también relacionada con su versión derivada siguiendo el procedimiento que explicamos a continuación, además de la relación de derivación que se ha expresado mediante el elemento PREMIS Relationship. Esto es necesario porque PREMIS no es METS, y se debe poder especificar desde el propio METS esta relación.

La relación se hará a través del atributo GROUPID del elemento FILE. El valor de este atributo debe es idéntico entre el máster y sus derivados para que se pueda representar sin ambigüedad esta relación.

9.8.2.13 Ficheros METS heredados.

No se admitirá en ningún la sustitución del fichero METS del AIP por ficheros METS heredados, ya complementen o ya sustituyan al fichero METS del AIP. Hemos de pensar que el fichero METS del AIP no cumple la misma función que estos ficheros METS que pueden proporcionar algunos remitentes generados directamente o no desde sus aplicaciones de gestión de sus bibliotecas o archivos digitales, por lo que los requisitos y exigencias de codificación no coinciden en su totalidad.

Los ficheros METS heredados se guardarán en la carpeta del paquete AIP metadatos heredados.

9.8.2.14 Ausencia de los elementos strucLink y behaviorSec.

No se permite el uso de los elementos strucLink y behaviorSec.

9.9 Procedimientos automatizados para la conversión de los paquetes SIP a paquetes AIP y su registro de datos en el sistema de gestión para poder generar automáticamente a partir de ellos los metadatos PREMIS y los informes de procesados.

Estos procedimientos se equiparan a los eventos de la entidad Eventos de PREMIS, aunque puede haber otros procedimientos no equiparables y que por tanto no deban quedar recogidos en los metadatos PREMIS del AIP. En el primer caso, sus datos deben quedar necesariamente registrados tanto en la declaración de metadatos PREMIS en el fichero METS correspondiente al AIP como en el sistema de gestión del sistema de preservación; en el segundo caso, quedarán sólo registrados en el sistema de gestión. Señalamos más abajo la obligatoriedad de representación como PREMIS de los procesos para cada uno de ellos.

9.9.1 Procesados de cada uno de los ficheros del SIP para su normalización en el AIP.

No necesariamente deben darse estos procesos en este mismo orden, ni incluso separados. El orden y unión o separación de procesos finales vendrán dados por la eficiencia del código de programación que se cree para la automatización de los procesados.

9.9.1.1 Chequeo antivirus para todos los ficheros ingestados o transformados.

Es metadato PREMIS de consignación obligatoria, bajo el tipo de evento "chequeo de virus ".

Se trata de detectar la presencia de virus o cualquier tipo de código malicioso en todos los ficheros del SIP justo antes de su almacenamiento en la unidad de disco donde van a ser procesados durante la transformación del paquete SIP a AIP.

Se debe actualizar o al menos comprobar el grado de actualización del antivirus de forma regular de manera que antes de cada procesado se asegure su máxima actualización.

9.9.1.2 Chequeo de integridad de cada fichero del SIP antes de cualquier otra modificación.

Es metadato PREMIS de consignación obligatoria, bajo el tipo de evento "chequeo de integridad".

Se hará contra el código hash aportado por el remitente o por el sistema de preservación en el momento de ingreso del PreSIP. O con el código hash creado en el momento del ajuste de los contenidos a la normativa de los SIP. Se hará para cada fichero del SIP, incluyendo el fichero o ficheros de metadatos o cualquier otro fichero de control aportado por el remitente.

9.9.1.3 Limpieza de paquete SIP. Comprobación y eliminación en su caso de ficheros o carpetas que figuran por error.

Es posible que a la hora de evaluar el PreSIP o de su ajuste como SIP, en su caso, al operador del módulo de ingreso se le haya pasado por alto la presencia de ficheros temporales que no fueron borrados en su día por error, o de otros ficheros sin información que es preciso eliminar. En ese caso se deberá "limpiar" el paquete SIP de manera que no se incluyan estos ficheros indeseados en el AIP.

9.9.1.4 Renombrado de ficheros y carpetas y reestructuración de carpetas.

Es metadato PREMIS de consignación obligatoria, bajo el tipo de evento "cambio de nombre".

Se hará de acuerdo a la normativa de nomenclatura y estructura de los AIP referida anteriormente.

9.9.1.5 Normalización de ficheros.

Es metadato PREMIS de consignación obligatoria, bajo el tipo de evento "migración".

Se trata de transformar el formato o la versión de formato para cumplir la normativa de formatos del plan de preservación aplicado. Es un proceso de migración que se aplica en el momento de conversión al AIP, por lo que deberá abrirse en los metadatos PREMIS una entrada para el evento de tipo migración.

9.9.1.6 Cálculo de código hash de los ficheros transformados durante el proceso de generación del AIP.

Es metadato PREMIS de consignación obligatoria, bajo el tipo de evento "código hash ".

Se calcularán, o reutilizarán para los ficheros heredados del SIP ya validados en integridad y que no han sufrido ninguna transformación, los códigos hash de todos los ficheros de contenido integrados en el paquete AIP, una vez realizados todos los procesos de ubicación en su carpeta correspondiente, de migración de formato o versión de formato y de chequeo antivirus. Se recogerán en la base de datos de gestión del repositorio los datos del método hash aplicado, la fecha de aplicación y el código resultante, de acuerdo al formato PREMIS.

9.9.1.7 Identificación de formato y versión de formato para todos los ficheros ingestados del SIP o para sus versiones transformadas.

Es metadato PREMIS de consignación obligatoria, bajo el tipo de evento "identificación de formato ".

Este proceso, como hemos referido en varias partes de este documento, se puede automatizar mediante herramientas de gran uso por parte de la comunidad de expertos y usuarios de la preservación digital, como son DROID y JHOVE.

Es importante que queden perfectamente registrados todos los datos que la normativa PREMIS indica para este proceso, pues la correcta identificación y documentación del formato de un fichero es uno de los aspectos principales de la preservación digital.

9.9.1.8 Validación de todos los formatos de ficheros creados, ingestados o transformados.

Es metadato PREMIS de consignación obligatoria, bajo el tipo de evento "validación de formato".

Se trata de comprobar que el fichero está bien formado de acuerdo a su especificación técnica. Este proceso es automatizable a través de la aplicación libre JHOVE.

9.9.1.9 Generación de los ficheros de control y mapeado de carpetas y ficheros entre SIP y AIP.

Se trata de los ficheros referidos en la normativa de empaquetamiento y representación de los AIP, ya referida más arriba: son los ficheros: listado.txt, tab_corp.txt, sip_estr_crp.txt, y Id_form_fich.txt. También incluimos aquí el fichero *check_aip.txt*.

9.9.1.10 Caracterización y extracción de metadatos de los ficheros.

Se trata de extraer los metadatos técnicos de las cabeceras de los ficheros contenidos en el AIP (no de los de control) para la representación de los metadatos PREMIS de objeto, evento y agente.

9.9.1.11 Creación y validación del fichero METS con los metadatos del paquete AIP.

El procedimiento ideal es que todos los metadatos necesarios para crear el fichero METS se extraigan automáticamente de sus fuentes, que son: las cabeceras de los ficheros del AIP de objeto, el sistema de archivos del sistema operativo del ordenador donde se almacenan los AIP, los ficheros de metadatos aportados por el remitente y la propia base de datos del sistema de gestión del repositorio (para los datos sobre los documentos que ya han sido grabados en el sistema de base de datos del repositorio durante la fase de ingesta, como, por ejemplo, los datos de propiedad intelectual). Para posteriormente registrarlos en el sistema de base de datos del repositorio, de manera que puedan generarse automáticamente a partir de estos datos registrados los ficheros METS.

Debido a la complejidad de este proceso, es conveniente que se valide la corrección del fichero METS creado de manera automática.

9.9.1.12 Creación de ficheros BagIT.

Son los ficheros TXT normativos de BagIT. Su automatización es fácil, pues existen no sólo herramientas software libre para hacerlo, sino una completa especificación de este estándar, tal y como comentamos más arriba en el epígrafe dedicado al sistema de empaquetamiento de los AIP.

9.9.1.13 Chequeo de validez de los ficheros BagIT y de control.

Para los ficheros Bag-It se pueden usar aplicaciones software libre ideados para realizar este procedimiento, accesibles desde la página Web del este estándar, como referimos más arriba.

Si así se considera relevante, sería también de utilidad poder validar la generación de los ficheros de control no Bag-It.

9.9.2 Datos a registrar por cada procesado de fichero ubicado dentro de la carpeta objetos.

Como indicamos más arriba, se deben registrar los mismos datos de control y de proceso de los AIPs que se registren en los metadatos PREMIS en la propia base de datos del sistema de gestión del repositorio, además de otros datos de procesos no equiparables a PREMIS. Es más, los metadatos PREMIS se deberían poder obtener automáticamente desde la base de datos de gestión. Si los datos en la base de datos de gestión del repositorio se estructuran de una forma compatible con PREMIS, el proceso de obtención de los metadatos PREMIS en formato XML compatible con METS se podrá automatizar con relativa facilidad.

Vamos a detenernos a continuación en el registro de los metadatos PREMIS, dada la necesidad de que los metadatos de preservación digital se ajusten lo más posible a este estándar.

En la terminología PREMIS cada proceso sufrido por un objeto digital a preservar recibe el nombre de evento, y es necesario que queden registrados por cada uno de ellos una serie de datos obligatorios, más la posibilidad de registrar otros que son opcionales.

Los datos a registrar por cada uno de los procesos (eventos PREMIS) que reciba un fichero serán obligatoriamente:

- Nombre de fichero procesado del AIP. Obligatorio.

- Si el fichero es producto o fuente de una derivación (si se da el caso de que el fichero deriva de otro fichero del SIP o es fuente para crear otro fichero del AIP). Obligatorio.
- Nombre del fichero fuente del SIP del que procede (cuando hay una derivación, si se da el caso de que procede de otro fichero del SIP que ha sido procesado para dar el AIP, por ejemplo en el caso de que haya que aplicar un proceso de migración desde el fichero SIP para conseguir un fichero AIP de acuerdo a las normas del plan de preservación que se le aplica). Obligatorio.
- Nombre del fichero del que es fuente (en el caso de derivación, si se da el caso de que a su vez el fichero AIP ha sido tomado como fichero fuente para crear otro fichero del AIP). Obligatorio.
- Tipo de identificador único de evento aplicado al evento (procesado). Obligatorio.
- Identificador único de evento que pondrá el sistema automáticamente (puede ser un código UUID u otro tipo de identificador único generado por el sistema de gestión del repositorio). Obligatorio.
- Tipo evento (tipo de procesado aplicado). Obligatorio.
- Detalle del evento (por ejemplo, el programa usado para generar el evento y la línea de comandos usada para activar una función del programa, tal como una línea de comandos de la aplicación ImageMagick). Este dato es opcional en PREMIS pero es relevante que se custodie.
- Fecha y hora de aplicación del evento. Obligatorio.
- Tipo de identificador de agente que activa el evento. Obligatorio.
- Valor del identificador de agente que activa el evento. Obligatorio.
- Resultado del evento en su caso (por ejemplo si se ha aplicado un generador de códigos UUID o algoritmo de hash, el código resultante). Es opcional en PREMIS pero es relevante que se consigne.

Si algún evento no tiene alguno de estos elementos no se registrará el dato, salvo los que aparecen como obligatorios, cuyos valores que deberán ser localizados. Los datos de información de agente deben poder ser repetibles, pues en un evento pueden participar varios agentes.

A partir del registro de todos estos datos en la base de datos del sistema de preservación podrá generarse automáticamente la declaración de metadatos PREMIS XML sobre eventos, objetos y agentes compatible con las secciones de metadatos de procedencia digital y técnicos de METS, y para cada objeto incluido en el AIP.

Para la mejor compatibilidad con PREMIS y otros estándares de preservación digital, es recomendable registrar los datos de fecha y hora de la manera más exacta posible, incluyendo la zona horaria que se toma como referencia.

Los valores admitidos para el campo tipo evento son[33]:

Tipo de evento	Definición
compresión	Se aplica un método de compresión para reducir su tamaño lo que implica una recodificación de los datos del fichero.
recodificación	Se cambia el sistema de codificación, por ejemplo un fichero de texto codificado en ANSI pasa a ser codificado en UTF-8.
cambio de nombre	Se cambia el nombre del fichero para adecuarlo a la normativa de nomenclatura del repositorio.
cambio de carpeta donde se ubica el fichero	Se cambia de carpeta donde se ubica el fichero.
UUID	Cálculo y asignación de un código UUID.
código hash	Cálculo y asignación de código hash al fichero.
chequeo de virus	Comprobación de no existencia de virus o códigos maliciosos.
identificación de formato	Se identifica el formato y la versión de formato que tiene el fichero.
ingesta	Entrada del fichero al sistema de preservación.
registro	Asignación de un código de registro del fichero en el sistema.

[33] Para su elaboración nos hemos basado en la propuesta de la propia especificación PREMIS, aunque esta propuesta no es normativa, pudiendo cada organización utilizar sus propios listados de evento.

creación	Creación de fichero.
chequeo de integridad	Chequeo de integridad para comprobar que el fichero no ha sido modificado (fixity check).
validación de formato	Validación de conformidad del formato a su especificación técnica.
captura	Procedimiento mediante el cual un repositorio obtiene un objeto de forma activa.
eliminación	Borrado de fichero.
descompresión	Proceso de revertir los efectos de la compresión.
descifre	Proceso de conversión a texto de los datos cifrados.
validación de firma digital	Proceso mediante el cual se determina que tras decodificar una firma digital el valor obtenido corresponde con el valor esperado.
migración	Transformación de un objeto a una nueva versión cuyo formato sea más actual. Se cambia el formato de fichero o la versión del formato.
normalización	Transformación de un objeto creando una versión más adecuada para la preservación.
replicar duplicar	Proceso de crear una copia idéntica al original.

Se debe crear asimismo un registro de datos por cada entidad Agente de PREMIS que pueda participar en procesos del repositorio. El registro debe hacerse con una estructura de datos compatible con PREMIS para que, del mismo modo, pueda generarse una salida automatizada de datos de agente de acuerdo con el formato

PREMIS en XML para insertarse en el fichero METS, como vimos en el ejemplo de más arriba.

Es conveniente que se use también un listado de valores admitidos para los campos de esa tabla. Los datos de PREMIS para los agentes son: tipo de identificador de agente, valor del identificador de agente, nombre del agente, tipo de agente. Podemos acceder a una versión en español de sistema de metadatos PREMIS en la Web de la Biblioteca Nacional[34].

Los valores admitidos para los tipos de agente, deben ser necesariamente los que enumeramos a continuación para el cumplimiento de PREMIS: persona, organización, software.

Todos los listados de valores admitidos de evento y de agente deberán ser actualizados cada vez que se presente un caso no contemplado en ellos, pero sin alterar los listados normativos del propio PREMIS.

Esta forma de registro de datos de eventos y agentes puede parecer compleja, pero sólo así podrán generarse automáticamente los metadatos de preservación de acuerdo al modelo PREMIS sobre XML y METS, que parece ser el requisito que pueden imponer muchos de los remitentes del sistema de preservación.

Por cada procesado de los enumerados en los epígrafes anteriores debe registrarse el dato sobre si han sido exitosos o, al contrario, no se han finalizado por haberse detectado un error.

El sistema de gestión de repositorio deberá dar la posibilidad de obtener listados de todos los procesos aplicados a los AIP a partir de los datos almacenados, indicando si fueron o no exitosos. Estos listados podrán ser entregados al remitente si así lo demanda.

En caso de errores en el procesado de un AIP, no podrá darse por válido el AIP, debiéndose analizar los errores y repetir el proceso hasta que se consiga la validación completa de todos ellos.

En los valores de los atributos o elementos PREMIS que permitan normalización, como por ejemplo, los tipos de evento, se usará el idioma inglés.

Los datos se captura se obtendrán automáticamente del fichero finalizado con la cadena de caracteres "_mdcapt" que, de acuerdo a la normativa de digitalización, debe ubicarse dentro de la carpeta donde se ubica cada fichero máster en los discos de entrega. Recordamos que ese fichero contiene tres datos separados por un carácter retorno de carro: el modelo de dispositivo usado para la captura expresado de la forma más completa posible pero sin usar ningún carácter de puntuación (ni coma, ni punto, ni punto y coma ni dos puntos o cualquier otro) para separar las palabras usadas; el nombre y versión de la aplicación o aplicaciones informáticas utilizadas para obtener, comprimir (en su caso) y dar formato de fichero al fichero máster , separadas éstas por coma, cuando haya dos o más (no usarán otros signos de puntuación entre las palabras de un nombre de aplicación, salvo el punto para expresar la versión exacta de ésta), aunque si ésta

[34] Accesible desde http://www.bne.es/es/Micrositios/Guias/DiccionarioPremis/resources/images/docs/PREMIS_es.pdf

última coincide únicamente con el controlador del escáner o la aplicación usada para hacer el revelado RAW, se hará constar sólo el nombre y versión de esta aplicación; el nombre y apellidos de la persona que realizó su captura; la empresa o institución responsable del trabajo de digitalización. Esta información podrá ser usada posteriormente obtener con facilidad los datos de evento de captura que se requieren en preservación digital y que no siempre aparecen incrustados en las cabeceras de los ficheros máster. Un ejemplo sería:

Escáner Epson Pefection 3170 Photo

Epson Scan 1.10, Adobe Photoshop CS3

José Martínez Pérez

Universidad Carlos III de Madrid

9.9.3 Procesado de los paquetes SIP para su conversión a paquetes AIP.

9.9.3.1 Verificación de corrección de sistema de empaquetado y estructura de carpetas del SIP.

Se trata de verificar en el momento de la ingesta del AIP al repositorio para los AIP que se cumple la normativa para los SIP en lo que respecta a estas dos cuestiones.

9.9.3.2 Normalización de carpetas y su estructura.

Se trata de renombrar las carpetas heredadas de acuerdo a la normativa para AIP, crear las que sean necesarias para cumplir la conformidad con las normas de empaquetamiento AIP y eliminar las carpetas vacías o que contengan otras carpetas vacías.

9.9.3.3 Asignación de UUID al AIP.

Se trata de conseguir el código UUID para aplicar a la carpeta padre del AIP.

9.9.3.4 Renombrado de la carpeta SIP como carpeta AIP.

Se trata de renombrar la carpeta padre del SIP incluyendo el código UUID obtenido anteriormente, tal y como se describe en las normas de empaquetamiento y representación de los AIP.

9.9.3.5 Inclusión de todos los contenidos del AIP.

Objetos y metadatos estructurados convenientemente de acuerdo a la normativa de empaquetamiento del AIP.

9.9.3.6 Validación AIP recién creado.

Incluye la validación de completitud y corrección de acuerdo a la normativa AIP. Es vital comprobar que no se ha perdido ninguno de los ficheros del SIP en el proceso de creación del AIP.

9.9.3.7 Proceso a seguir con los AIP erróneos.

Se dejarán en una carpeta que identifique su contenido como tal dentro de la carpeta del subrepositorio de remitente, a la espera del análisis del fichero log de errores y de la resolución de los problemas técnicos que impidieron la generación del AIP bien formado. Se adjuntará el fichero log de errores, que en su nombre deberá incluir el nombre del AIP erróneo para su mejor identificación.

El encargado del sistema deberá analizar los log de errores y los AIPs afectados para su resolución.

Se deberá registrar la carpeta donde se ubican los AIP defectuosos en el registro de datos de conversión a AIP correspondiente a ese AIP erróneo, y una vez resuelto el problema y consumada la conversión a AIP bien formado el dato de finalización correcta del proceso.

Una vez subsanada la conversión de los AIPs defectuosos, se borrarán de esa carpeta, pero no los ficheros log de errores, ya que estos pueden ser usados con el tiempo para mejorar el rendimiento del sistema. Y en el sistema de gestión se registrarán sólo para ese AIP los datos del procesado exitoso, pudiéndose borrar los anteriores correspondientes al procesado defectuoso, pues la información sobre errores estará almacenada permanentemente en el fichero log de errores.

9.9.3.8 Ingesta del AIP.

Una vez transformado el SIP en AIP y registrados los datos de la transformación en el sistema de gestión del repositorio se ingestará el AIP en el repositorio, en la unidad de disco y carpeta de subrepositorio y sus unidades correspondientes que deberán haber sido creadas anteriormente a este proceso.

9.9.3.9 Registro de datos de ingesta del AIP.

Se registrará la fecha de ingesta del paquete AIP, su identificador (nombre de carpeta padre que incluye el UUID asignado al AIP), la ruta en el sistema de archivos de la unidad donde se almacena físicamente en el sistema de almacenamiento digital, y la salida del proceso de validación de conformidad del paquete AIP con la normativa de empaquetamiento y representación, y en su caso, el enlace al fichero log de errores. Deberá ser posible enlazar estos datos con todos los datos de los procesados aplicados al AIP y a sus contenidos.

Una vez subsanados los errores de un AIP erróneo y hecha de nuevo la ingesta definitiva, los datos del sistema de gestión serán exclusivamente los del proceso exitoso. El fichero log de errores se mantendrá almacenado indefinidamente en la carpeta donde se van ubicando los AIP erróneos y permanecerán allí después de borrados estos tras su subsanación.

9.9.4 Datos a registrar por cada proceso aplicado en el procedimiento de conversión del SIP al AIP.

Para cada uno de los procesos de AIP descritos anteriormente se deberá registrar: nombre del paquete AIP, fecha de aplicación del proceso, su resultado (con dos posibles valores: error, correcto), y en su caso, un enlace al fichero de log de errores detectados.

Cada registro de datos se separará enfila aparte.

Los AIP correctos e ingestados en el sistema de preservación tendrán siempre como dato de resultado el valor correcto. El valor error sólo lo tendrán los AIPs que resultaron erróneos en su procesado y están a la espera de ser reprocesados para conseguir un AIP válido.

9.9.5 Creación de fichero log de errores de conversión de SIP a AIP.

En caso de fallo en alguno de los procesos se creará un listado de errores encontrados que se almacenará en un fichero TXT, usándose una fila por error, identificando el nombre del proceso afectado y, si es posible, más datos sobre el error que pueda haber generado la aplicación auxiliar empleada en el proceso.

Ese fichero se almacenará permanentemente en la carpeta creada para el almacenamiento de los AIP defectuosos. Quedará vinculado en el sistema de gestión con los datos de los AIP y no se borrará aunque se borren los AIP erróneos de esa carpeta ya subsanados.

9.10 Actuación en caso de descartes de SIP por problemas técnicos o defectos no detectados anteriormente.

Cuando un SIP sea rechazado de forma definitiva por haberse detectado algún tipo de disconformidad grave, no detectada previamente y achacable al incumplimiento de las pautas de ingreso dadas al remitente, se deberá documentar en el sistema de gestión este hecho, identificando el SIP rechazado y los motivos de rechazo. Es vital hacerlo así, para poder demostrar en cualquier momento que no ha habido pérdida de paquetes en el proceso de conversión de SIP a AIP.

10 Normas específicas para el tratamiento de cada tipo de objeto artístico ya en formato digital.

10.1 Ficheros pertenecientes a documentos textuales administrativos o personales multipágina o página simple.

Todo lo que se precisa para su ingesta, empaquetamiento y gestión de preservación digital ha sido desarrollado en la normativa general, más arriba.

10.2 Ficheros de fotografías.

Todo lo que se precisa para su ingesta, empaquetamiento y gestión de preservación digital ha sido desarrollado en la normativa general, más arriba.

10.3 Ficheros de Vídeo.

Todo lo que se precisa para su ingesta, empaquetamiento y gestión de preservación digital ha sido desarrollado en la normativa general, más arriba.

10.4 Ficheros de Audio.

Todo lo que se precisa para su ingesta, empaquetamiento y gestión de preservación digital ha sido desarrollado en la normativa general, más arriba.

10.5 Ficheros de gráficos no fotográficos.

Todo lo que se precisa para su ingesta, empaquetamiento y gestión de preservación digital ha sido desarrollado en la normativa general, más arriba.

10.6 Ficheros multimedia interactivos.

10.6.1 Opciones de trabajo.

Esta tipología incluye ficheros en formatos multimedia que procuran experiencias interactivas y navegación hipertextual. Es el caso de ficheros en formato Flash o DIR de Adobe o Macromedia Director.

Cuando estos ficheros se vuelven obsoletos, al no poder ser ejecutados en las versiones disponibles de las aplicaciones o plugins de navegador Web específicos para su ejecución o edición que aún funcionan en los sistemas operativos actuales, nos encontramos con un riesgo alto de pérdida de la obra.

En estos ficheros estamos ante una situación en que la migración es poco factible, porque no hay aplicaciones que permitan pasar automáticamente todas las posibilidades de interacción y contenido a un formato no obsoleto. O si se consigue para algunas de sus versiones, no se evita el riesgo de que en poco tiempo vuelvan a quedarse obsoletos, dada el abandono progresivo por parte de la industria del software de este tipo de aplicaciones y formatos hipermedia.

La única opción de preservación digital válida con perspectivas de medio y largo plazo es recrear la obra original en un nuevo fichero escrito con un lenguaje no obsoleto y del que se prevea un largo soporte futuro por parte de la industria del software, tal como HTML5 o un motor de videojuegos como Unity, de manera que se puedan ir migrando en el futuro, sin pérdida alguna de funcionalidad y contenido, las obras a los formatos multimedia interactivos sucesores de los actuales. Ello implica una labor de rescate y de reprogramación informática obra a obra que es costosa, y, por tanto, poco viable económicamente. Mediante estos procedimientos se extraen de los ficheros obsoletos los contenidos audiovisuales, almacenándolos en ficheros individuales, y se crean diagramas de interactividad. Posteriormente se crea una aplicación informática o fichero HTML5 que permita recrear la obra original. La obra original consistiría en un fichero ejecutable más un conjunto de ficheros audiovisuales, o en un único fichero ejecutable [35].

Ante esta situación, nos planteamos como una posible solución alternativa la emulación o la virtualización[36]. Las opciones son:

---

[35] Estas opciones han sido estudiadas y arbitradas por Fred Adams en el curso de sus trabajos para el rescate de sus obras en formatos del software Macromedia y Adobe Director. La descripción de estos procedimientos y soluciones se ha hecho en el documento inédito de este autor titulado *Actualización de las Obras interactivas del MIDE creadas con Macromedia Director*.

[36] La virtualización que aquí proponemos no deja de ser un tipo de emulación, pues la aplicación que genera una máquina virtual emula un hardware concreto sobre el que funciona el sistema operativo sobre el que se necesita trabajar. En el contexto de la Informática la virtualización tiene un mayor alcance, ya que este término se usa para referir la creación a través de un software de una versión virtual de cualquier recurso informático hardware o software, e incluso sesiones de trabajo de usuario o recursos en red. La virtualización también puede ser empleada para poder aprovechar recursos hardware antiguos para la ejecución de aplicaciones que trabajan sólo con sistemas operativos y recursos actuales, emulando, consiguientemente, sistemas no obsoletos dentro de sistemas ya obsoletos.

El interés para la preservación digital de esta tecnología es que la máquina virtual que crea el software de virtualización permite ejecutar en la propia máquina virtual un sistema operativo obsoleto (huésped) sin necesidad de instalarlo en el ordenador anfitrión en el que se ha instalado la máquina virtual. El sistema operativo huésped virtualizado puede manejar los recursos hardware del ordenador anfitrión como si se tratara del entorno hardware obsoleto para el que fue diseñado el sistema operativo huésped. Así podemos instalar y hacer funcionar en ordenadores actuales aplicaciones ya obsoletas que no pueden ser instaladas ni ejecutadas en los sistemas operativos actualmente funcionales.

a) La creación o uso de un emulador de la aplicación informática que lee los ficheros, en este caso, Macromedia Director o Adobe Director. La opción puede ser más costosa que reprogramar en HTML5 o Unity obra por obra, pero más versátil al procurar una aplicación que pueda leer todos los ficheros. El emulador se puede crear en Unity[37].
b) Uso de una aplicación de virtualización de sistemas operativos obsoletos que permita crear máquinas virtuales que trabajen con el sistema operativo obsoleto bajo el que corre la aplicación obsoleta desde un ordenador actual. Tenemos a nuestra disposición potentes generadores y gestores de máquinas virtuales de uso libre, como Oracle VirtualBox[38], que permiten ejecutar virtualmente y de forma simultánea en un mismo ordenador actual diferentes sistemas operativos ya obsoletos y presentes a comienzo de los años 1990. También tenemos aplicaciones especializadas que permiten virtualizar en PCs actuales versiones anteriores de Windows, tal como Windows Virtual PC[39] de Microsoft o el programa de uso libre Dioscuri[40]. Otro ejemplo es el programa libre DosBox, que emula un entorno informático compatible con el sistema operativo MS-DOS, de manera que facilita la ejecución de videojuegos de hace varias décadas sin modificación[41]. La opción de la virtualización tiene la ventaja de que nos da la posibilidad de ejecutar, sin tener que instalar, en cualquier ordenador actual sistemas operativos obsoletos en los que funcionan las aplicaciones que necesitamos ejecutar para poder leer e interoperar con los ficheros interactivos multimedia.

La alternativa de la virtualización de sistemas operativos requiere que el AEMA tenga siempre listas las máquinas virtuales con la aplicación de trabajo con los ficheros obsoletos instalada, de manera que cualquier usuario pueda interaccionar con las obras en modo local cuando lo solicite. También deberá suministrar a las instituciones que demanden una copia legal de la obra debidamente autorizada las instrucciones para la creación de las máquinas virtuales con la aplicación lectora instalada. Las necesidades de cara a la preservación digital con esta alternativa son:

---

[37] Estas opciones han sido estudiadas y arbitradas por Fred Adams en el curso de sus trabajos para el rescate de sus obras en formatos del software Macromedia y Adobe Director, en el mismo documento inédito citado anteriormente.

[38] Disponible para su descarga en https://www.virtualbox.org/ .

[39] Se trata de un producto gratuito para usuarios con Windows y descargable desde https://www.microsoft.com/es-es/download/details.aspx?id=3702 . Su última versión soporta o es compatible con prácticamente todas las versiones obsoletas del sistema operativo Windows.

[40] Podemos obtener más información en http://coptr.digipres.org/Dioscuri y http://dioscuri.sourceforge.net/ .

[41] Descargable desde https://www.dosbox.com/ . Este programa es usado por Internet Archive para la virtualización de juegos obsoletos desde la Web.

- Tener y preservar digitalmente los discos de instalación de los sistemas operativos a ejecutar o, de forma alternativa, imágenes de disco de esos discos de instalación, que deberán ser preservadas. Dada la poca fiabilidad de los soportes ópticos a largo plazo, es muy preferible la opción de preservar digitalmente imágenes de cada uno de los discos de la aplicación, estando el formato de esas imágenes estandarizado. Por ello recomendamos la creación y preservación digital de imágenes de disco en formato ISO. Todavía se pueden descargar desde la Web imágenes de disco con los sistemas operativos de uso más común en las últimas décadas, por lo que de momento la virtualización en ordenadores actuales de sistemas operativos antiguos sobre los que correr las aplicaciones ya obsoletas es una posibilidad muy viable, incluso a medio plazo, para poder ejecutar las obras en formatos multimedia interactivo ya obsoletos[42]. Es importante que, en el caso de aplicaciones que se distribuyen en varios discos, se hagan las imágenes exactas de los discos CD o disquetes de su distribución. Si las imágenes no son de los discos de su distribución, al instalar el instalador irá pidiendo los discos de instalación, uno a uno, por lo que, no podrá ser instalado ni el sistema operativo ni la aplicación si las imágenes no corresponden a cada disco de instalación.
- Tener y preservar digitalmente la aplicación informática que permite trabajar con el fichero con plena funcionalidad. Las recomendaciones para esta tarea son las aportadas en el punto anterior.
- Preservar digitalmente el fichero obsoleto en su estado de obsolescencia. Es algo contradictorio con un sistema de preservación digital mantener información obsoleta, pero como es un requerimiento para la virtualización se debe permitir.

A largo plazo la emulación y la virtualización no se pueden considerar como buenas estrategias de preservación, ya que se tendrá una alta dependencia de la disponibilidad futura de los emuladores y virtualizadores que puedan virtualizar los sistemas operativos obsoletos y de los formatos de las imágenes de disco futuros producto de la migración desde los formatos no obsoletos actuales. El plan de preservación se ve obligado a incluir como formato a preservar el de las imágenes de disco preservadas digitalmente.

Como venimos comentando, el éxito de esta alternativa reside no sólo en la disponibilidad de programas de virtualización de sistemas operativos u ordenadores concretos, sino en los formatos de imágenes de disco. Existen decenas de formatos de imagen de disco propietarios de aplicaciones informáticas o de

[42] Podemos encontrar tutoriales que nos indican como configurar y ejecutar máquinas virtuales con sistemas operativos obsoletos en Youtube, tal como https://www.youtube.com/watch?v=ThXYroFsjuA

empresas concretas[43], pero nos decantamos por el basado en el estándar ISO 9660 para las imágenes de soportes ópticos, o procedentes de otro tipo de soportes de almacenamiento pero a las que se va a dar el formato de soporte óptico. El mencionado estándar desarrolla los requisitos de un formato para el almacenamiento de archivos en soportes de tipo disco compacto. Este formato de imagen de disco se denomina habitualmente como imagen ISO. Su extensión es ".iso".

Una fichero de imagen de disco, ISO o de otro formato, adopta la forma de un fichero informático que contiene una copia exacta de la estructura y contenido de un soporte de almacenamiento completo, aunque también pueden generarse para contener carpetas con ficheros o, simplemente, un conjunto de archivos. Estos archivos pueden ser montados en unidades de disquete, CD o DVD virtuales, de manera que sean tratadas por el sistema operativo como unidades de disco reales[44]. Se distribuyen actualmente muchos programas libres para crear imágenes de cualquier tipo de soporte de almacenamiento[45]. Una de ellas es discwizard[46]. Pero estos programas sólo trabajan con unidades enteras, pues se usan para hacer copias de seguridad mediante imágenes de disco. Hay programas que permiten hacer imágenes de disco, como si fueran CDs o disquetes, de conjuntos de archivos en una carpeta, como UltraISO[47] o PowerISO[48]. No todas las aplicaciones crean imágenes ISO de disquetes, por lo que si la máquina virtual sólo admite imágenes de disquetes deberemos usar una que permita crear este tipo de imágenes, como Floppyimage, WinImage o MagicIso[49]. Los formatos de imagen de disco de disquetes se suelen crear con el formato y extensión ".img" o ".ima".

Se trata de bajar de la Web la aplicación obsoleta que lee el archivo obsoleto y hacer una imagen de disco de todos sus archivos, o varias para sus disquetes o CDs si se suministra con varias carpetas que simulan los soportes originales de instalación. Luego se hace la imagen o imágenes de disco y se llaman desde la

---

[43] Podemos obtener un listado de los más comunes en el Web de FileInfo.com, apar tado Disk Image Files, disponible en https://fileinfo.com/filetypes/disk_image .

[44] Este procedimiento de montaje de un fichero de imagen de disco es sencillo. En las últimas versiones del sistema operativo Windows, consiste en hacer doble clic de ratón sobre el fichero imagen desde el Explorador de archivos. Así se instalan y se ven como si fueran una unidad de CD, con su nombre de unidad determinado. El sistema operativo trabaja con el contenido de la imagen de disco como si fuera una unidad de almacenamiento concreta, pero sólo permite la lectura, no se pueden añadir ficheros. Si lo que deseamos hacer es extraer todos los contenidos de la imagen con una sola operación, podemos abrir los ficheros de imagen con una aplicación de compresión, tal como 7-Zip, Winrar o Winzip. De esta manera podemos extraer el contenido completo del archivo .iso a una carpeta. No siempre es posible montar la imagen de disco de un disquete, por lo que para poder trabajar con sus contenidos hemos de usar una aplicación como 7-Zip para extraerlos a una carpeta desde la que poder trabajar con los ficheros.

[45] Podemos encontrar un directorio de aplicaciones de este tipo en la Web de Hipertextual https://hipertextual.com/2016/12/imagenes-de-disco-windows

[46] Disponible en https://www.seagate.com/es/es/support/downloads/discwizard/

[47] Disponible en https://www.ezbsystems.com/ultraiso/download.htm

[48] Disponible en https://www.poweriso.com/

[49] Disponible en http://www.magiciso.com/tutorials/miso-createfloppyimage.htm

aplicación de virtualización con la máquina virtual del sistema operativo bajo el que corre la aplicación obsoleta funcionando. De esta manera las imágenes de disco se cargarán en el sistema operativo virtualizado y se generarán unidades virtuales con el contenido de las imágenes de disco cargadas. Podemos ir a esas unidades desde el propio sistema operativo virtualizado para poder trabajar con sus ficheros y carpetas.

La potencia de un fichero imagen de disco es que puede ser usado para la distribución rápida de una aplicación, junto a todo su entorno software de trabajo preinstalado, sin necesidad de tener que instalar cada uno de sus componentes. Por ejemplo, el conocido software de preservación digital Archivematica se llegó a distribuir en versiones anteriores bajo la forma de un fichero imagen de disco duro virtual en formato VMDK o OVF[50]. Este fichero contenía ya preinstalado no sólo las aplicaciones Archivematica y Atom, sino el propio sistema operativo Linux Ubuntu, PHP, MySQL y Apache. De esta manera podemos crear una máquina virtual en un ordenador Windows sobre la que ejecutar Archivematica sin tener que hacer ningún tipo de instalación. La máquina virtual trabajará sobre el fichero de imagen de disco duro virtual como si fuera un disco duro con el sistema operativo, aplicaciones y ficheros de datos empaquetados en la imagen de disco.

Un proyecto de interés que igualmente delega la estrategia de preservación de ficheros ejecutables en la virtualización es el proyecto Olive. Este proyecto es fruto de la colaboración entre la Carnegie Mellon university e IBM. Su finalidad es desarrollar una metodología y tecnología para crear máquinas virtuales ejecutables desde la Web o en modo local para poder acceder y trabajar sin pérdida de contenido, apariencia y funcionalidad con contenido ejecutable interactivo. La virtualización a través de la Web se realiza con la tecnología de Internet Suspend/Resume[51], que permite hacer streaming de imágenes de máquinas virtuales almacenadas en un servidor.

Los contenidos ejecutables a preservar y difundir se incluyen en máquinas virtuales completas, que tienen todo el software necesitado para su ejecución correcta. La unidad de preservación no es, por tanto, el contenido a preservar sino la máquina virtual completa que lo contiene.

Los usuarios pueden incluso cambiar el contenido de los objetos a preservar de forma local en sus PCs, pero sin alterar la versión archivada. Ante cada carga de la misma obra, el usuario accederá a su versión con los cambios que realizó anteriormente.

---

[50] VMDK (.vmdk) (Virtual Machine Disk) y Open Virtualization Format (OVF) son formatos abiertos para empaquetar y distribuir software a ejecutar en máquinas virtuales bajo la forma de una imagen de disco. Otros formatos similares son VDI y VHD. Estos formatos pueden contener discos duros virtuales, con la instalación completa de un sistema operativo, aplicaciones concretas que corren sobre él y ficheros de datos a ser usados por las aplicaciones. Las aplicaciones de virtualización suelen crear discos duros virtuales de este tipo con alguno de estos formatos cuando el usuario crea una máquina virtual para un sistema operativo concreto. En esos discos duros virtuales (fichero imagen de disco), que son de lectura y escritura almacenan en instalan el sistema operativo y las aplicaciones de usuario y guardan los ficheros de datos que el usuario vaya creando.

[51] The Internet Suspend/Resume® (ISR) project. Disponible en http://isr.cmu.edu/

El concepto de máquina virtual es de gran interés porque independiza el software del hardware concreto para el que fue diseñado, de esta manera, el contenido interactivo puede ejecutarse en cualquier generación de ordenadores que admita el sistema de máquina virtual preservado.

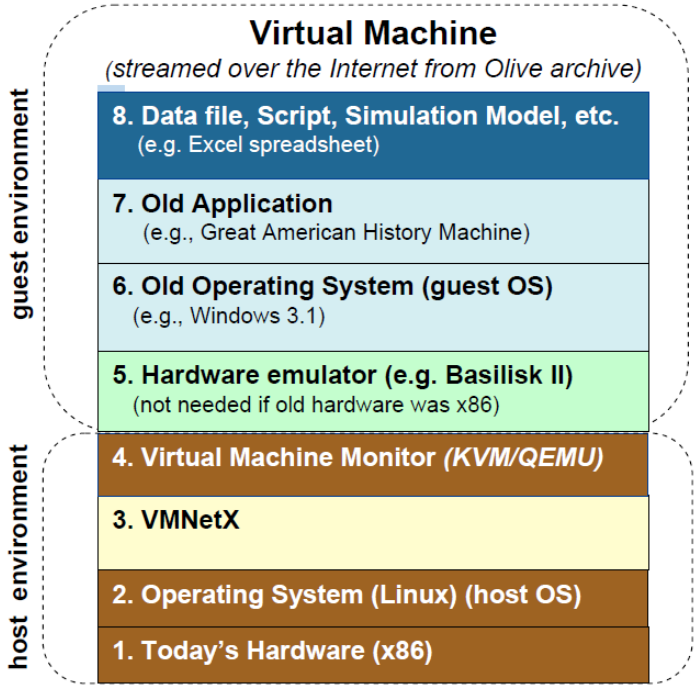
El sistema Olive se basa en un modelo cliente servidor:

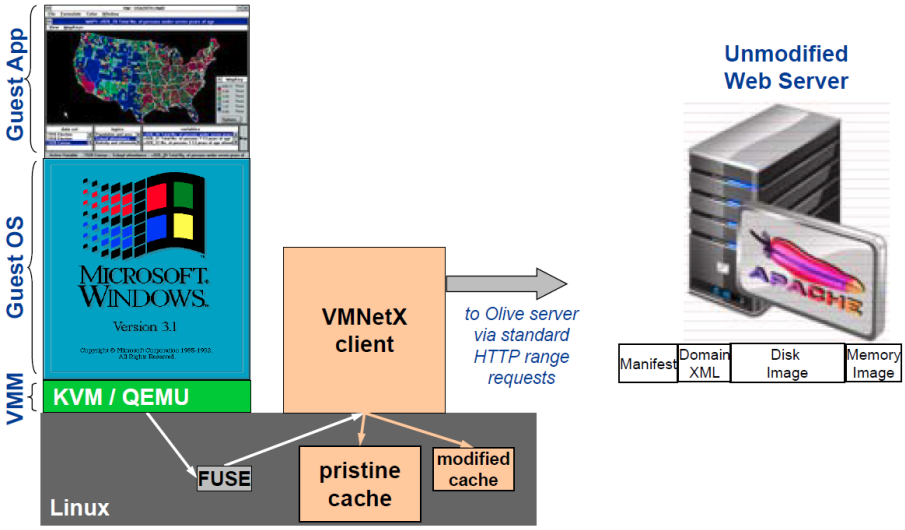
a) Lado cliente. El software cliente es ejecutado en modo local por el usuario que quiere hacer uso de los objetos interactivos preservados, como, por ejemplo, un vídeo juego. Se denomina VMNetX y se distribuye bajo la modalidad GNU General Public License, versión 2. El software correo bajo Linux, pero se distribuye también una versión más ligera para Windows[52]. Ese software descansa en el virtualizador KVM/QEMU. El usuario activa la descarga del objeto con el que quiere interaccionar desde una página Web que dé servicio a la descarga de estos objetos con la tecnología Olive. Para ello usa su navegador Web. Al activar el enlace del objeto se abre la aplicación cliente, que crea una máquina virtual en modo local. La máquina virtual local se conecta a la Web de descarga mediante este software para recibir una imagen de la máquina virtual almacenada mediante streaming. El usuario no debe esperar a que se descargue la máquina virtual completa para empezar a interaccionar con ella en local, gracias al sistema de streaming. La aplicación cliente se basa en el software virtualizador KVM/QEMU, que permite crear máquinas virtuales sobre Linux.
b) Lado servidor. El software servidor se encarga de hacer streaming de las máquinas virtuales archivadas que almacena el servidor a petición de las aplicaciones clientes.

Los siguientes esquemas muestran gráficamente la arquitectura de la máquina virtual que se crea cuando la aplicación cliente carga la imagen de máquina virtual desde el servidor[53].

[52] Las descargas están disponibles desde https://olivearchive.org/docs/vmnetx/install/

[53] Satyanarayanan, M. *et al. One-Click Time Travel,* June 2015, p. 5. Disponible en https://olivearchive.org/static/documents/CMU-CS-15-115.pdf





Las capas de host environment contienen los elementos del ordenador cliente que no son emualdos, esto es, los reales: el hardware del ordenador local, su sistema operativo, el plug-in VMNetX, y el software virtualizador que usa VMNetX (KVM/QEMU). Las capas guest environment contienen aquellos elementos de software que son virtualizados: el fichero del objeto interactivo (por ejemplo, un fichero en formato de la primera versión Macromedia Director), la aplicación que los puede ejecutar correctamente (Old Application, por ejemplo, la primera versión de Macromedia Director) y el sistema operativo virtualizado (por ejemplo,

Windows 3.1). En el caso de que el ordenador antiguo obsoleto virtualizado no tuviera un procesador de arquitectura Intel x86, es preciso incluir un emulador de hardware del procesador obsoleto; sería la capa Hardware Emulator.

Los elementos 5 a 8 son encapsulados en las denominadas *archival VM images* (AVMI). Estas imágenes de disco contienen por tanto: el fichero de la aplicación interactiva a preservar, el sistema operativo obsoleto, la aplicación obsoleta, y el emulador de hardware (si se necesita). El servidor lo que sirve a demanda del cliente es estas imágenes, una por cada objeto interactivo demandado. Además cada paquete de preservación contiene una AVMI además de los metadatos XML que permiten interpretarla.

Un sistema parecido, de tipo cliente/servidor sobre la Web, ha sido implementado como Emulación As a Service[54]. Se trata de un Web que permite a cualquier usuario con un navegador actual y desde su propio ordenador conectado a Internet, emular aplicaciones interactivas obsoletas. Otro sistema similar es el usado por Internet Archive para la distribución de videos juegos antiguos online[55]. Otro proyecto que puede ser inspirador de cara al desarrollo futuro de la estrategia de la emulación en el AEMA es el proyecto KEEP[56], cuyos resultados vieron la luz en el año 2012. Este proyecto ha desarrollado un emulador universal que permite emular un amplio número de arquitecturas hardware (x86, Commodore 64, Amiga, BBC Micro y Amstrad CPC). En realidad, el proyecto no ha desarrollado sus propios emuladores, sino que lo que hace, es utilizar emuladores de uso libre y solventes que ya están en el mercado. El sistema de KEEP se basa en un flujo de trabajo a partir del fichero que aporta el usuario. El propio sistema identifica el formato automáticamente y realiza la selección de los componentes técnicos necesarios para abrir el fichero sin pérdida de funcionalidad, emulando el hardware y sistema operativo bajo el que se creó el fichero y configurando la aplicación que permite ejecutarlo, tal y como vemos en la siguiente figura[57]. Se basa también en el uso de imágenes de disco de soportes de almacenamiento (disquetes, CDs…) que contienen los ficheros con los que se quiere trabajar y las aplicaciones obsoletas que los abren, si éstas últimas no están ya preinstaladas. Por ello se precisa de estos dos elementos para que el sistema funcione, aunque el propio emulador puede traer ya preinstaladas versiones libres de aplicaciones que leen determinados formatos de fichero y da la posibilidad de que los usuarios puedan ir preinstalando aplicaciones a medida que se van necesitando, para evitar la instalación durante el proceso de lectura o trabajo con el fichero obsoleto.

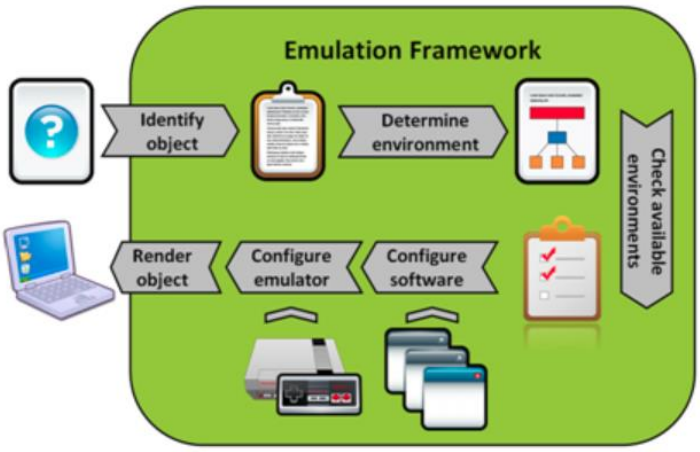
---

[54] Disponible en http://eaas.uni-freiburg.de/

[55] Internet Archive ofrece diferentes tipos de video juegos para su emulación online desde un navegador Web actual, tal como Internet Arcade (https://archive.org/details/internetarcade) o MD-DOS Games (https://archive.org/details/softwarelibrary_msdos_games).

[56] Emulation Framework Project. Su sitio Web es accesible desde http://emuframework.sourceforge.net/ . Se puede descargar libremente el emulador desde este sitio.

[57] Obtenida en http://emuframework.sourceforge.net/about.html



En el contexto de los videojuegos antiguos podemos encontrar múltiples aplicaciones para emular las plataformas más usadas desde los años 1980[58]. El usuario sólo tendrá que descargar el emulador y el ROM del video juego. Un ROM es un archivo digital que contiene los datos que en su día se suministraban a los clientes de videojuego en un cartucho, cinta, disquete o CD de juego. El fichero ROM es el fichero que se tiene que abrir desde el emulador de la plataforma de videojuegos o desde el sistema operativo huésped.

10.6.2 Sistema de empaquetamiento.

10.6.2.1 Arquitectura

La arquitectura de empaquetamiento se basa en encapsular el fichero obsoleto con información e instrucciones necesarios para la emulación y con una o varias imágenes de disco de los discos de instalación del sistema operativo y de la aplicación que puede trabajar con el fichero, tal y como se distribuyeron originalmente. Estas últimas serán imágenes, en su caso, de los discos CD o disquetes de su distribución. Si las imágenes no son de los discos de su distribución, al instalar el instalador irá pidiendo los discos de instalación, uno a uno, por lo que, no podrá ser instalado ni el sistema operativo ni la aplicación si las imágenes no corresponden a cada disco de instalación. Para trabajar desde la máquina virtual, por ejemplo, desde Oracle VirtualBox hay que montar en la unidad virtual de CD o disquetera los discos con los que se va a trabajar, para que el sistema operativo virtualizado muestre el contenido de la imagen como si fuera el de la unidad de almacenamiento con la que se está trabajando. Para ello, se va al menú Dispositivos y se elige el tipo de unidad y desde ahí se llama a la imagen de disco que queremos cargar como si fuera un soporte de almacenamiento en la unidad a: o d:, etc.

[58] Un ejemplo es el conocido emulador Stella, para la consola de video juegos Atari 2600. Se puede descargar desde http://www.emulator-zone.com/doc.php/a2600/stella.html .

La estructura de un paquete de preservación AIP será como describimos a continuación[59]:

- Una **carpeta padre** para la obra interactiva, cuya denominación será el nombre de fichero que tienen los ficheros máster correspondientes en el fondo digital actual, sin incluir la extensión de fichero, más un código UUID separado del anterior por un guion medio.

    - Una **carpeta hija** denominada "data" que contendrá los ficheros que se empaquetan, sus metadatos y los ficheros de control que indique la normativa de empaquetamiento más actualizada, de acuerdo a la siguiente estructura:

        - Una **carpeta hija** denominada "logs_datos_sip" y acorde a la normativa general de los paquetes AIP.
        - Una **carpeta hija** denominada "metadatos_recibidos" y acorde a la normativa general de los paquetes AIP. Al menos deberá contener los metadatos producto de la catalogación bajo la forma de un fichero DC extendido en formato XML y con extensión ".xml".
        - Una **carpeta hija** denominada "objetos" que contendrá:

            - Carpeta "obra". Contendrá el fichero o ficheros ejecutables o en el formato determinado de la aplicación bajo la que se creó, sin ningún tipo de cambio de formato ni transformación, salvo que se añadirá un código UUID a continuación del nombre de fichero o de cada fichero, separado de éste por un guion medio. Dado que las aplicaciones de virtualización pueden llegar a limitar el acceso a los contenidos desde los sistemas operativos huésped como imágenes de disco, se podrá incluir además una imagen de disco como unidad CD o disquete, según corresponda al tamaño y antigüedad de la obra, en formato ISO para discos ópticos o ".img" o ".uma" para disquetes, que llevará su código UUID en el nombre de fichero.

[59] En esta descripción obviamos la explicación y detalle de los elementos que son comunes al resto de tipos de paquetes AIP y que ya hemos descrito más arriba.

- Carpeta “so”. Contendrá las imágenes de los discos de instalación del sistema operativo bajo el que puede ser ejecutada la aplicación que abre el fichero a preservar. El formato de las imágenes será ISO, con extensión “.iso” si son de tipo de soporte óptico, y “.img” o “.ima” si son de tipo disquete. Se añadirá un código UUID a continuación del nombre de fichero o de cada fichero, separado de éste por un guion medio.
- Carpeta “ap”. En caso de que sea necesario, por no tener el fichero de la obra la forma de fichero ejecutable directamente desde el sistema operativo huésped, se creará esta carpeta, que contendrá las imágenes de los discos de instalación de la aplicación que abre el fichero a preservar. El formato de las imágenes será ISO, con extensión “.iso”, para soportes ópticos, e “.img” o “.ima” para disquetes. Se añadirá un código UUID a continuación del nombre de fichero o de cada fichero, separado de éste por un guion medio.

  Podríamos tener el caso de que la institución posee ya un emulador para aplicación o plataforma hardware donde se creó la obra. Por ejemplo, un emulador de una videoconsola concreta de los años 1980[60], y que ese emulador corre sobre los sistemas operativos actuales y no necesita de la emulación de los sistemas operativos obsoletos. En esta situación, se usará esta carpeta (ap) para almacenar el fichero ejecutable del emulador, que deberá llevar su propio código UUID. En la carpeta “documentos” se incluirá en formato PDF/A un documento explicativo que permite trabajar con la obra desde este emulador, así como otro en el que figuren todos los datos técnicos sobre el emulador.
- Carpeta “documentos”. Contendrá, si se dispone de ello y en formato PDF/A:

  - El manual de la aplicación que abre el fichero Las instrucciones que se estimen precisas para la instalación y apertura del fichero desde la aplicación o para la ejecución del programa en que consiste la obra desde el sistema operativo huésped (guest), si se considera que el manual

[60] Como podría ser, por ejemplo, la aplicación de uso libre Stella, para la consola de video juegos Atari 2600.

no es suficientemente explicativo o si no se dispone de manual.
    - De forma obligatoria, un documento informativo sobre:
        - El formato del fichero de la obra y su versión.
        - El nombre y versión de la aplicación que lo reconoce y que está empaquetada en su imagen o imágenes de disco correspondientes.
        - El nombre y versión del sistema operativo que está empaquetado en su imagen o imágenes de disco correspondiente.
        - Cualquier otro dato que se considere relevante para documentar los anteriores.
    - Las instrucciones del propio autor de la obra sobre la forma de exhibición del ordenador desde el que los usuarios acceden a la obra, o cualquier otro aspecto relevante para el trabajo con la obra.
    - Otros documentos que el autor de la obra considere oportuno preservar. Pueden ser documentos explicativos de la obra, su proceso de creación o significado, o de otro tipo, por ejemplo, un vídeo que con la grabación de una sesión de usuario interaccionando con la obra, y que no haya sido considerado previamente como una obra a ser preservada de forma diferenciada, por tener su propio autor y metadatos descriptivos.

Se añadirá a todos los ficheros un código UUID a continuación del nombre de fichero o de cada fichero, separado de éste por un guion medio.

- Carpeta “vm”. Contendrá:
    - Un fichero PDF/A con todas las instrucciones necesarias para generar una máquina virtual que pueda ejecutar el sistema operativo y la obra de acuerdo a la tecnología existente en el

momento de la ingesta de la obra al sistema de preservación.

- Si es necesario, porque la aplicación o sistema operativo a virtualizar hubieran sido creados desde un ordenador con un procesado de arquitectura no x86, en esta carpeta se incluirá una aplicación emuladora de la arquitectura del procesador obsoleto que pueda trabajar en sistemas operativos actuales, y en concreto desde el sistema operativo anfitrión[61]. Cuando se acuda a esta opción se incluirá además un fichero en formato PDF/A que contendrá las explicaciones necesarias para poder trabajar con este software de emulación desde el entorno de virtualización que se haya considerado para esta obra en concreto.
- De forma opcional, y si así se decide por parte de los responsables de preservación digital del AEMA y de forma muy justificada, se almacenará la aplicación de virtualización que permite en la actualidad generar las máquinas virtuales sobre las que poder emular el entorno de la aplicación que abre el fichero a preservar. Está opción no es muy recomendable debido a que la propia aplicación de virtualización se irá quedando obsoleta con el paso del tiempo. También hemos de pensar que este tipo de aplicaciones suele derivar en plazos breves de tiempo en versiones con mejores prestaciones y que corren con las nuevas versiones o los nuevos sistemas operativos. También se ha de considerar en que, dado el alto uso que se hace de esta tecnología, con el tiempo irán surgiendo nuevos productos con más prestaciones, mejores rendimientos y capacidad para enfrentarse a entornos software y hardware obsoletos.

[61] Un ejemplo sería el software emulador Basilisk II, que permite virtualizar sobre Windows y otros sistemas operativos el sistema operativo obsoleto 68k MacOS. Esto permite ejecutar desde sistemas operativos actuales aplicaciones realizadas décadas atrás para este entorno. Pero al ser una aplicación de virtualización, el usuario necesitará disponer de imágenes de disco del sistema operativo MacOS y, por ejemplo, si se trata de emular un videojuego, del ROM para Macintosh. Información y descargas disponibles en https://basilisk.cebix.net/

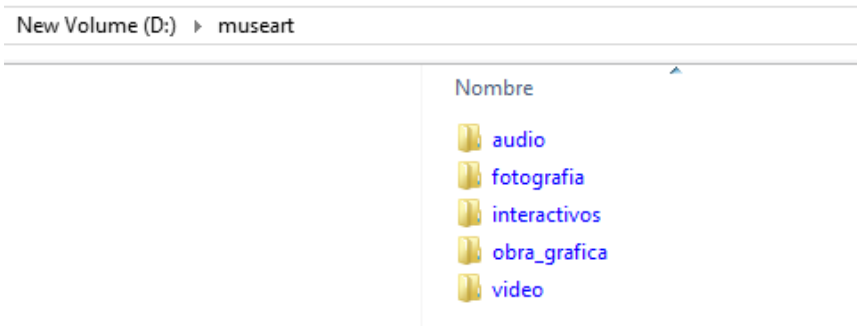
        - Fichero de metadatos en formato METS. Seguirá las recomendaciones generales del AIP.

    - Fichero bag-info.txt. Contendrá la información normativa del estándar BagIT para este fichero.
    - Fichero bagit.txt. Contendrá la información normativa del estándar BagIT para este fichero.
    - Fichero manifest-xxxx. Contendrá la información normativa del estándar BagIT para este fichero. Los x serán sustituidos por el identificador del método hash usado para obtener los códigos hash, usándose tantos caracteres como se necesite.
    - Fichero tagmanifest-xxxx. Contendrá la información normativa del estándar BagIT para este fichero. Los x serán sustituidos por el por el identificador del método hash usado para obtener los códigos hash, usándose tantos caracteres como se necesite.

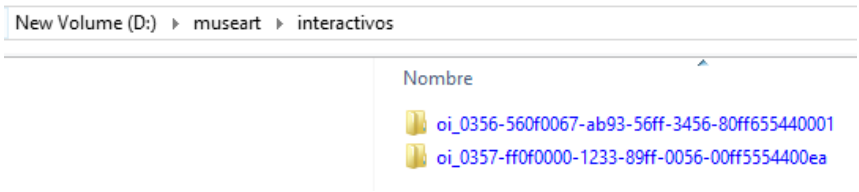
Las aplicaciones a empaquetar deberán ser de uso libre o contar con una licencia de la empresa que las distribuye.

Veamos un ejemplo de navegación por un paquete de este tipo de obras de acuerdo al sistema de empaquetamiento descrito.

Si vamos a la carpeta padre de la institución veremos carpetas hija que contendrán los paquetes de preservación de cada una de las colecciones de la institución, que en este ejemplo hemos nombrado como *museart*.

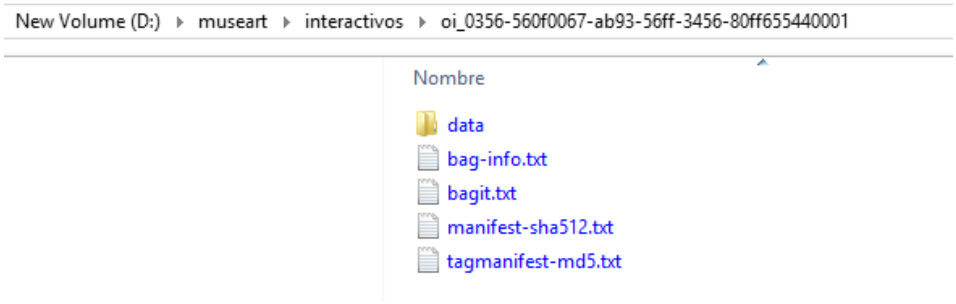


La carpeta interactivos contiene dos obras interactivas que de momento se precisa que sean conservadas en el sistema de preservación digital.

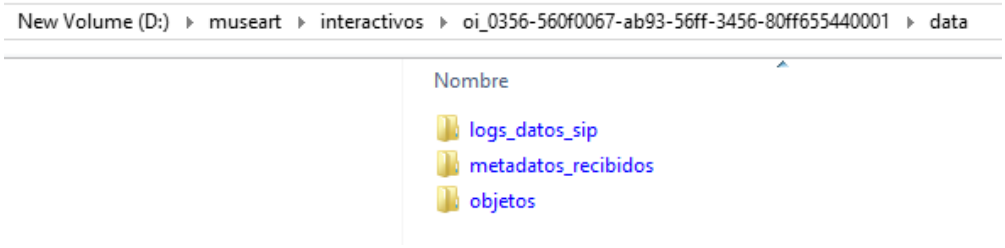


Cada carpeta tiene un nombre acorde a estas especificaciones, tiene su identificador único de objeto de acuerdo a su número de inventario y separado por un guion medio un código UUID.

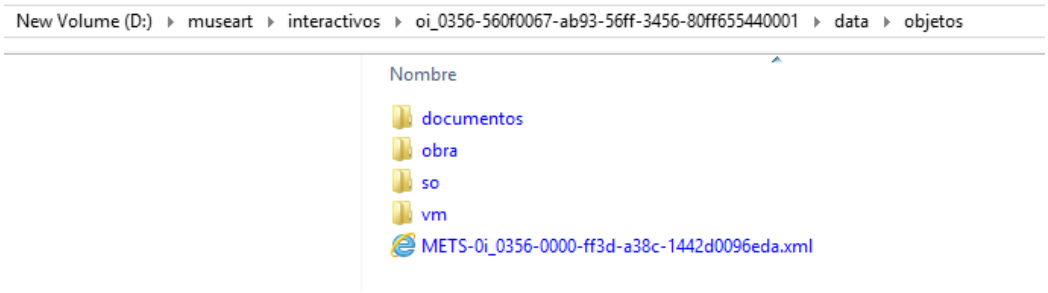
Si accedemos al contenido de la primera carpeta de objeto, vemos la estructura de carpetas que es normativa en un paquete Bagit.



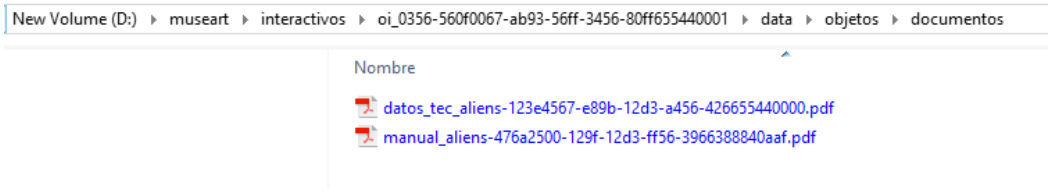
Si accedemos a la carpeta data, vemos la estructura que hemos establecido como obligatoria a nivel general para los paquetes de preservación.



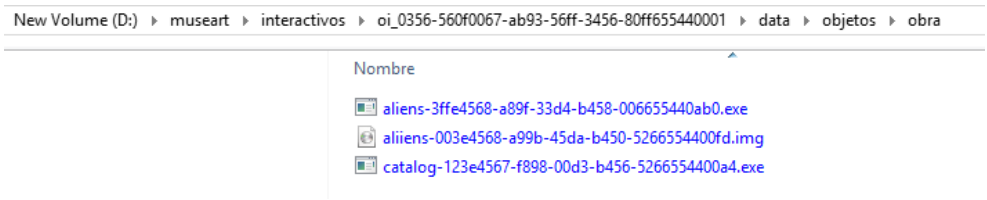
Si accedemos a la carpeta objetos, vemos la estructura que hemos establecido como obligatoria para las obras interactivas.



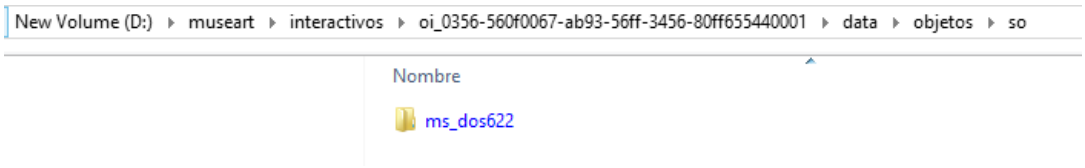
Si accedemos a documentos.



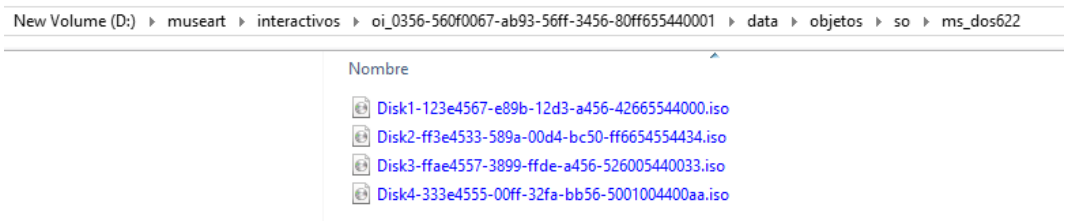
Si accedemos a obra.



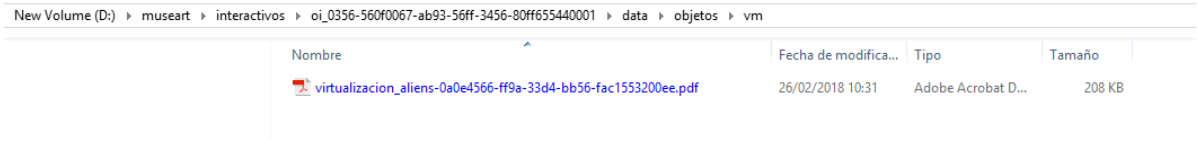
Si accedemos a so, vemos una carpeta que contendrá las imágenes de disco en formato ISO del antiguo sistema operativo MS-DOS en su versión 6.2.2.



Si accedemos a la carpeta del sistema operativo vemos las 4 imágenes de disco usadas para su distribución a través de cuatro disquetes de dos caras y alta densidad.



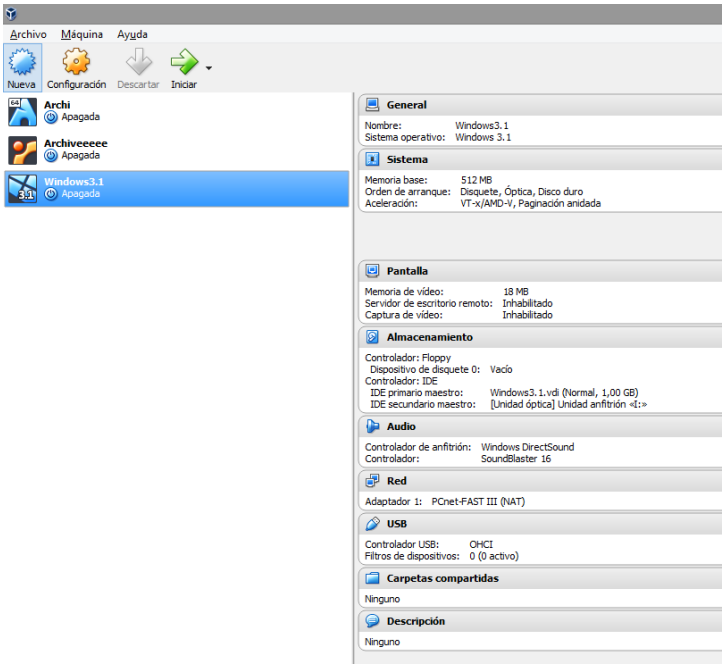
Si accedemos a vm, vemos los ficheros que nos explican cómo virtualizar la obra desde un sistema operativo, actual, en este caso sobre Windows 10 y con la aplicación Oracle Virtual Box.



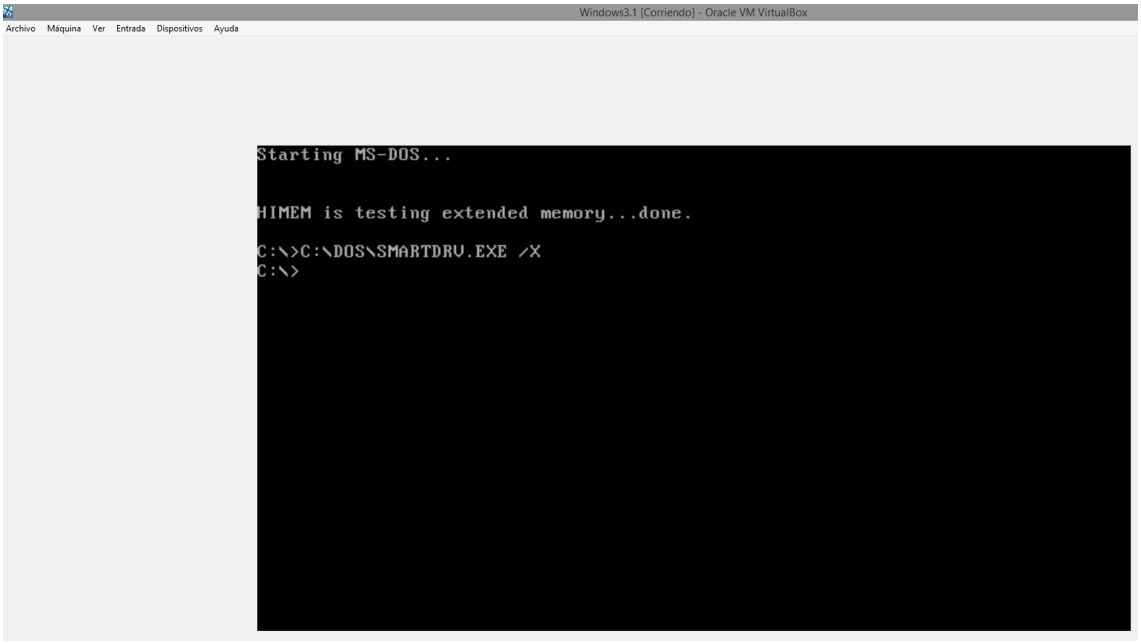
Mostramos a continuación el resultado de virtualizar esta obra con estas herramientas, que adopta la forma de video juego para un PC de finales de la década de 1980.

Iniciamos Virtual Box y creamos una máquina virtual que permita emular el sistema operativo MS-DOS en la versión 6.2.2 y también Windows 3.1. Para ello hemos preinstalado desde la propia máquina virtual estos dos sistemas operativos, a través de sus imágenes de disco. Las imágenes para MS-DOS las hemos obtenido desde el propio paquete AIP de la obra que queremos virtualizar, al estar ahí almacenadas. No sería necesaria la instalación de Windows 3.0, pero lo hemos hecho para aprovechar la misma máquina virtual para poder trabajar con otras aplicaciones sobre Windows 3.1, aprovechando que esta versión de Windows necesita tener preinstalado MS-DOS.

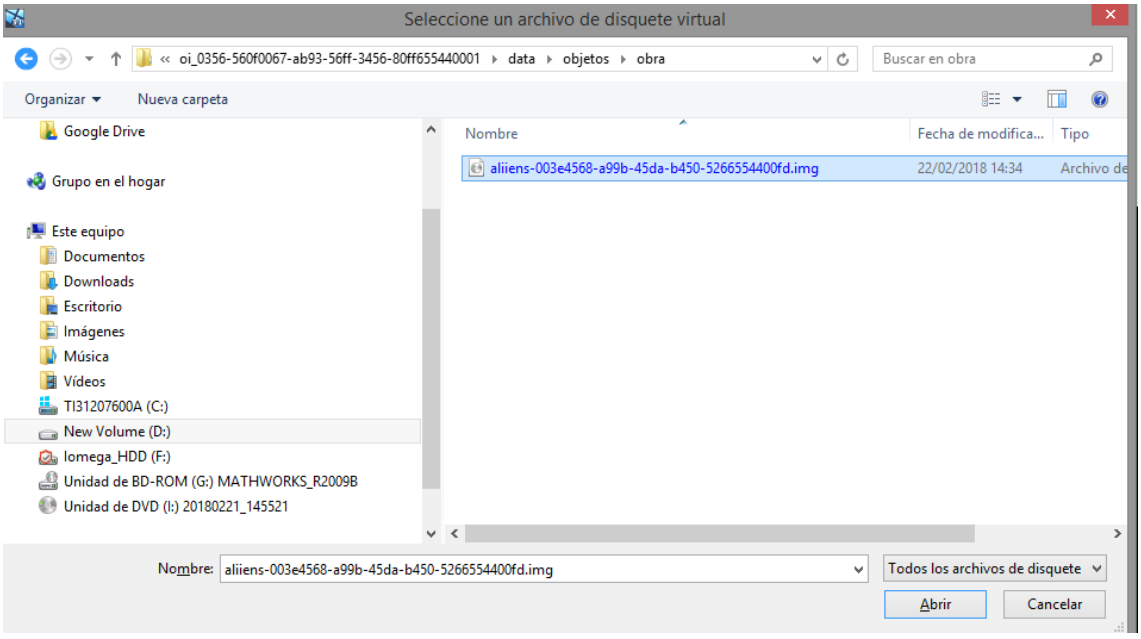
En la siguiente imagen, mostramos la máquina virtual apagada, una vez configurada.



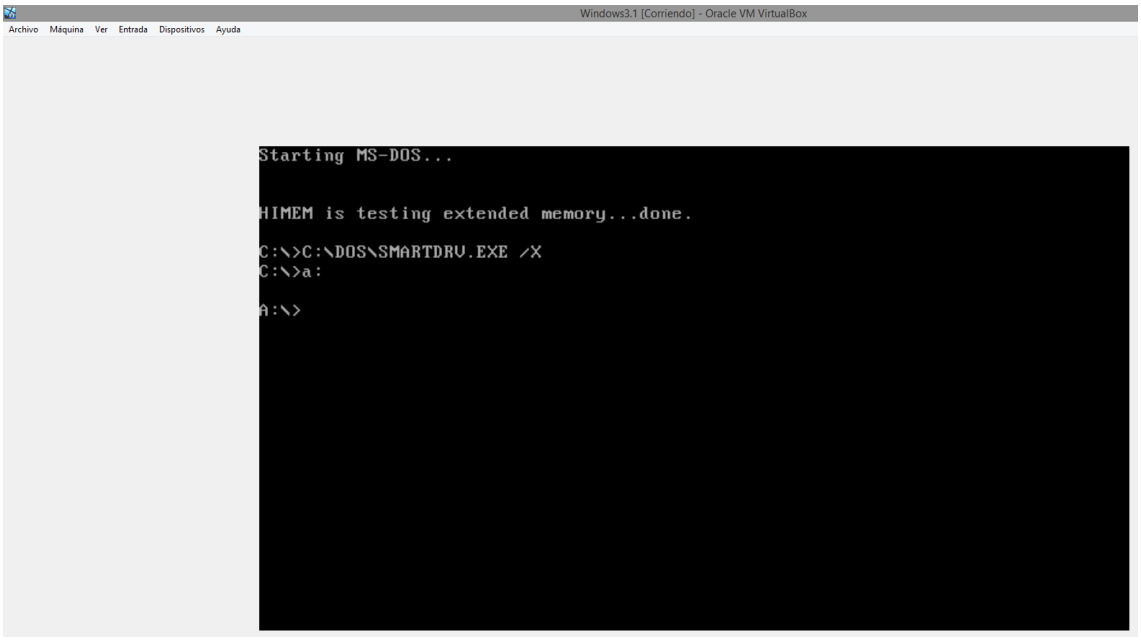
En la siguiente imagen mostramos la máquina virtual emulando el sistema operativo MS-DOS.



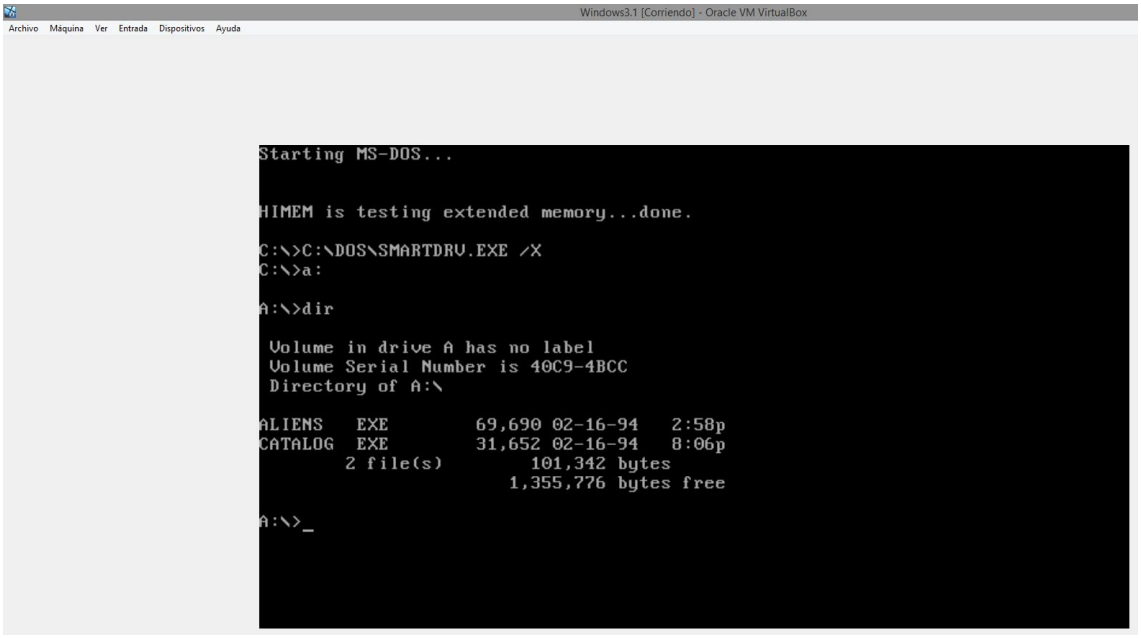
En la siguiente imagen, llamamos a una unidad de disco virtual desde Oracle Virtual Box, que será la imagen del disquete que ha sido incluida en el paquete AIP contiendo la propia obra, que adopta la format de video juego.



Una vez montada la imagen de disco, el virtualizador la reconocerá como disquetera virtual con un disquete que contiene el fichero ejecutable del videojuego. Desde el prompt del MS-DOS cambiamos a la unidad A:.



Hacemos un comando dir para ver el contenido del disquete virtual.

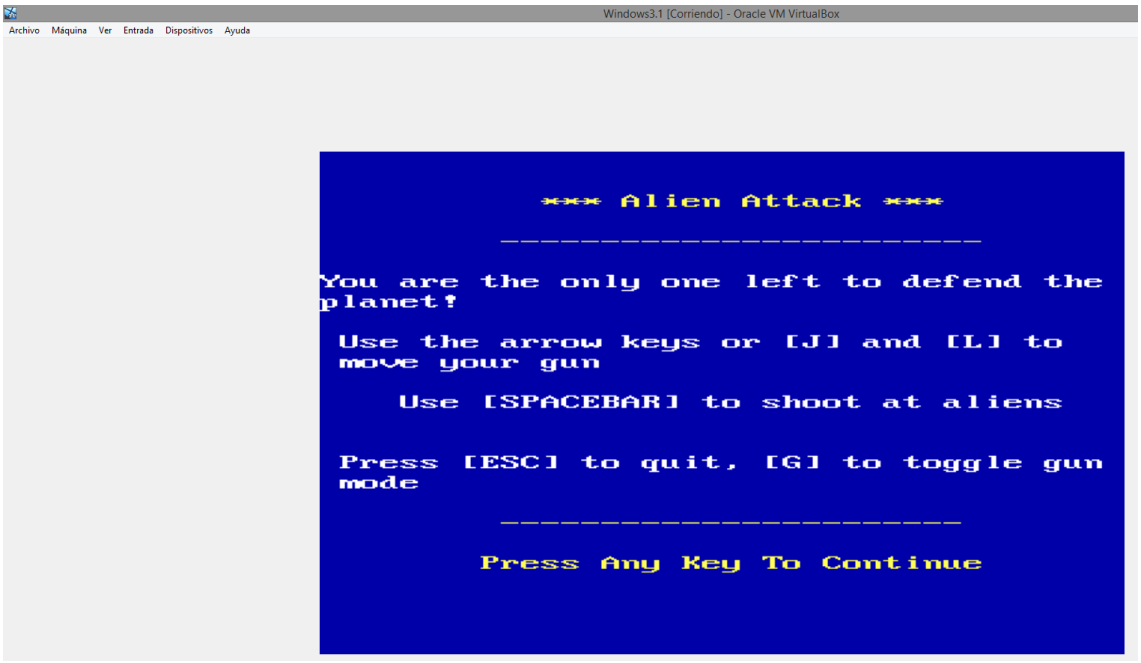


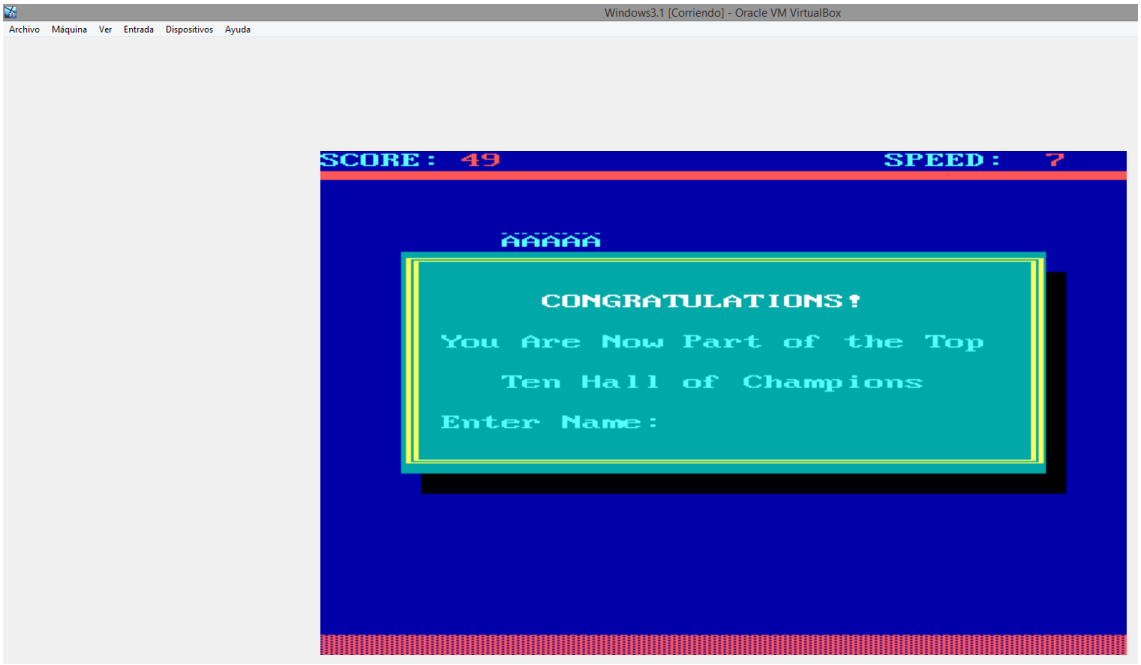
Tenemos dos ejecutables, pero el que nos interesa ahora para interaccionar con el videojuego es “aliens.exe”. Escribimos el nombre del fichero y pulsamos Intro.

```
ALIENS    EXE         69,690 02-16-94    2:58p
CATALOG   EXE         31,652 02-16-94    8:06p
        2 file(s)          101,342 bytes
                         1,355,776 bytes free

A:\>aliens
```

A continuación se ejecuta la aplicación, que nos permite interaccionar con toda su funcionalidad desde nuestro ordenador.





10.6.3 Listado de emuladores de arquitecturas hardware de uso libre que pueden ser incorporados cuando se precise en los paquetes AIP.

- Qemu, 8086+. Es un emulador y virtualizador para arquitecturas x86 de Intel. Puede descargarse desde https://www.qemu.org/ . Tiene versiones open source y de uso libre para Linux, Windows y MacOS.
- Dioscuri, 8086. Es un emulador para arquitecturas x86 de Intel escrito en Java. Permite emulación remota vía VNC. Puede descargarse desde https://sourceforge.net/projects/dioscuri/files/latest/download .
- UAE, Amiga. Permite emular los sistemas de Amiga A500, A500+, A600, A1200, A1000, A3000 y A4000. Puede descargarse desde https://fs-uae.net/.
- BeebEm, BBC Micro. Es un emulador para ordenadores BBC Micro y Master 128. Puede descargarse desde http://www.mkw.me.uk/beebem/ .
- Vice, Commodore. Permite emular una amplia gama de ordenadores Commodore: C64, C64DTV, C128, VIC20, los modelos PET, PLUS4 y CBM-II (aka C610/C510). Puede descargarse desde http://vice-emu.sourceforge.net/ .
- JavaCPC, Amstrad. Emulador de ordenadores Amstrad en Java. Puede descargarse desde https://sourceforge.net/projects/javacpc/ .

10.6.4 El fichero METS de preservación y sus metadatos.

10.6.4.1 Sección Descriptiva (dmdSEC).

Llevará los metadatos descriptivos bibliográficos y los datos custodia de la obra siguiendo la normativa general descrita más arriba.

10.6.4.2 Sección Administrativa (amdSec).

Los metadatos técnicos de objeto Premis serán completos, tal y como se especifican más arriba para los ficheros de documento, y lo más completo posibles para los ficheros de las obras y las imágenes de disco en la medida de lo posible. Para todos ellos existe un código identificador único en Pronom y un tipo MIME.

En los metadatos de procedencia digital (digiprovMD), se incluirá en los eventos el evento de creación de la obra, como tipo de evento creación, y como agente el artista que ha creado la obra.

Será obligatorio incluir en los metadatos Administrativos (en secciones *mets:amdSec*) y en formato Premis XML una serie de elementos que aportan información sobre el entorno informático en el que se pueden hacer funcionar las imágenes de disco de los sistemas operativos y de las aplicaciones que pueden abrir y trabajar con los ficheros preservados en el paquete AIP.

Premis 3.0 exige que los metadatos de entorno técnico de objetos a preservar se expresen como entidades de tipo Objeto y subtipo Entidad Intelectual. Esto es así, porque se sobreentiende que un producto software o hardware es una entidad intelectual que puede tener diferentes manifestaciones físicas (ficheros, máquinas o distribuciones concretas). Ello exige crear varias secciones administrativas <mets:amdSec> adicionales para describir estos nuevos objetos de Entidad Intelectual de entorno. Estas nuevas secciones administrativas contendrán un elemento para metadatos técnicos (<mets:techMD> que a su vez contendrá un elemento <premis:object>. El tipo de Object será intellectualEntity (atributo de Object xsi:type="intellectualEntity" > . Ejemplo de código:

```
<mets:amdSec ID="amdSec-0012d3d7-000c-4b53-a792-dccfe721e827">
  <mets:techMD ID="tech-0012d3d7-000c-4b53-a792-dccfe721e827">
   <mets:mdWrap MDTYPE="PREMIS:OBJECT">
    <mets:xmlData>
     <premis:object xsi:type="IntellectualEntity">
...
     </premis:object>
```

Estos elementos Object se relacionarán con elemento Object del fichero a preservar o entre sí, como veremos más abajo, mediante el elemento <premis:relationship>, que deberá ser incluido en el Object del fichero a preservar o de la Entidad Intelectual para describir el entorno tipo aplicación que abre el fichero a preservar o de su sistema operativo, y que será del tipo *dependency*. El subtipo de esta relación estará en función del tipo de dependencia que haya entre el fichero y el entorno o entre los entornos relacionados, como explicamos a continuación más abajo.

Veamos un ejemplo de relaciones entre un objeto fichero y varios objetos para describir metadatos de entorno (obtenido del documento de la especificación Premis 3.0)[62]:

Vemos el esquema de las relaciones entre ficheros, entorno de tipo aplicación y entorno de tipo sistema operativo o máquina virtual ideado por Premis y que aplicamos en estas especificaciones como modelo. Por tanto, tenemos que establecer en los metadatos dos tipos de relaciones:

---

62 *PREMIS Data Dictionary (full document), Version 3.0*. Disponible en https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf

a) **Relación entre el fichero y la aplicación** bajo la que se puede trabajar con el fichero. Por ejemplo, entre un fichero en formato Macromedia Director y la propia aplicación Macromedia Director.
b) **Relación entre la aplicación y el entorno** bajo el que puede ser ejecutado esta, que puede ser un sistema operativo, un emulador de sistema operativo, una máquina virtual, etc. Por ejemplo, entre Macromedia Director y el sistema operativo Windows 3.11, y entre Windows 3.11 y una aplicación de virtualización concreta, como podría ser Oracle Virtual Box.

Por cada relación (elemento <premis:relationship>) de tipo a), incluida en el elemento Object del fichero a preservar, entre el mismo objeto y un Objeto de entorno del tipo aplicación informática se incluirán los siguientes elementos de metadatos:

relationshipType . Su valor será *dependency* .

relationshipSubType. Su valor será *requires*.

relatedObjectIdentifier

relatedObjectIdentifierType. Su valor será UUID, puesto que se usará este sistema de indentificación.

relatedObjectIdentifierValue. Su valor será el código UUID del objeto Entidad Intelectual de entorno con el que se están estableciendo la relación, que será del tipo aplicación informática.

relatedEnvironmentPurpose. Su valor será *render* .

relatedEnvironmentCharacteristic. Se expresará la necesidad de uso de la aplicación en inglés, usando uno de los valores recomendados por Premis, según corresponda: known to work, minimum, recommended, unspecified[63]. En caso de que necesidad absoluta de esta aplicación se usará el término known to work.

Por cada relación (elemento <premis:relationship>) de tipo b) incluida en el elemento Object de Entidad Intelectual de la aplicación que trabaja con el objeto a preservar y un Objeto de Entidad Intelectual de entorno de tipo sistema, operativo, emulador o máquina virtual, o entre un Objeto de sistema operativo o emulador y máquina virtual, se incluirán los siguientes elementos de metadatos:

relationshipType . Su valor será *dependency* .

relationshipSubType. Su valor será *requires*.

[63] Podemos obtener información sobre el uso de estos valores en http://id.loc.gov/vocabulary/preservation/environmentCharacteristic.html

relatedObjectIdentifier

relatedObjectIdentifierType. Su valor será UUID, puesto que se usará este sistema de indentificación.

relatedObjectIdentifierValue. Su valor será el código UUID del objeto Entidad Intelectual de entorno con el que se están estableciendo la relación.

relatedEnvironmentPurpose. Su valor *run*.

relatedEnvironmentCharacteristic. Se expresará la necesidad de uso de la aplicación en inglés, usando uno de los valores recomendados por Premis, según corresponda: known to work, minimum, recommended, unspecified[64]. En caso de que necesidad absoluta de esta aplicación se usará el término known to work.

En cada elemento nuevo del tipo <premis:object xsi:type="intellectualEntity"> usado para describir un entorno se deberán incluir los siguientes subelementos Premis:

objectIdentifier

objectIdentifierType

objectIdentifierValue

environmentFunction

environmentFunctionType. Se dará siempre el máximo nivel de especificidad, usando la terminología aportada por la propia especificación Premis 3.0 en inglés, según corresponda: software application, software library, software driver, operating system, plugin, hardware architecture, hardware peripheral, chip.

environmentFunctionLevel. Es el nivel dentro de una pila entornos. Para aplicaciones se dará el nivel 1, para sistemas operativos el 2, para emuladores el 3, y para virtualizadores el 4.

environmentDesignation

environmentName. El nombre del producto en su denominación comercial.

environmentVersion. La versión del producto en su denominación comercial, antecedido de la palabra *Version*.

environmentOrigin. La empresa u organización creadora o responsable del producto.

[64] Podemos obtener información sobre el uso de estos valores en http://id.loc.gov/vocabulary/preservation/environmentCharacteristic.html

environmentDesignationNote. Se dará en inglés, cualquier información adicional requerida para mejorar la especificación correcta del entorno. Por ejemplo, que corre sobre procesadores de 32 bits.

environmentDesignationExtension. Si se dispone de esta información, se usará este elemento para aportar cualquier información adicional (como fecha de lanzamiento, fechas de soporte para el software, período de fabricación para hardware, etc.) o una expresión más detallada de la información de la versión (como el número de compilación, el número de serie o las pautas de instalación).

environmentExtension. Se utilizará para aportar características de entorno adicionales no cubiertas por PREMIS. No se usará nunca como reemplazo para las unidades especificadas en PREMIS. Se incluirá en este elemento la siguiente información:

- La fecha en que se han hecho las pruebas de instalación de la aplicación y el montaje de una máquina virtual que permite trabajar correctamente con la obra a preservar.
- Un enlace de descarga (o la carpeta donde está preservada una versión o versiones de la aplicación dentro del AEMA para sistemas operativos de uso común, preferiblemente Windows, Linux y Mac OS)

Además, en cada elemento nuevo del tipo <premis:object xsi:type="intellectualEntity">, cuando corresponda, se incluirá los metadatos de relación especificados más arriba.

En la descripción del entorno figurarán, de acuerdo a la forma de representación explicada en los párrafos anteriores, datos de:

- Una aplicación de virtualización de uso libre (como por ejemplo, Oracle Virtual Box o VMware WorkStation 12 Player) que permita crear las máquinas virtuales sobre las que ejecutar el sistema operativo a emular y la aplicación que permite trabajar con el fichero de la obra a preservar.
- El sistema o sistemas operativos sobre los que hay versión disponible de la aplicación.
- Una aplicación que abre el fichero. Se elegirá la que trabaja con él con menor riesgo de pérdida de contenido o funcionalidades, en el caso de haber varias a disposición del repositorio.
- En caso de necesidad, aplicaciones de emulación de microprocesadores o de sistemas operativos.

La información de entorno quedará obsoleta con el paso del tiempo, al menos lo que atañe a aplicaciones de virtualización para crear máquinas virtuales y de emulación, debido a la aparición periódica que se presumen de nuevas versiones

con más altas prestaciones o incluso al desarrollo de nuevas tecnologías de virtualización o emulación. Por ello, la recomendación es que se actualice al menos cada 5 años, o cuando se presuma que ha habido un cambio en la tecnología de virtualización pensada inicialmente que puede provocar que ésta quede en breve plazo obsoleta. La aparición de nuevos sistemas operativos de uso común, puede ser un buen momento para la revisión del estado de obsolescencia de la tecnología de virtualización registrada en los metadatos de preservación de la obra.

La actualización de esta información requerirá que los responsables de preservación del AEMA realicen pruebas con las nuevas aplicaciones de virtualización que dejan obsoletas a las que se utilizaron para las pruebas anteriores, por lo que se deberá actualizar también la fecha de la realización de las nuevas pruebas.

10.6.4.3 Sección de datos de ficheros (fileSec).

FileSec debe referir a cada uno de los elementos a preservar del paquete AIP y que se ubican en la carpeta objetos: imágenes de disco, objeto interactivo a preservar, ficheros de instrucciones en PDF/A, etc.

10.6.4.4 Sección de mapa estructural (structMap). Normas generales.

El mapa estructural debe ser físico (carpetas) del paquete AIP y referir todos los ficheros a preservar ubicado dentro de la carpeta objetos.

10.6.5 Ejemplo de instrucciones para el desempaquetado y virtualización que permite ejecutar el contenido preservado.

Presentamos a continuación un ejemplo de instrucciones de manejo del paquete AIP para conseguir virtualizar la obra:

1. Instalar el virtualizador, por ejemplo, Oracle Virtual Box.
2. Copiar el paquete de preservación en el disco duro local desde el sistema de preservación.
3. Extraer del paquete de preservación las imágenes virtuales de los discos de instalación del sistema operativo, o llamar posteriormente a esas imágenes desde el mismo paquete una vez en la máquina virtual.
4. Crear una máquina virtual para el sistema operativo obsoleto desde el virtualizador.

5. Instalar dentro de la máquina virtual el sistema operativo obsoleto usando las imágenes de disco.
6. Instalar dentro de la máquina virtual y sistema operativo obsoleto la aplicación obsoleta, si se precisara, usando las imágenes de disco.
7. Ejecutar la aplicación obsoleta y llamar al fichero obsoleto desde la aplicación. O si es un ejecutable interactivo, ejecutar directamente el fichero ejecutable desde el prompt del sistema operativo huésped.

10.7 Ficheros pertenecientes a obras basadas en instalaciones artísticas que incluyen media art.

10.7.1 Alcance de la preservación digital de este tipo de obras.

Hemos de reconocer que ofrecer pautas suficientemente desarrolladas y sistemáticas para la preservación digital y física de este tipo de obras queda fuera del alcance de este proyecto, debido a su complejidad. Estas manifestaciones artísticas suelen presentar varios elementos de naturaleza física junto a medios y componentes electrónicos analógicos o digitales; también suelen constar de unos medios de interacción con los espectadores que procuran experiencias sin las que la obra pierde todo su sentido. Tampoco hemos de perder de vista el impacto del espacio de exhibición o, en su caso, de las experiencias sensoriales que la obra procura a los visitantes y que difícilmente pueden ser vehiculados en los registros gráficos o audiovisuales con los que usualmente se documentan las obras de cara a su custodia museística.

Unos resultados sistemáticos sobre este aspecto requeriría un proyecto de investigación específico, dado el gran tiempo y recursos que se requieren para conseguir unos resultados exhaustivos y generalizables. También se necesita la participación de un equipo multidisciplinar, muy especializado en este tipo de obras, que incluya diferentes perfiles: historiadores del arte, conservadores de museo de arte contemporáneo, expertos en preservación digital, artistas especializados en instalaciones, técnicos informáticos y audiovisuales, etc. En este proyecto lo más que podemos hacer es dar unas pautas básicas para preservar los objetos digitales que nos lleguen de estas instalaciones, por ejemplo, un vídeo que documente la instalación o la interacción de espectadores con la obra o la documentación gráfica y textual que da pautas para su exhibición.

La preservación digital de las instalaciones entra dentro del campo de la conservación de obras de arte, en concreto de la especializada en este tipo de manifestaciones artísticas. Ya ha habido proyectos de investigación internacionales de gran alcance y recursos específicos para abordar este aspecto, como *Inside Installations*[65], *PRACTICs* y *DOCAM*[66]. Tenemos asimismo una amplia

[65] Podemos obtener más información sobre este proyecto en http://www.nimk.nl/eng/inside-installations-preservation-and-presentation-of-installation-art

bibliografía publicada sobre cómo abordar la conservación de este tipo de obras desde una perspectiva amplia, no sólo como estrategia de preservación digital. En ella podemos ver cómo se han llegado a definir al menos 5 estrategias de conservación[67]:

- Emulación. Consiste en imitar la apariencia de la obra mediante medios diferentes. Un caso típico de emulación es la sustitución de un componente de la obra por otro que imita sus características. Por ejemplo, sustituir un aparato de TV obsoleto en una instalación por otro diferente, pero que imita la calidad de imagen y la apariencia del antiguo.
- Migración. Consiste en actualizar el equipo y material fuente. Un ejemplo sería la captura digital de un video analógico.
- Almacenamiento. Consiste en guardar en las mejores condiciones y mantener en el tiempo la funcionalidad de los elementos originales de la obra de arte. Su mayor ventaja es que no se pierden las características intrínsecas de originalidad e integridad de la obra. Pero es obvio que los componentes tecnológicos quedarán obsoletos y no funcionales en un lapso de tiempo determinado.
- Reinterpretación. Se redefine la obra en un medio contemporáneo con el valor metafórico de un medio obsoleto[68].
- Recreación. Se recrea la obra completamente con otros medios, pero respetando la imagen perceptual de la obra original y se usa una tecnología no obsoleta[69]. El problema principal es la pérdida de autenticidad y el riesgo de pérdida de integridad y contenido. Un ejemplo sería reprogramar una obra multimedia interactiva creada con las primeras versiones de Macromedia Director en HTML 5.
- Reconstrucción. Se recrea la obra manteniendo los elementos originales que sea posible.

Hemos de considerar que el cambio de los medios originales empleados por el artista con la intención de poder digitalizar los contenidos y preservarlos digitalmente, puede llegar a alterar la esencia y significado de la obra gravemente.

---

[66] Sitio web disponible en: http://www.docam.ca/ Esta Web presenta mucha información relacionada con la preservación de Media Art.

[67] DOCAM. Conservation Guide. *A Preservation Guide for Technology-Based Artwork*s. Disponible en: http://www.docam.ca/en/conservation-guide.html

[68] GARCÍA MORALES, Lino. *Refectum#1.6. Dé-Coll/age de Wolf Vostel. Un caso de studio*. Disponible en: https://www.teseopress.com/typamuseos/chapter/refectum1-6-tv-de-collage-de-wolf-vostell-un-caso-de-estudio/

[69] GARCÍA MORALES, Lino, *Ibidem*.

Por ello, la digitalización de los sonidos, imágenes fijas o en movimiento empleados por el artista en las instalaciones, o de los usados para documentarlas, y la consiguiente preservación digital de los ficheros digitales resultantes no deben ser entendidas como la preservación de la obra, sino como la preservación de contenidos o evidencias que permitirán a las generaciones actuales y futuras conocer las técnicas y elementos de esa obra, y en algunos casos, permitirán la recreación de la obra mediante el uso de los medios transcodificados digitalmente junto a los medios físicos originales que hayan sido preservados en las instalaciones de la institución depositante.

La misión del sistema de preservación del AEMA, consiguientemente, no es tanto la misión de un conservador de museo, esto es, la conservación de la obra, sino la preservación de las representaciones digitales de los elementos tangibles que han perdurado de la obra, o de documentos que aportan evidencias de la obra, contenidas en los fondos artísticos de las instituciones depositantes del AEMA. En ningún caso se pretende suplir la tarea de la conservación museística de las obras artísticas, sino la de conservar y aportar evidencias digitales de la obra; aquellas evidencias que han podido ser digitalizadas y custodiadas en el AEMA.

Por ello, para la preservación de todos los objetos y documentos de una instalación artística, hay que definir bien cómo es lo tangible en estos proyectos, esto es, los objetos físicos y los datos analógicos y digitales de la instalación que se custodian para poder reconstruir la instalación o saber cómo era:

- Contenidos audiovisuales de la obra. Pueden ser cintas analógicas sonoras o de vídeo del propio autor que aportan contenidos a la obra y , en su caso, sus ficheros digitales resultantes de su digitalización; o ficheros nacidos digitales.
- Fotografías físicas usadas para ser proyectadas o aportar contenido a la instalación.
- Gráficos 2D o 3D digitales que se usan como contenido de la obra.
- Planos de la instalación.
- Textos descriptivos del autor con instrucciones para el montaje o exhibición de la obra, o con otro contenido, como podría ser una explicación redactada por el propio artista sobre el significado de su obra.
- Bocetos o dibujos que representan contenidos de la obra o la disposición de sus elementos.
- Contenidos informáticos ejecutables o interpretables, tales como programas de ordenador, scripts, ficheros creados con lenguajes de marcado, ficheros multimedia interactivos (creados con aplicaciones como, por ejemplo, Flash o Adobe Director), etc.
- Documentación textual o gráfica del propio autor sobre los programas informáticos desarrollados para su obra o de terceros utilizados.
- Documentación audiovisual y gráfica de la instalación o de sus componentes físicos: grabaciones videográficas, sonoras o fotografías que

documentan una experiencia de espectador concreta con la instalación o la propia instalación desde diferentes puntos de vista.

- Entrevistas realizadas por los conservadores de museo al artista con la finalidad de obtener datos para poder documentar y recrear la obra en el futuro.
- Otros documentos que forman el archivo personal del autor de la instalación que tienen que ver con su proceso creativo.
- Documentos externos referidos a la obra. Podrían ser, por ejemplo, recortes de prensa, o publicaciones donde se cita o estudia la obra.

Insistimos en la idea de que un riesgo de la preservación digital de instalaciones que incluyen media art es que la digitalización de los medios y su preservación digital no permitan capturar las características esenciales de la obra. En ese caso, la información digital a preservar tendrá carencias de cara al cumplimiento de su misión esencial para la Historia del Arte, que es mantener evidencias suficientes que permitan conocer e interpretar y dar significado artístico y social a la obra. Hemos de considerar que este tipo de obras es más que la suma de objetos de imagen, sonido o audiovisuales y otros componentes físicos, y que las experiencias de interactividad o sensoriales que procuran son parte esencial de su contenido. Estas obras tienen un significado que puede perderse pese a que se consiga mantener en formato digital los contenidos de los medios o documentación que permite reconstruir los diferentes componentes de la obra. Hay unas características esenciales de la obra que son fundamentales para poder llegar a su significado que deben ser comprendidas y documentadas suficientemente en su conservación museística y, consiguientemente, en su preservación digital[70]. No tenemos solución para este problema, salvo que la información sobre estas características llegue ya en formato digital y suficientemente organizada dentro del conjunto de ficheros digitales sobre la obra, pues, como hemos indicado más arriba, la misión del servicio de preservación digital del AEMA sólo debe ser la de preservación los contenidos digitales que le sean suministrados sobre una obra concreta.

En la bibliografía tenemos algunas aproximaciones al problema de la "esencia" en el caso de la conservación de obras escénicas de cualquier tipo, esto es, a la dificultad de poder conseguir el registro de información esencial de la obra que nos aproxime mejor a la obra en el momento de su exhibición[71]. En estos trabajos se manifiesta la poca representatividad de los registros gráficos, audiovisuales o sonoros que se usan habitualmente para tratar de registrar la totalidad o lo que se

---

[70] Winget, Megan. Digital Preservation of New Media Art Through Exploration of Established Symbolic Representation Systems. Disponible en: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.5394&rep=rep1&type=pdf

[71] ABBOTT, D.; JONES, S.; ROSS, S. *Curating digital records of performance*. 2008 Annual Conference of CIDOC Athens, September 15 - 18, 2008. Disponible en: https://www.researchgate.net/publication/254846847_Curating_Digital_Records_of_Performance ; JONES, S. ; ABBOTT, D.; y ROSS, S. Redefining the performing arts archive. Archival Science, 2009, 8 (3). ISSN 1389-0166. Disponible en: http://eprints.gla.ac.uk/7627/

consideran eventos relevantes de una obra performativa, incluso la distorsión de la obra que estos registros provocan, clamándose por el registro complementario de información que permita a un estudioso acercarse a la esencia de esa obra, como pueden ser las relaciones entre el artista y la audiencia. Pero difícilmente puede captarse en un conjunto limitado de registros y datos el conocimiento subjetivo, experiencial e intangible que conforma la esencia de este tipo de obras. Esto es, la experiencia personal, subjetiva de una performance no puede ser recreada de ninguna manera a partir de los puntuales registros fotográficos, sonoros o audiovisuales que habitualmente se usan para conseguir evidencias documentales que formen parte de los archivos especializados en este tipo de manifestación artística. En las instalaciones artísticas, tenemos también un componente subjetivo experiencial o sensorial difícilmente registrable mediante las representaciones gráficas y audiovisuales que tratan de documentar experiencias de usuario y el montaje de la instalación.

10.7.2 Cuestiones relacionadas con los metadatos.

Hemos de considerar que sólo podemos preservar digitalmente lo que esté en formato digital, independientemente de que en los metadatos descriptivos de la obra se mencionen los elementos físicos u otras informaciones que permitan conocer todos los elementos que componen la obra original. Para facilitar este objetivo, de manera que no se pierda ningún dato del contexto de la obra en los paquetes de preservación digital, se recomienda que en los metadatos, ya sea en un campo de notas u otro, se incluya la siguiente información en el caso de la custodia de ficheros digitales de instalaciones artísticas, o al menos que se refiera que esa información está almacenada en los ficheros digitales que conforman el contenido a preservar digitalmente de la obra[72]:

- Información de los elementos que componen la instalación:
    - Lista de contenidos multimedia (indicando por cada uno de ellos: el formato, cuándo se creó, para qué se creó y para qué se utiliza).
    - Equipos de visualización o de otro tipos usados en la instalación.
    - Manuales de los equipos o instrucciones de manejo.
    - Componentes esculturales.
- Especificaciones de montaje y exhibición, indicando los planos, diagramas o instrucciones del artista.
- Listado histórico de exhibiciones de la obra, especificaciones concretas de cada exhibición y material que las documenta e ilustra.

[72] Información acorde a las recomendaciones de *Matters in Media Art. Documenting Media Art*, 2015. Disponible en: http://mattersinmediaart.org/assessing-time-based-media-art.html . En estas recomendaciones aparecen más elementos de metainformación orientadas a la catalogación completa de las instalaciones artísticas, que pueden ser tomados como modelo si se requiere la catalogación integral de este tipo de obras.

10.7.3 Pautas generales para el empaquetamiento de preservación digital.

Se seguirá la normativa de empaquetamiento de preservación descrita más arriba hasta donde pueda ser aplicada. La normativa se irá adaptando a nivel particular de acuerdo a la naturaleza diferenciada de las ingestas de este tipo de obras al sistema de preservación.

Todos los objetos digitales relativos a una instalación, ya sea nacidos digitales o resultados de la digitalización, que ingresen al repositorio de preservación, deberán tener la consideración de ficheros de una misma obra, requiriéndose que todos estén dentro de una carpeta padre de obra (la asignada a la instalación) y que estén organizados de acuerdo a una estructura que permita identificar su tipología o función dentro de la obra. Dada la amplia variedad de contenidos que podemos llegar a encontrar no normalizamos en estas especificaciones las estructuras ni nombres de carpetas usadas para el empaquetamiento de los objetos digitales en este tipo de obras. Aunque sí que exigimos que ante la presencia de versiones másteres, derivadas o raw de digitalizaciones de contenidos, se use la terminología que más arriba indicamos para el nombrado de las carpetas que contienen los ficheros en los paquetes de preservación, y que haya separación en carpetas diferentes de estas versiones.

Cada objeto en formato digital recibirá el tratamiento de preservación digital normativo que le corresponda de acuerdo a su tipología.

Se exigirá al menos un fichero de metadatos descriptivos de la obra a la que pertenecen los objetos digitales a preservar, que deberá seguir la normativa de metadatos dada más arriba. Por ello, todos los metadatos adicionales que pueden requerir estas obras para el mejor conocimiento de su naturaleza y contexto deberán estar contenidos en el formato Dublin Core extendido y codificados en XML. Los ficheros de metadatos deberán estar en una carpeta cuyo nombre será metadatos_recibidos.

11 Tratamiento de las versiones de una obra dentro del repositorio de preservación.

Es posible que una obra ya ingestada en el respositorio de preservación sea versionada en el futuro, o que en el momento de la ingesta se conserven varias versiones de una misma obra. De acuerdo a la normativa establecida en estas especificaciones, las versiones será tratadas desde la perspectiva de la preservación digital como obras diferentes, correspondiendo a los metadatos descriptivos el registro de los datos que permiten conocer la identificación de la versión concreta de la obra almacenada en el paquete AIP, la fecha de la versión, su motivación, o cualquier otro dato relevante para su interpretación. De esta manera, el repositorio de preservación conservará todas las versiones de una obra disponible, que podrán ser recuperadas y extraídas por los usuarios que deseen tener constancia de la evolución de una obra concreta, mediante las opciones de búsqueda del buscador del repositorio.

12 Tareas de preservación digital de realización continua y periódica.

12.1 Procedimientos de sincronización de copias con separación geográfica.

Cada vez que se añadan contenidos o se efectúan procesos de migración en los ficheros preservados dentro de la unidad de almacenamiento custodiada en la sede donde se ubica el Archivo Español de Media Art, se tiene que sincronizar el sistema de respaldo separado geográficamente, verificándose durante el proceso que las dos versiones son idénticas.

12.2 Controles de integridad.

De forma periódica se harán chequeos de integridad de todos los objetos digitales y ficheros de control y metadatos incluidos en los paquetes AIP a partir de los códigos hash. Este chequeo detecta cualquier problema de corrupción o daño de datos en el sistema de almacenamiento. El propio sistema de preservación decidirá la regularidad de estos chequeos. Ante cada acto de reescritura del fichero en el mismo disco o en otro disco (copiado o reemplazo) se aplica una función hash de verificación.

Los datos de los chequeos de integridad cuando se produce el traspaso de los ficheros a otros soportes, cuando se cambie el sistema de almacenamiento, deberán ser registrados en el sistema de gestión del repositorio y en los metadatos PREMIS empaquetados junto a los contenidos de cada paquete de información.

12.3 Informes periódicos de actividad y estado del sistema de preservación.

Cada vez que haya una modificación o actualización en los contenidos de los paquetes AIP se deberá generar un informe de actividad, donde se reflejarán todos los cambios y su motivación desde la última fecha de actualización: nuevos ingresos, incidencias, actualizaciones de contenidos, procesos de migración, informes de seguimiento tecnológico, mejoras del servicio, etc.

12.4 Alertas de preservación digital.

Se deberá activar un sistema humano de seguimiento de riesgos de obsolescencia, que detectará situaciones problemáticas en este sentido, alertando a los responsables del sistema de preservación de este hecho y haciéndoles llegar una propuesta de aplicación de estrategias de preservación digital.

12.5 Actualización del plan de preservación digital.

El plan de preservación se irá actualizando de manera continua a medida que el seguimiento del entorno tecnológico detecte cambios de criterios por parte de la comunidad de expertos en preservación digital con respecto a lo que puede considerarse como seguro de cara la preservación digital en lo relativo a todos los aspectos técnicos de la confirmación de los objetos digitales y los metadatos custodiados en el repositorio: medios de almacenamiento, formatos de archivo y sus versiones, sistemas de compresión, formatos de codificación, lenguajes y metalenguajes de marcado, esquemas y formas de representación de metadatos, formas de empaquetamiento de objetos a preservar y aspectos a considerar en las buenas prácticas de preservación digital.

12.6 Realización de procesos de migración.

La única estrategia de preservación digital activas que requiere la transformación de los ficheros de objeto digital es la migración. El cambio de soporte o sistema de almacenamiento de los AIP no implica un proceso de migración.

En el modelo de sistema de preservación aquí planteado un proceso de migración supone una transformación del paquete AIP necesariamente, que puede implicar a alguno, a varios o a todos estos elementos: los ficheros de los objetos digitales a preservar, los ficheros de control, los ficheros de metadatos, las nomenclaturas de ficheros o carpetas, la estructura de carpetas, y el sistema de empaquetamiento. Cualquier cambio que sufran los objetos digitales a preservar o el fichero METS de empaquetamiento implicará necesariamente la actualización de los metadatos técnicos y de procedencia digital. Estos últimos registrarán el evento o eventos que se apliquen y su agente,

Cuando un paquete AIP sufre un proceso de migración se genera una nueva versión del AIP no obsoleta, pasando a ser considerada la versión de partida como obsoleta, y por tanto, sin uso ni valor.

Los procesos de migración quedarán descritos con todo el nivel de detalle posible en los planes de preservación digital a aplicar a los contenidos del repositorio. Y su calendario de aplicación, además, en el sistema de gestión del repositorio.

Se recabará la autorización y acuerdo del remitente para aplicar las estrategias de migración, pues el remitente puede querer usar otras estrategias o validar por su parte que la estrategia aplicada le parece correcta y no modifica atributos importantes de la información y el fichero.

Los técnicos del sistema de preservación considerarán también en los planes de preservación digital la necesidad de actualizar las normas de empaquetamiento y de registro de datos de control, resultado de lo cual puede ser la necesidad de migrar los propios paquetes a sistemas de empaquetamiento más acorde con el estado tecnológico del momento, o los propios formatos de los ficheros de control.

En las futuras versiones de los AIP que se tengan que crear debido la aplicación de procesos de migración se irán creando los ficheros siguientes en la carpeta del paquete AIP denominada "logs_datos_sip", que se añadirán a los ficheros de control ya heredados desde la primera versión del AIP, sin que se admita en ningún caso sustitución de este tipo de ficheros:

- Fichero denominado "listadoAIP_Mxx.txt" que contenga un listado de carpetas y sus ficheros contenidos en el AIP obsoleto que ha sido tomado como origen de la migración al AIP migrado, con los datos de nombre, tamaño, fecha y hora de última modificación, tamaño en bytes y la identificación de si es carpeta o fichero. Las xx del nombre de fichero se sustituirán por un número indicativo del ciclo de migración aplicado. A la primera migración le corresponderá el número "01" a la segunda "02", de manera que se pueda identificar en cualquier momento el proceso de migración a que refiere cada fichero. Al comienzo del fichero se deberá describir a modo de comentario mediante un texto muy breve, no más de una línea, que es lo que contiene este fichero, precedido de la cadena "Comentario: ".
- Fichero denominado "tab_corpAIP_Mxx.txt" que contenga una tabla de correspondencia entre los nombres de ficheros y carpetas del AIP origen de la migración correspondiente con los nombres de ficheros y carpetas del AIP migrado, con el siguiente formato: cada fila tendrá los datos de una correspondencia de ficheros o carpetas en la forma *nombre en AIP origen, nombre AIP resultante*. Las filas se separarán por un carácter Intro. Si a una sola carpeta de origen en AIP corresponden varias en el AIP resultante se repetirá la fila tantas veces como carpetas correspondan en el AIP resultante, teniendo la columna para el AIP origen el mismo valor de nombre de carpeta. Si ocurre a la inversa, se repetirán también la fila pero ahora el valor común será para la carpeta AIP resultante. Al comienzo de

este fichero se abrirá una línea extra que contendrá dos elementos, de izquierda a derecha: normativa_AIP, seguido del nombre de fichero identificador único de la normativa de empaquetamiento y representación AIP vigente en el momento de la migración aplicada. Este dato deberá ser registrado asimismo en el registro de datos correspondiente para el AIP resultante en el sistema de gestión del repositorio. Las xx del nombre de fichero se sustituirán por un número indicativo del ciclo de migración aplicado. A la primera migración le corresponderá el número "01" a la segunda "02", de manera que se pueda identificar en cualquier momento el proceso de migración a que refiere cada fichero. Al comienzo del fichero se deberá describir a modo de comentario mediante un texto muy breve, no más de una línea, que es lo que contiene este fichero, precedido de la cadena "Comentario: ".

- Fichero que denominado "sip_estr_crpAIP_Mxx.txt" que contenga la estructura original de carpetas y ficheros del paquete AIP origen de la migración. Las xx del nombre de fichero se sustituirán por un número indicativo del ciclo de migración aplicado. A la primera migración le corresponderá el número "01" a la segunda "02", de manera que se pueda identificar en cualquier momento el proceso de migración a que refiere cada fichero. Al comienzo del fichero se deberá describir a modo de comentario mediante un texto muy breve, no más de una línea, que es lo que contiene este fichero, precedido de la cadena "Comentario: ".
- Fichero "migración_metadatosxx.txt". Sólo se usará en el caso de que se migren los ficheros de metadatos incrustados en el fichero METS de preservación que representa al paquete AIP o el propio fichero METS a una versión METS nueva. Al comienzo del fichero se deberá describir a modo de comentario mediante un texto muy breve, no más de una línea, que es lo que contiene este fichero, precedido de la cadena "Comentario: ". Los caracteres xx se sustituirán por un número indicativo del ciclo de migración aplicado. Su contenido será un texto explicativo que indique los cambios de versión o de esquema de metadatos aplicados a los metadatos bibliográficos (descriptivos) de derechos de propiedad intelectual, de preservación (Premis) y técnicos (Premis u otros). Cuando lo que se migre sea el propio fichero METS se hará consignar así indicando la versión de formato de salida y la del destino.

Todos los ficheros en formato texto y con extensión TXT anteriores deberán llevar al comienzo una línea que indique el sistema de codificación de caracteres aplicado.

Gracias a estos ficheros adicionales de control se podrá reconstruir o descodificar de una manera sencilla, ya sea automática o visualmente, desde cualquier AIP que

haya sufrido distintos procesos de migración cualquiera de las versiones obsoletas del AIP previamente custodiadas en el repositorio.

12.7 Necesidad de procesos de emulación o virtualización de sistemas operativos desfasados para acceso a contenidos obsoletos.

Cuando se estime que se requiere la aplicación de emulación o el uso de máquinas virtuales sobre las que corran sistemas operativos obsoletos como estrategia de preservación y acceso a contenidos o aplicaciones a preservar, no sólo se deberá definir la infraestructura de emulación o virtualización que funcione en la actualidad con los sistemas operativos actuales, sino que hay que preservar, y dar pautas para ello, las propias aplicaciones emuladoras o virtualizadoras y las imágenes de disco de los sistemas operativos y aplicaciones a virtualizar, actualizándolas en el tiempo de manera que se pueda seguir emulando o virtualizando de manera indefinida.

13 Protocolo de actualización de paquetes AIP.

El sistema de preservación deberá permitir la actualización de paquetes AIP, al estimarse que los metadatos descriptivos pueden sufrir variación a lo largo del período de existencia del AEMA. La actualización de metadatos supone:

- La reescritura o sustitución del fichero de metadatos descriptivos almacenado en la carpeta de metadatos recibidos.
- La rescritura de la sección de metadatos descriptivos del fichero METS de preservación.
- La actualización de los ficheros Bagit que refieren los ficheros editados o sustituidos: manifest-xxxxxx.txt y tagmanifest-xxxxx.txt.

Tal y como se obliga más arriba, ante la modificación de los metadatos descriptivos, deberá crearse un evento Premis en el fichero METS de preservación del tipo METADATA_MODIFICATION.

La base de datos del buscador, deberá poderse actualizar periódicamente en correspondencia con la actualización de los metadatos.

14 Especificaciones del buscador del sistema de preservación.

Se deberá crear un buscador para los contenidos preservados, que debe estar instalado en cada unidad de almacenamiento del repositorio de preservación.

El buscador se alimentará de los metadatos en formato DC XML incluidos en los paquetes AIP. Permitirá buscar por los datos de la obra (contenido, autoría, institución depositaria, procedencia, fecha de obra o documento, tipo de medio) y datos de fichero (nombre, formato y fecha).

Los resultados de la búsqueda deben proporcionar acceso a todos los elementos de los AIP.

Desde el buscador se deberá poder hacer una selección sobre los resultados de búsqueda y la descarga de dicha selección a una unidad de disco seleccionada por el usuario.

15 Implementación final del sistema de preservación y manual de usuario y administrador.

Una vez implementada la solución de preservación digital, se describirá en este epígrafe el manual de uso del sistema con pantallazos. El manual contemplará todas las funciones de ingesta, preservación activa, buscador y salida/exportación de ficheros y metadatos.

16 Bibliografía.

ABBOTT, D.; JONES, S.; ROSS, S. *Curating digital records of performance*. 2008 Annual Conference of CIDOC Athens, September 15 - 18, 2008. Disponible en: https://www.researchgate.net/publication/254846847_Curating_Digital_Records_of_Performance

ADAMS, Fred. *Actualización de las Obras interactivas del MIDE creadas con Macromedia Director*. 2017. Indédito.

Digital Preservation Coallition. *Digital Preservation: Sustaining Media Art (A Matters in Media Project)*. 9 September 2016. Disponible en: https://www.dpconline.org/events/digital-preservation-awards/sustaining-media-art

DOCAM. *Conservation Guide. A Preservation Guide for Technology-Based Artworks*. Disponible en: http://www.docam.ca/en/conservation-guide.html

GARCÍA MORALES, Lino. *Refectum#1.6. Dé-Coll/age de Wolf Vostel. Un caso de studio*. Disponible en: https://www.teseopress.com/typamuseos/chapter/refectum1-6-tv-de-collage-de-wolf-vostell-un-caso-de-estudio/

ISO. Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model. ISO 14721:2012. Geneva: ISO, 2012. Equivalente a: CCSDS. Reference Model for an Open Archival Information System (OAIS). Recommended practice CCSDS 650.0-M-2. MAGENTA BOOK. June 2012 [en línea]. Washington, DC: CCSDS, 2012. Disponible en: http://public.ccsds.org/publications/archive/650x0m2.pdf

JONES, S. ; ABBOTT, D.; y ROSS, S. Redefining the performing arts archive. *Archival Science*, 2009, 8 (3). ISSN 1389-0166. Disponible en: http://eprints.gla.ac.uk/7627/

KUNZE, J., et al. The BagIt File Packaging Format (V0.97). Draft-kunze-bagit-14. October 21, 2016. Disponible en: https://tools.ietf.org/html/draft-kunze-bagit-14

REAL, William A. Toward Guidelines for Practice in the Preservation and Documentation of Technology-Based Installation Art. Fall/Winter 2001. *Journal of the American Institute for Conservation*. Vol. 40, no. 3, pp. 207-225. Disponible en: http://www.eai.org/user_files/supporting_documents/william_real.pdf

Subdirección General de Coordinación Bibliotecaria. Ministerio de Educación, Cultura y Deporte. Descripción del Sistema de Preservación a Largo Plazo de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

(SIPREDI_SGCB). 6 de octubre de 2017. Disponible en: http://travesia.mcu.es/portalnb/jspui/handle/10421/9003

Sustaining Media Art. *Matters in Media Art*, 2015. Disponible en: http://mattersinmediaart.org/sustaining-your-collection.html

WINGET, M. Digital Preservation of New Media Art Through Exploration of Established Symbolic Representation Systems. Disponible en: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.5394&rep=rep1&type=pdf