

**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

Subdirección General de Coordinación Bibliotecaria



**Descripción del Sistema de Preservación de las Bibliotecas Digitales de
la Subdirección General de Coordinación Bibliotecaria**

CONTROL DE CAMBIOS

Versión	Fecha	Resumen cambios
1.0	27-02-2017	Versión inicial
1.1	06-10-2017	Inclusión del punto <i>Control Documental de los depósitos PrePIT</i>

Documento elaborado por:

- **Julio Cordal Elviro ; Domingo Arroyo Fernández** (*D.G. de Bellas Artes y Patrimonio Cultural, Ministerio de Educación, Cultura y Deporte*)
- **Jesús Robledano Arillo** (*Universidad Carlos III de Madrid*)

© Ministerio de Educación, Cultura y Deporte, 2017



El presente informe pertenece a la Subdirección General de Coordinación Bibliotecaria del Ministerio de Educación, Cultura y Deporte y está bajo una licencia de **Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional** (http://creativecommons.org/licenses/by-nc-sa/4.0/deed.es_ES) y por ello está permitido copiar, distribuir y comunicar públicamente esta obra bajo las condiciones siguientes:

Reconocimiento — Debe reconocer los créditos de la obra de la manera especificada haciendo referencia expresa a la Subdirección General de Coordinación Bibliotecaria del Ministerio de Educación, Cultura y Deporte y al identificador <http://hdl.handle.net/10421/9003> en <http://travesia.mcu.es>. Dicho reconocimiento no podrá en ningún caso sugerir que el Ministerio de Educación, Cultura y Deporte presta apoyo a dicho tercero o apoya el uso que hace de su obra.

Contenido

Acrónimos	5
Introducción.....	7
1 Aspectos Generales.....	9
1.1 Roles dentro de SIPREDI_SGCB.....	9
1.2 Metadatos implicados	9
2 Contenidos en SIPREDI_SGCB	11
2.1 Contenidos que se incluyen en SIPREDI_SGCB	11
2.1.1 Biblioteca Virtual de Prensa Histórica	11
2.1.2 Biblioteca Virtual de Patrimonio Bibliográfico	11
2.2 Flujo de los datos en el sistema de preservación	12
2.3 Control documental de los depósitos PrePit	13
2.4 Aspectos normativos de los paquetes ingresados.....	14
2.5 Sistema de organización de SIPREDI_SGCB	15
2.6 Sistemas de empaquetamiento de SIPREDI_SGCB	15
2.6.1 Introducción	15
2.6.2 Fichero de comprobación general de todo el depósito	17
2.6.3 Sistema de empaquetamiento a nivel de documento físico individual (número de publicación seriada o volumen)	17
2.6.4 Sistema de empaquetamiento a nivel de publicación periódica o libro presentado en varios volúmenes. 21	
2.7 Tratamiento dado a los metadatos en el paquete de preservación PIA.....	21
2.7.1 Metadatos del PIA, sus esquemas y formas de codificación admitidas.	21
2.7.2 Perfiles METS específicos para los ficheros METS de preservación del paquete PIA.....	25
3 Tareas de preservación digital a realizar en SEPREDI_SGCB.....	51
3.1 Procedimientos de sincronización de copias con separación geográfica	51
3.2 Controles de integridad	51
3.3 Informes periódicos de actividad y estado del sistema de preservación	51
3.4 Alertas de preservación digital	51
3.5 Actualización del plan de preservación digital.....	51

**Descripción del Sistema de Preservación de las Bibliotecas Digitales de
la Subdirección General de Coordinación Bibliotecaria**

3.6	Realización de procesos de migración	52
4	Utilidades de creación de PIC.....	55
4.1	Buscador de contenidos preservados	55
4.2	Restauración de paquetes prePIT	55
5	Equipamiento informático que soporta SIPREDI_SGCB	57
5.1	Sistema de copias con separación geográfica	57
5.2	Características del sistema	57
5.3	Acceso al sistema	57
	ANEXO I Ejemplo del sistema de organización del depósito de preservación de SIPREDI_SGCB	59
	ANEXO II Perfil de preservación digital de SEPREDI_SGCB.....	65
	ANEXO III Buscador de contenidos preservados	69

Acrónimos

SGCB	Subdirección General de Coordinación Bibliotecaria
SIPREDI_SGCB	Sistema de Preservación Digital de la SGCB
OAIS	Sistema Abierto de Información de Archivo
IC	Información de Contenido (<i>Content Information - CI</i>)
IDC	Información de Descripción de Conservación (<i>Preservation Description Information - PDI</i>)
PIA	Paquete de Información de Archivo (<i>Archival Information Package – AIP</i>)
PIC	Paquete de Información de Consulta (<i>Dissemination Information Package – DIP</i>)
PIT	Paquete de Información de Transferencia (<i>Submission Information Package – SIP</i>)

Introducción

Como en cualquier otra biblioteca, los gestores de las bibliotecas digitales tienen la obligación de conservar sus fondos. La garantía del acceso a los contenidos digitales para las generaciones futuras es un criterio fundamental especificado por la *Ley 10/2007, de 22 de junio, de la lectura, del libro y de las bibliotecas*, en su art.4.

La preservación de los contenidos digitales culturales, comparados con la conservación de los documentos plasmados en los soportes tradicionales de papel, es una tarea mucho más compleja, ya que supone el reto de preservar toda su riqueza y su complejidad más allá de la tecnología informática vigente en el momento de su creación.

Este documento describe el sistema de preservación construido en la Subdirección General de Coordinación Bibliotecaria con el fin de preservar a largo plazo los contenidos de la Biblioteca Virtual del Patrimonio Bibliográfico (<http://bvpb.mcu.es>) y de la Biblioteca Virtual de Prensa Histórica (<http://prensahistorica.mcu.es>), resultado de importantes inversiones realizadas desde 2006 por el Ministerio de Educación, Cultura y Deporte para la digitalización de materiales del patrimonio bibliográfico.

Este documento intenta crear un marco lo más estable posible para el desarrollo de la compleja operativa requerida en las tareas de preservación digital. Los riesgos para la preservación digital de nuestros fondos, aparte de falta de recursos humanos y financieros, derivan de aspectos organizativos que no están resueltos y que no son de nuestra competencia. Por una parte existe una falta de planificación en la Administración para la preservación de sus propios contenidos digitales culturales que hace que cada unidad o centro directivo busque su particular forma de solucionar el problema. Por otra parte supone un gran reto adaptar la delicada operativa de preservación a los requisitos de contratación pública, pues éstos conllevan un cambio de roles en las operaciones de preservación que pueden dejar el sistema en manos de diferentes empresas adjudicatarias. Además es previsible que en alguna ocasión pudiera retrasarse o denegarse la contratación del servicio.

Este documento pretende minimizar los riesgos derivados de estos aspectos organizativos, el sistema se ha diseñado de manera que los cambios en los roles afecte lo menos posible al servicio y a los contenidos preservados.

El sistema de preservación está basado en el modelo de referencia OAIS¹, lo que implica que ha implementado la arquitectura, terminología y los procesos propuestos por este modelo.

1 UNE-ISO 14721:2015. Sistemas de transferencia de datos e información espaciales. Sistema abierto de información de archivo (OAIS). Modelo de referencia.

1 Aspectos Generales

1.1 Roles dentro de SIPREDI_SGCB

Se consideran los siguientes roles:

- **Productor**, es el Área dentro de la SGCB responsable de los proyectos de digitalización origen de los Paquetes de Información a preservar.
- **Usuario**, es la propia SGCB, que explotará esos datos dentro de su propia actividad o bien canalizará las peticiones de los interesados en obtener imágenes de alta calidad, volcados de metadatos etc.
- **Dirección**: Dentro de la SGCB, la persona que proporciona la definición y alcance del sistema y dirige los procesos de evaluación del rendimiento.
- **Administrador interno**. Es la figura dentro la SGCB que intermedia con el contratista de servicio de preservación digital.
- **Administrador externo**. Es la mercantil adjudicataria del contrato de servicio de preservación. Se ocupa de realizar todos los trabajos relacionados con los procesos indicados en el apartado 2.3 de este documento.

1.2 Metadatos implicados

Los metadatos implicados son los relativos a los documentos cuyas versiones digitales van a ser depositadas en el repositorio de preservación y a los propios objetos digitales que los representan. Se consideran también los metadatos 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. 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 deben ser empaquetados junto a los objetos digitales.
- Metadatos externos. Son metadatos que se van recogiendo durante los procesos de ingreso, salida y gestión en SEPREDI_SGCB para facilitar la eficiencia y eficacia de todas las operaciones del servicio de repositorio. 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 PIA.

SIPREDI_SGCB utiliza la última versión del estándar de la Library of Congress METS (Metadata Encoding and Transmission Standard) como método de encapsulación para normalizar la representación y vinculación de metadatos, objetos digitales y documentos de procedencia; y PREMIS es el utilizado para normalizar metadatos específicos de preservación digital². Para seguir la mayor conformidad con estos últimos estándares citados empleamos esta clasificación de metadatos implicados:

² Using PREMIS with METS y PREMIS XML Usage Examples, disponibles en: <http://www.loc.gov/standards/premis/premis-mets.html> y <http://www.loc.gov/standards/premis/examples.html> .

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

- **Descriptivos.** Son datos de tipo bibliográfico, que describen física y formalmente los documentos 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...).
- **Administrativos.** Son datos que ayudan al conocimiento técnico y gestión de los objetos digitales.
 - **Técnicos.** Datos sobre las características técnicas de los objetos digitales.
 - **Procedencia digital.** Datos sobre los eventos que ha sufrido un objeto digital desde su creación o captura.
 - **Derechos de propiedad intelectual.** Información sobre los poseedores de los derechos de explotación de las obras y los permisos para su utilización.
 - **Estructurales.** Datos que permiten vincular estructuralmente, ya sea a nivel físico o lógico los objetos digitales.

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

2 Contenidos en SIPREDI_SGCB

SIPREDI_SGCB consiste en una serie de servicios que se aplican al repositorio de preservación. Aquí veremos la perspectiva de los contenidos a preservar, habiendo sido pensada su organización para facilitar las tareas de preservación.

2.1 Contenidos que se incluyen en SIPREDI_SGCB

Actualmente se preservan los contenidos indicados de los siguientes proyectos de digitalización desarrollados en la SGCB:

2.1.1 Biblioteca Virtual de Prensa Histórica

Se preservan los siguientes tipos de ficheros:

- TIFF. Versiones másteres de cada página.
- JPEG. Versiones derivadas de cada página.
- ALTO. Archivos OCR de las páginas.
- PDF. Versiones derivadas multipágina de cada número.
- Carpeta Base. Contiene listados en PDF de contenidos de discos y bases de datos Access sobre las que se generan los listados.
- EDM o ESE. Ficheros de metadatos que se usan para la recolección de registros bibliográficos por parte de Hispana y Europeana. Cada número de prensa o revista tiene un fichero en formato EDM o ESE.
- MARC21. Registros bibliográficos y de ejemplar de cabeceras.
- METS de carga. Ficheros para la carga de contenidos en el sistema de biblioteca digital. METS de preservación de origen. Ficheros para la preservación de contenidos.
- EPUB. Ficheros para eReaders. En algunos soportes con contenidos de la BVPH hay archivos en formato Epub, algunos de estos Epubs, los que indique la SGCB, hay que conservarlos.

2.1.2 Biblioteca Virtual de Patrimonio Bibliográfico

Se preservan los siguientes tipos de ficheros:

- TIFF. Versiones másteres de cada página.
- JPEG. Versiones derivadas de cada página.
- Carpeta Base. Contiene listados en PDF de contenidos de discos y bases de datos Access sobre las que se generan los listados.
- EDM o ESE. Ficheros de metadatos que se usan para la recolección de registros bibliográficos por parte de Hispana y Europeana. Cada número de prensa o revista tiene un fichero en formato EDM o ESE.
- MARC21 Registros bibliográficos y de ejemplar a nivel de obra.
- METS de carga. Ficheros para la carga de contenidos en el sistema de biblioteca digital. Hay un fichero por cada volumen.
- METS de preservación de origen. Ficheros para la preservación de contenidos. Hay un fichero por cada volumen.
- PDF. Versiones derivadas multipágina de cada volumen.

2.2 Flujo de los datos en el sistema de preservación

Los datos se transfieren mediante transmisiones discretas (Sesión de Transferencia de Datos) de Paquetes de Información, llamadas así en OAIS. Los Paquetes de Información se definen como contenedores de Información de Contenido (el contenido original a conservar) e Información de Descripción de Conservación (IDC). En OAIS es obligatorio que un Paquete de Información esté vinculado física y lógicamente a su información normativa. Un Paquete de Información debe contener:

- Información de Contenido (IC). Es aquella información que constituye el objetivo original de la conservación. Se compone de:
 - Objeto de Datos de Contenido (esto es, todos los bits que conforman un fichero digital a conservar).
 - Información de Representación asociada, toda la información que se necesita para que los usuarios puedan comprender el contenido.
- Información de Descripción de Conservación (IDC). Son los metadatos necesarios para la preservación y uso en términos legales de la IC. Deben incluir:
 - Información de Procedencia: que permite documentar la historia de IC. Registrando origen, creador, tecnología y cambios diacrónicos técnicos, de contenido y de custodia.
 - Información de Contexto: describe como se relaciona la IC con otras ICs, por qué se ha creado, etc.
 - Información de Referencia, permite asegurar que la IC es identificable de manera única.
 - Información de Fijeza, permite asegurar su autenticidad e integridad.
 - Información de Derechos de Acceso, la legalidad de cualquier operación de procesado técnico o de contenido requerido para la preservación o difusión, o de uso público o privado.
- Información de Empaquetado. Son los datos que agrupan la IC con la IDC 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. Son metadatos para facilitar la recuperación del objeto y la realización de búsquedas. Por ejemplo, las fichas descriptivas en un catálogo.

Dependiendo de la naturaleza de cada transmisión a SIPREDI_SGCB, cada Paquete de Información puede o bien corresponder a un documento físico entendido como unidad individual (por ejemplo, un volumen de libro, un número de publicación seriada, una fotografía, un mapa...), o bien referirse a una unidad a nivel de obra compuesta por varias unidades individualizables (por ejemplo todos los números de una publicación seriada concreta, todos los volúmenes de una obra o todas las fotografías integradas en un álbum fotográfico).

Se considera el siguiente flujo de los datos en SIPREDI_SGCB.

- **Paquete de Información de Transferencia (PIT).** Es el paquete de información enviado a SIPREDI_SGCB para su preservación. El PIT debe estar normalizado en cuanto a nomenclatura y rutas de los ficheros. Previa a la Sesión de Transferencia de Datos al PIT, se considera una Sesión de Transferencia de Datos PrePIT donde se depositan los ficheros de la SGCB a la espera de comprobación del cumplimiento de la normativa o que están a la espera de que el servicio de repositorio haga la transformación a PIT para que la cumplan y puedan ser ingresados en el repositorio.

- **Paquete de Información de Archivo (PIA)** El paquete ya procesado y almacenado. En la transformación de PIT a PIA, se le se le añade la información necesaria para su gestión y preservación a largo plazo, o sea un conjunto completo de IDC sobre la Información de Contenido. Más adelante se describe la arquitectura de empaquetamiento que incorpora SIPREDI_SGCB.
- **Paquete de Información de Consulta (PIC)** En la actualidad el Sistema de Preservación no sirve objetos a los sistemas de difusión directamente. No obstante SIPREDI_SGCB dispone de un buscador de los contenidos preservados y una utilidad de restauración de los paquetes PrePIT. Estas herramientas identifican los PIA de interés y los asignan en Paquetes de Información de Consulta (PIC). Ambos se describen más adelante en el punto 4.

2.3 Control documental de los depósitos PrePit

La SGCB transferirá “Depósitos PrePIT” al Administrador externo. Estos depósitos consisten en un conjunto de soportes o medios de almacenamiento (discos duros USB, DVDs, pendrives, cintas magnéticas, etc.) con contenidos que se han de transferir o cargar en SIPREDI_SGCB. Una vez realizada la carga, el Administrador externos devolverá los materiales a la SGCB.

Las transferencias de depósitos PrePIT han de estar debidamente controladas y documentadas

En este sentido, cualquier entrega o devolución de materiales ha de documentarse con notas de entrega que incluyan como mínimo la siguiente información:

- Entidad que realiza la entrega y que la recibe
- Identificador del depósito PrePIT. El identificador incluirá la numeración cronológica del depósito.
- Relación de cajas entregadas con indicación de los soportes que incluye cada una de ellas. Cada caja lleva escrito un identificador o bien lleva una etiqueta con su identificador. Cada soporte lleva una etiqueta con su código identificador.
- Descripción general del contenido del depósito PrePIT, por ejemplo: digitalizaciones de prensa histórica del año 2012.
- Lugar y Fecha de entrega.
- El documento incluirá un “Recibí” en donde la persona que recoja los materiales pondrá su firma, nombre y DNI.

Este documento, una vez firmado, será fotocopiado, quedándose en poder del original la persona que entrega los materiales y con la fotocopia la persona que los recoge.

La SGCB entregará al Administrador externo, como material adjunto, todas las listas de contenidos de los soportes que forman parte del depósito PrePIT de las que disponga, así como cualquier otra información que considere que pueda facilitar el trabajo al Administrador externo.

El Administrador externo, o cualquier otra persona o institución que esté a cargo del depósito PrePIT ha de llevar un riguroso control de los materiales que se les han confiado, inventariándolos para poder localizarlos en cualquier momento. Este inventario incluirá datos como:

- Persona o institución que ha entregado el depósito PrePIT.
- Identificador del depósito.
- Fecha de entrada.
- Descripción general del depósito.

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

- Estado del depósito (en espera, en copia, terminado, etc.).
- Localización del depósito y de las cajas y soportes que lo conforman (por ejemplo, mediante signaturas).
- Fecha de salida.

Los materiales han de almacenarse con las condiciones ambientales y de seguridad que garanticen su conservación.

El Servicio de Repositorio realizará automáticamente un listado de contenidos copiados de cada disco duro externo que integre en su sistema de almacenamiento. Estos listados se pasarán a formato PDF y se entregarán junto a los informes de carga a la SGCB.

Habida cuenta de que algunos ingresos supondrán decenas de miles de ficheros, no parece viable que el listado de ficheros entregados e ingresados se haga sobre papel impreso, ya que se pueden superar fácilmente las miles de páginas. Por ello se hará el listado en un fichero PDF multipágina, que deberá estar protegido de modificación.

2.4 Aspectos normativos de los paquetes ingresados

Los soportes de un depósito prePIT contienen paquetes PrePIT, que corresponden cada uno de ellos a una obra digitalizada o digital. Cada paquete contendrá el archivo o los archivos que conforman la obra digital y sus metadatos asociados.

El Administrador externo del sistema ha de comprobar que los paquetes PrePIT cumplan con con la normativa para PIT en la totalidad de todos su ficheros y metadatos. Si no cumplen esa normativa han de ser transformado para que la cumplan y puedan ser ingresados en el sistema de preservación.

Al proceso de comprobación de la normativa PIT, transformación de paquetes PrePIT y transferencia de archivos de los soportes a SIPREDI_SGCB que se realiza sobre los soportes prePIT le llamaremos Sesión de Transferencia de Datos PrePIT.

La Sesión de Transferencia de datos incluye los siguientes trabajos:

- Normalización de nombres de ficheros y carpetas del depósito PrePIA. Con el fin de buscar la máxima compatibilidad con sistemas operativos comunes del empaquetamiento de ficheros de los PIA resultantes, los ficheros y carpetas del depósito PrePIA serán normalizados para eliminar los siguientes problemas:
 - Uso de caracteres no alfanuméricos, a excepción de guion, guion bajo, y un punto para separar el nombre del fichero y la extensión del fichero.
 - Uso de puntos en el nombre del fichero o carpeta, si no son para separar el nombre de la extensión. Nunca deberá usarse el punto como primer carácter del nombre del fichero o carpeta.
 - Uso de caracteres especiales, como: / > < + = ' ^ | \ { } [] # , ; ? ! \$ * &, usualmente prohibidos para los nombres de fichero o carpeta en los sistemas operativos.
 - Uso de espacios en blanco, eñes y tildes.
 - Extensiones de fichero en mayúsculas o no válidas.
 - Nombres de ficheros con más de 128 caracteres, incluyendo el propio nombre, el punto separador de la extensión y la extensión.
 - Rutas de fichero (combinación de nombre de fichero con su extensión y nombre de la vía de acceso) de más de 172 caracteres. Ningún fichero máster o derivado a preservar digitalmente tendrá una longitud de ruta mayor a 172 caracteres.

El Administrador externo debe disponer de una aplicación informática que haga la evaluación de cumplimiento de normativa PIT de nomenclatura y extensión de rutas y el ajuste automático de los nombres de fichero y carpetas y haga 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.

- Además se puede requerir que el Administrador de SIPREDI_SGCB realice tareas asociadas al control de calidad de la transferencia: chequeos de integridad de los ficheros de datos del PIT, documentación de error en la transferencia de ficheros o de lectura/escritura en el soporte, etc.

2.5 Sistema de organización de SIPREDI_SGCB

La organización de SIPREDI_SGCB separa

1. **Los ficheros que contienen los listados** de contenidos de los discos (los ficheros PDF y MDB ubicados dentro de la carpeta “base”) han sido separados de la estructura de organización de las obras a preservar, pues su contenido no se organiza por obras sino por discos entregados en los contratistas del servicio de digitalización. Los ficheros contenidos en las carpetas base se organizan de acuerdo a la siguiente estructura jerárquica: Año de digitalización-Proyecto [BVPB o BVPH] -> Disco (esta carpeta tiene la función de carpeta padre de paquete PIA del listado) -> Carpetas propias del modelo de empaquetamiento, junto los ficheros de control y los ficheros de listados en PDF y Access a preservar.
2. **De los ficheros de los documentos** (imágenes de páginas y metadatos) se organizan de acuerdo a la lógica del documento:
 - Libro -> volumen
 - publicación periódica -> Año -> Mes -> Número

El Anexo I muestra un ejemplo del sistema de organización del depósito de preservación.

2.6 Sistemas de empaquetamiento de SIPREDI_SGCB

2.6.1 Introducción

El sistema de organización de los paquetes PIA de SIPREDI_SGCB se basa en los requerimientos de OAIS y TRAC (*Trustworthy Repositories Audit and Certification Criteria*)³ y el estándar *BagIt*⁴, de la Library of Congress de los Estados Unidos. Es un sistema relativamente sencillo de implementar y de entender a nivel humano y a nivel máquina, lo que

³ 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>)

⁴ <http://www.digitalpreservation.gov/documents/bagitspec.pdf>

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

permite simplificar los procesos humanos y automatizados, no requiere de una aplicación específica para su descodificación ni codificación y cualquier usuario puede tanto empaquetar como desempaquetar los contenidos con las aplicaciones típicas de escritorio disponibles desde cualquier 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).

De acuerdo a *BagIt* se reúnen en un directorio de sistema operativo el objeto u objetos digitales relacionados con un documento u otra unidad documental. Dentro de esta carpeta padre se crean carpetas hijo para poder organizar de una forma más estructurada los ficheros. Además de estas carpetas y los objetos digitales, se 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. De acuerdo a *BagIt* los caracteres de los ficheros TXT se deben codificar en UTF-8 (8-bit Unicode Transformation Format).

Los paquetes no podrán ser serializados en un fichero único (TAR y/o ZIP) en el Repositorio de Preservación.

En un paquete BagIt pueden existir los siguientes contenidos:

- 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-xxxxx.txt”). Contiene una relación de los nombres de fichero incluidos en el Bag precedidos 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.

En el sistema de empaquetamiento PIA se fuerza al uso de códigos UUID⁵. Se ha modificado el estándar UUID para añadir elementos de control que pueden ayudar en su momento a la localización de ficheros perdidos. La implementación del sistema de códigos UUID aplicado en SIPREDI_SGCB es el siguiente. Como el número UUID tiene una gran longitud, se usará parte de él con un código propio que sigue garantizando la independencia y permite facilitar la localización y la búsqueda de ficheros en caso de problema técnico en el sistema de almacenamiento.

El formato definitivo de código UUID es: CCCAAAAA-PPPP-4XXX-XXXX-XXXXXXXXXXXX.

En las C se ubica un código entidad, en las A el código de PIA para esa institución, y las P el número de objeto dentro del PIA, las X serían el código UUID (El 4 es obligatorio para el tipo de UUID aleatorio completo). El fragmento aleatorio del código UUID sigue teniendo 2⁷⁶ posibles UUID, y este orden de magnitud garantiza la unicidad en el sistema de repositorio y en cualquier otro sistema de almacenamiento ajeno en el que se puedan ingresar los paquetes PIA.

La estructura y normativas del paquete PIA se ha pensado para permitir generar automáticamente el PIT original que entrega la SGCB, exactamente con la misma estructura de carpetas, nombres de ficheros y sus formatos. Y al mismo

⁵ Leach, P.; Mealling, M.; Salz, R. (July 2005). A Universally Unique Identifier (UUID) URN Namespace. Internet Engineering Task Force. RFC 4122. <https://tools.ietf.org/html/rfc4122>

tiempo permitir generar una salida del PIA que mantenga la estructura original del PIT pero incluyendo los ficheros no obsoletos tras aplicarles las funciones de preservación digital.

2.6.2 Fichero de comprobación general de todo el depósito

La información de todos los paquetes que forman el depósito se encuentra en la base de datos de la aplicación. Para el caso en que esta base de datos o sus copias pudieran estropearse y para garantizar que la estructura completa pueda ser verificada por cualquier aplicación externa, se ha generado un fichero de texto adicional al que llamaremos **check_aip.txt**.

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.

2.6.3 Sistema de empaquetamiento a nivel de documento físico individual (número de publicación seriada o volumen)

El PIA contiene obligatoriamente y con este sistema de estructuración:

- Una **carpeta padre**, cuya denominación es el nombre de la carpeta usada para reunir todos los ficheros del documento (normalmente, el código del campo 001 del registro MARC correspondiente) más un código UUID separado del anterior por un guion medio. El código UUID se debe calcular automáticamente sobre la marcha en el momento de la conversión del PIT al PIA. El nombre aportado por la SGCB 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 carpeta contiene todos los ficheros de un documento junto a sus metadatos.
 - Una **carpeta hija** denominada “data” que contiene 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 contiene los ficheros que incluyen los datos del paquete PIT original y sus ficheros. La finalidad de esta información es permitir un proceso de reconstrucción íntegra y casi exacta del PIT a partir del PIA, así como documentar perfectamente el PIT en el PIA. Esos ficheros y sus contenidos son:
 - Fichero denominado “listado.txt” que incluye un listado de carpetas y sus ficheros contenidos en el paquete PIT correspondiente al PIA con los datos de nombre, tamaño, fecha y hora de última modificación, tamaño en bytes y la identificación de si es carpeta o fichero. Al comienzo del fichero se describe, a modo de comentario, mediante un texto muy breve que es lo que contiene este fichero, precedido del carácter #. La línea de comentario finaliza 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 PIT no coincidan con los del PrePIT, se usarán los nombres del PrePIT para garantizar que se registran los nombres de fichero y carpeta tal y como son aportados por la SGCB, sin la normalización de nomenclatura de la fase PIT. En

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

el caso de que el paquete PIT 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 contiene una tabla de correspondencia entre los nombres de ficheros y carpetas del PIT correspondiente con los nombres de ficheros y carpetas del PIA, con el siguiente formato: cada fila tiene los datos de una correspondencia de ficheros o carpetas en la forma *nombre en PIT, nombre PIA*. Las filas se separan por un carácter Intro. Se usan las rutas completas de los ficheros. Si a una sola carpeta de origen en PIT corresponden varias en el PIA se repite la fila tantas veces como carpetas correspondan en el PIA, teniendo la columna para el PIT el mismo valor de nombre de carpeta PIT. Si ocurre a la inversa, se repite también la fila, pero ahora el valor común será para la carpeta PIA. Al comienzo de este fichero se abre una línea extra que contiene dos elementos, de izquierda a derecha: normativa_PIA, seguido del nombre de fichero identificador único de la normativa de empaquetamiento y representación PIA aplicada. Al comienzo del fichero se describe 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 finaliza con un carácter Intro (Retorno de carro y salto de línea). En el caso de que los nombres de ficheros y carpetas del paquete PIT no coincidan con los del PrePIT, se usan los nombres del PrePIT para garantizar que se registran los nombres de fichero y carpeta tal y como son aportados por la SGCB, sin la normalización de nomenclatura de la fase PIT.
- Fichero denominado “sip_estr_crp.txt” que contiene la estructura original de carpetas y ficheros del paquete PIT. 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
```

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

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 describe 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 finaliza obligatoriamente con un carácter Intro (Retorno de carro y salto de línea).

Se añade 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 PIT no coincidan con los del PrePIT, se usarán los nombres del PrePIT para garantizar que se registran los nombres de fichero y carpeta tal y como son aportados por la SGCB, sin la normalización de nomenclatura de la fase PIT.

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

- Fichero denominado “*Id_form_fich.txt*” que contiene el resultado de la identificación del formato de los ficheros del PIT correspondiente, con el formato: nombre de fichero, nombre del formato y nombre de versión del formato; pudiéndose añadir también si no constan estos datos en los metadatos PREMIS *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 PIT. Se usa como separador de fila el carácter Intro. Al comienzo del fichero se describe 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 finaliza con un carácter Intro (Retorno de carro y salto de línea). En el caso de que los nombres de ficheros y carpetas del paquete PIT no coincidan con los del PrePIT, se usan los nombres del PrePIT para garantizar que se registran los nombres de fichero y carpeta tal y como son aportados por la SGCB, sin la normalización de nomenclatura de la fase PIT. 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 se hace a través del sistema de registro PRONOM⁶.
- Una **carpeta hija** denominada “*metadatos_recibidos*” que contiene los ficheros de metadatos correspondientes al documento al que se dedica el paquete PIA recibidos desde la SGCB. Se incluye también el fichero XML “*datos_custodia.xml*” que puede crear el Administrador interno o externo del SEPREDI_SGCB siguiendo la normativa establecida en

⁶ <http://apps.nationalarchives.gov.uk/PRONOM/Default.aspx>

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

este mismo documento. Esta carpeta se ha estructurado a su vez en subcarpetas indicativas del formato de metadatos que contiene.

- Una **carpeta hija** denominada “objetos” que contiene los ficheros del documento empaquetado con la estructura que se indica en los siguientes párrafos, considerando que si no existe algún tipo de objeto no se creará la carpeta 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 al que se dedica el paquete PIA. El contenido de esta carpeta es:
 - Carpeta “maestros”
 - Ficheros maestros. Se respeta el nombre de los maestros aportados por la SGCB pero ajustados a la normativa de nomenclatura de los PIA. Cada fichero maestro lleva un código UUID a continuación del nombre del maestro, separado de éste por un guion medio. Dentro de esta carpeta se podrán crear carpetas específicas para contener elementos adicionales de los ficheros maestros, como pueden ser ficheros de cartas de control de color o tono y sus ficheros con datos de referencia.
 - Carpeta “derivados”
 - Ficheros derivados. Se respeta el nombre de los derivados aportados por la SGCB pero ajustados a la normativa de nomenclatura para derivados de los PIT. Los derivados proporcionados por la SGCB llevan su propio código UUID si son generados por la SGCB y no por el servicio de repositorio a petición de ésta. Su nombre, a excepción del código UUID, debe ser igual al de los maestros a que corresponden. Dentro de esta carpeta se crean tantas carpetas hijo como formatos de derivados haya que conservar: alto, jpeg, pdf, epub...
- Fichero de metadatos en METS. Su contenido es explicado con detalle en el epígrafe dedicado a metadatos, más abajo. Este fichero contiene codificados en METS todos los metadatos en toda la tipología exigida por la normativa. Sólo contiene metadatos relativos al documento al que corresponde el paquete PIA y a los ficheros digitales que lo conforman y están, por tanto, almacenados en la carpeta “objetos”. Su extensión es “xml” y su nombre coincide con el nombre de la carpeta padre del PIA con la cadena “mets-” justo delante. El nombre incluye también obligatoriamente el código UUID de la carpeta padre. Este fichero METS no incluye información sobre ficheros de control o BagIT. 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 PIA.
 - Fichero bag-info.txt. Contiene la información normativa del estándar BagIT para este fichero.
 - Fichero bagit.txt. Contiene la información normativa del estándar BagIT para este fichero.

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

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

2.6.4 Sistema de empaquetamiento a nivel de publicación periódica o libro presentado en varios volúmenes.

En este caso, y para un mayor control, se crearán al nivel de la carpeta padre contenedora de la publicación periódica o del libro presentado en varios volúmenes los siguientes elementos:

- Los ficheros preceptivos del estándar Bagit: bag-info.txt, bagit.txt, manifest-md5.txt (conteniendo una línea sólo por cada fichero contenido en la carpeta DATA a su mismo nivel) y tagmanifest-md5.txt. El contenido de estos ficheros ha sido explicado en el epígrafe anterior.
- Una carpeta denominada “data” que contendrá dos subcarpetas:
 - Una carpeta denominada “logs-datos-sip”, que contendrá los siguientes ficheros de control:
 - listado.txt, que sólo listará los ficheros conservados en la carpeta METADATOS_RECIBIDOS.
 - sip_estr_cr.txt, que contendrá una representación de la estructura de carpetas del PIT a nivel de obra multivolumen o de cabecera.
 - tab_corp, que sólo representará los ficheros conservados en la carpeta METADATOS_RECIBIDOS.
 - Una carpeta denominada “metadatos_recibidos” que contendrá los metadatos recibidos a nivel de cabecera o libro multivolumen. Esta carpeta podrá contener subcarpetas diferentes para cada tipo de metadatos, debiendo ser su nombre suficientemente significativo de su contenido. Esta carpeta contendrá los ficheros de control adicionales que se hayan considerado como necesarios para poder reconstruir la estructura original de los discos en caso de la necesidad de procesos complejos de organización de la información procedente de varios discos. De entrada, se parte de la necesidad de incluir dos ficheros de control adicionales: correspondencia_archivos_jpeg.txt y correspondencia_archivos_pdf.txt. Aunque en el caso de que la publicación presente también ficheros OCR en formato ALTO o ficheros epub se podrán crear también sus correspondientes ficheros de control: correspondencia_archivos_alto.txt o correspondencia_archivos_epub.txt. Estos ficheros de control adicionales informan respectivamente de las operaciones de movimiento de ficheros necesarias para la organización automática de los fondos antes de su ingreso en el repositorio de preservación.

2.7 Tratamiento dado a los metadatos en el paquete de preservación PIA.

2.7.1 Metadatos del PIA, sus esquemas y formas de codificación admitidas.

Los metadatos se registran 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 usa la última versión de este estándar.

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

Recordamos que sólo se generan metadatos administrativos o de cualquier tipo dentro del fichero METS del PIA 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 ser referidos estos objetos.

En los epígrafes siguientes describimos los tipos de metadatos que se incorporarán al archivo METS.

2.7.1.1 Metadatos descriptivos (bibliográficos) (dmdSec METS).

Estos metadatos contienen la descripción bibliográfica y de contenido de los documentos a los que corresponden los objetos digitales resultado de la digitalización o de las unidades intelectuales nacidas digitales.

Es obligatoria la presencia de metadatos descriptivos para cada documento a nivel de ítem o de cabecera u obra, y que se aporten en alguno de los esquemas de metadatos y sistemas de codificación admitidos por la comunidad bibliotecaria. Estos metadatos descriptivos se insertan en una o varias secciones dmdSec del fichero METS.

En los METS de preservación del PIA a nivel de publicación periódica (cabecera) y de libro con varios volúmenes (obra completa) se incluyen los metadatos en formato MARCXML a estos niveles, creados desde los ficheros en formato ISO 2709, que se ubican en las carpetas así denominadas. Se incluye el registro bibliográfico y el de fondos.

Para el proyecto BVPH, a pesar de que se suministran los registros bibliográficos en varios formatos, en los METS de preservación del PIA a nivel de número de publicación periódica se incluirán los metadatos en formato MODS, tal y como aparecen en los METS de preservación suministrados.

Para el proyecto BVPB, en los METS de preservación del PIA a nivel de volumen de libro o documento simple (fotografía, mapa, ilustración...) se incluirán los metadatos en formato MARCXML, tal y como aparecen en los METS de preservación suministrados por la SGCB. Se debe incluir el registro bibliográfico y el (o los) de fondos.

2.7.1.2 Metadatos administrativos (amdSec METS).

Incluyen una amplia tipología de metadatos:

- Técnicos (techMD METS). Describen parámetros técnicos de los ficheros digitales, usándose los formatos PREMIS (Entidad Object) y NISO Technical Metadata en su sistema de representación XML con el esquema MIX⁷. Se representan todas las unidades semánticas PREMIS Object de las que se disponga de información técnica y cuyo contenido se pueda extraer automáticamente, garantizando que están todas las unidades que son obligatorias en PREMIS cuando corresponde su aplicación al tipo de objeto procesado, 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).

⁷ NISO Technical Metadata for Digital Still Images Standards in XML. Disponible en <http://www.loc.gov/standards/mix/>

- 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. Se representan 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).
 - 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. Las normativas de preservación digital⁸ obligan a la representación de este tipo de metadatos, por lo que se considera como obligatorio en el repositorio de preservación digital. Estos metadatos se registran de acuerdo al esquema METSRights⁹.

Cualquier estrategia de preservación que se lleve a cabo en el servicio de repositorio, como es el caso de las migraciones de formatos, metadatos y sistemas de empaquetamiento, se registra en la base de datos del sistema empleando los metadatos de Eventos PREMIS correspondientes. Estos metadatos se actualizan también en los ficheros de metadatos METS incluidos en los paquetes de los PIA. En la generación del nuevo paquete PIA tras la aplicación de las estrategias de preservación pertinentes, se generarán de nuevo los códigos hash para cada fichero.

El idioma de los valores predeterminados de los metadatos PREMIS será el inglés.

2.7.1.3 Metadatos estructurales (fileSec y structMap METS).

Permiten vincular lógicamente los ficheros que forman parte de un mismo objeto digital. Se incluyen las secciones fileSec y structMap, de manera que todos los objetos incluidos en la carpeta “objetos” del paquete PIA queden vinculados correctamente y en la secuencia que les corresponde en el documento original.

2.7.1.4 Codificación de la estructura lógica del fondo (biblioteca, colección, archivo, sección, serie, etc.) en los metadatos METS.

⁸ Por ejemplo, el estándar OAIS exige el registro para cada PIA de al menos: las restricciones de acceso a los contenidos, el marco legal de regulación de propiedad intelectual que se aplica, y los términos de la licencia de uso, distribución y control de acceso de los objetos de contenido. Esta información de acuerdo a este estándar debería quedar reflejada en el contrato o documento legal que regula el envío de documentos al repositorio de preservación, así como los permisos de custodia y procesado de preservación digital que habilitan al envío y preservación digital de los documentos en el servicio de repositorio.

⁹ Accesible en <http://www.loc.gov/standards/rights/METSRights.xsd>

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

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

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 PIA. 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 biblioteca o institución documental que conserva el
                                   documento digitalizado de acuerdo al sistema de códigos establecido en el proyecto
                                   BVPH o BVPB.

<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) separados por comas.

</datos_custodia>

```

Los datos de nombre de disco en la etiqueta en papel y el año de digitalización son relevantes, por lo que se incluyen en los metadatos del paquete PIA, dentro de los datos de custodia en dos elementos nuevos específicos del sistema adaptado a la SGCB.

Un 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="Ministerio de Educación, Cultura y Deporte"/>

```


Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

```

<unidad valor="Subdirección General de Coordinación Bibliotecaria" nivel="1"/>
<fondo valor="Biblioteca Virtual de Prensa Histórica" nivel="2"/>
<fondo valor="Hemeroteca de Madrid" nivel="3"/>
<codigoBiblioteca valor="HEM-M"/>
<digitalizacion valor="2012"/>
<soporte
valor="HD_TIFF_006_BVPH_2012,HD_JPEG_006_BVPH_2012,HD_OCR_002_BVPH_2012,CD_METADATOS_M
CU"/>

```

</datos_custodia:datos_custodia>

Los metadatos de custodia sólo se almacenarán a nivel del mets de número de publicación periódica o del volumen de libro. De esta manera, en el caso de que una publicación periódica o libro presentado en varios volúmenes haya sido digitalizado por varias bibliotecas, cada número de publicación o volumen de libro lleva asociado sus metadatos de custodia correspondientes a la biblioteca que conserva los ejemplares digitalizados. Consiguientemente, los metadatos de custodia no se incluirán en los ficheros METS a nivel de cabecera o de obra no seriada presentada en varios volúmenes.

2.7.2 Perfiles METS específicos para los ficheros METS de preservación del paquete PIA.

2.7.2.1 Perfil METS específico del fichero METS de preservación digital del paquete PIA a nivel de número de revista, volumen de libro o ítem individual.

2.7.2.1.1 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"?>
```

2.7.2.1.2 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.

2.7.2.1.3 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 PIA en los dispositivos de almacenamiento usados, sin que pierdan su sentido cuando se procede a traspasar los PIA a otros dispositivos o configuraciones de almacenamiento donde pueden cambiar las denominaciones de las unidades o carpetas contenedoras, o su estructura.

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

El elemento raíz incorporará necesariamente las declaraciones de xmlns de los esquemas XML siguientes: XSI, XLINK, PREMIS y METS. Y lo hará de la forma que muestra el siguiente ejemplo, aunque siempre con las actualizaciones requeridas para referir correctamente a las versiones de los esquemas que se utilicen en cada momento:

```
<mets:mets xmlns:mets="http://www.loc.gov/METS/" xmlns:premis="http://www.loc.gov/premis/v3"
xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/version111/mets.xsd
http://www.loc.gov/premis/v3 http://www.loc.gov/standards/premis/v3/premis.xsd ">
```

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.

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 en acuerdo con la SGCB.

Los esquemas del resto de elementos, como pueden ser los MIX, MARC, MODS, DC, etc. se referirán en su etiqueta padre correspondiente.

No se admitirá en los metadatos PREMIS la repetición en el elemento padre OBJECT, EVENT y AGENT de los atributos de espacio de nombres y localización del esquema.

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>
```

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

```
<collection xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  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>
```

```
[...]
```

2.7.2.1.5 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 del creador del fichero METS, tal y como vemos en el siguiente ejemplo:

```
<mets:agent TYPE="ORGANIZATION" ROLE="CREATOR">
  <mets:name>Empresa externa</mets:name>
  <mets:note>Creado en nombre de [Institución Cliente]</mets:note>
</mets:agent>
```

Admitiéndose también la forma siguiente:

```
<mets:agent ROLE="CREATOR" TYPE="ORGANIZATION">
  <mets:name>Ministerio de Educación, Cultura y Deporte</mets:name>
</mets:agent>
<mets:agent ROLE="EDITOR" TYPE="ORGANIZATION">
  <mets:name>xxxxxxx</mets:name>
</mets:agent>
```

En el caso de que se desee incluir más metadatos de cabecera, deberá comunicarlo al servicio de preservación digital, proporcionando al mismo tiempo los datos necesarios para su codificación.

2.7.2.1.6 Sección Descriptiva (dmdSEC)

En esta sección se incluirán, de forma interna (sin referencias a ficheros externos), los metadatos descriptivos del documento al que corresponde el paquete PIA (número de publicación periódica, libro, documento manuscrito, fotografía...), en los formatos y esquemas admitidos en la normativa de repositorio (MARCXML para las monografías y MODS para los números de revista o prensa) y, en todo caso, codificados en XML. Recordamos que no se admitirá la preservación digital de documentos 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 consignará el espacio de nombres (namespace) adecuado al caso, 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"
        xmlns: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 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. Esta redundancia se mantiene por considerarla positiva de cara a la preservación digital futura de los objetos digitales referenciados en el fichero METS, pudiéndose eliminar si se considera necesario.

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

En el caso del proyecto BVPH los metadatos descriptivos de carácter bibliográfico a incluir en el METS del PIA serán siempre los metadatos aportados en el METS de preservación o de carga del PIT (aportado por la SGCB) en la sección dedicada a los metadatos descriptivos en formato MODS ("DMDMODSISSUE"), independientemente de que el fichero METS de preservación que aporta la SGCB incluya la expresión de estos metadatos en otros formatos bibliográficos. Para el proyecto BVPB los metadatos descriptivos de carácter bibliográfico a incluir en el METS del PIA serán los metadatos aportados en el METS de preservación o de carga del PIT (aportado por la SGCB) en la sección dedicada a los metadatos descriptivos en formato MARCXML ("DMDMARC").

La declaración de metadatos de custodia que se incluye en una sección <dmdSec> específica incorporará, además de los datos establecidos por defecto por el sistema de preservación digital, los datos de nombre de disco del que se han extraído los contenidos digitales, el año de digitalización, el código de proyecto (BVPH o BVPB) y la institución que ha digitalizado los contenidos, con su nombre completo y su código de acuerdo al sistema de códigos de institución asentado por la SGCB.

2.7.2.1.7 Sección Administrativa (amdSec)

Se incluirá, al menos, una sección administrativa para cada uno de los objetos del PIA contenidos en la carpeta "objetos", ya sean másteres, derivados u objetos con otros tipos de función. Se usarán obligatoriamente las siguientes subsecciones, para cada uno de estas versiones:

- Metadatos técnicos (techMD). Se usará obligatoriamente el estándar PREMIS, que podrá ser complementado en el elemento <premis:objectCharacteristicsExtension> con el estándar MIX, o cualquier otro esquema acordado con la SGCB, que preferiblemente deberá estar sujeto a un estándar *de iure* o *de facto* y con un esquema público y bien documentado.
- 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.

Es obligatorio incluir al menos una sección de metadatos administrativos con metadatos de derechos (rightsMD). Se usarán los elementos para derechos de propiedad intelectual de PREMIS o de cualquier otro esquema XML estándar o de uso común que se requiera, preferiblemente METSRights. Si la SGCB no aporta este tipo de metadatos, el Servicio de Repositorio de preservación creará los metadatos mínimos de propiedad intelectual codificados en PREMIS XML que indica esta normativa y los incrustará en esta sección. De modo general, los metadatos de derechos se conectarán a los ficheros de la sección fileSec de la siguiente manera: se creará una amdSec nueva que contendrá como único elemento el rightsMD que a su vez contendrá todos los metadatos de derechos aplicables al documento. 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</rightsStatementId
entifierType>
    <rightsStatementIdentifierValue>info:nyu-dl/x-
v1/pr/xpnvx139/rmd/v0001
  </rightsStatementIdentifierValue>
</rightsStatementIdentifier>
  <rightsBasis>license</rightsBasis>
  <licenseInformation>
    <licenseIdentifier>
      <licenseIdentifierType>NYU-Legal
      </licenseIdentifierType>
      <licenseIdentifierValue>0123456789</licenseId
entifierValue>
    </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"/>

```

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

```

</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 obstante, se podrán convenir otras formas de conexión de este tipo de metadatos, en cuyo caso se creará en este manual de procedimiento un epígrafe específico que explique al detalle esta forma diferente de conexión.

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 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.

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 PIA), salvo que la SGCB proporcione ya este tipo de metadatos codificados en XML (aptos y sin errores) para ser introducido en el fichero METS del PIA en la subsección sourceMD dentro de la sección amdSec.

Tal y como obliga el propio estándar METS, los elementos techMD, rightsMD y digiproMD, 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 adecuadamente¹⁰, 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.

¹⁰ 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>

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 adecuadamente¹¹, usándose los siguientes posibles valores, según corresponda al tipo de evento, ya sea en su versión inglesa o española:

- 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.

2.7.2.1.8 Sección de datos de ficheros (fileSec)

Sólo se incluirán en esta sección los ficheros contenidos dentro de la carpeta objetos y sus carpetas hijas. No se admitirá, por tanto la inclusión de los ficheros contenidos directamente en la carpeta padre del paquete PIA ni en sus hijas "data", "logs_datos_sip" y "metadatos_recibidos". 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 PIA. Cada uno de estos fileGrp incluirá la información relativa a los archivos que conforman la obra a preservar en un formato determinado. Por ejemplo, el fileSec de un documento METS típico de PIA 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 que contenga una versión en pdf de la obra.

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\TIFF_editado\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">
```

¹¹ 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>

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

```
< 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" GROUPIP="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 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 documento 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 actualización¹². En el caso de documentos de texto y si está disponible el dato se 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 GROUPIP 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 PIA.

Para los dos proyectos de la SGCB, BVPH y BVPB, la sección fileSec se ajustará obligatoriamente a los requisitos siguientes, que sustituyen a los equivalentes en el perfil METS base:

1. Habrá una única declaración de metadatos de propiedad intelectual en el documento METS, en formato METSRights, que se hará en una sección de metadatos administrativos específica.
2. Los metadatos administrativos de propiedad intelectual (METSRights), previamente declarados en una sección de metadatos administrativos específica, se vincularán únicamente a los elementos fileGrp que organizan los elementos file en la sección de ficheros (fileSec). Por lo que no se incluirá el valor del atributo ID de la <admSec> de derechos en el atributo ADMID de los elementos file. En su lugar, se usará el atributo ADMID en los

¹² Accesible desde <http://www.iana.org/assignments/media-types/media-types.xhtml>

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

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 datosCustodia-0010000e-0000-41d8-869e-30d3dba619a2" 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 datosCustodia-0010000e-0000-41d8-869e-30d3dba619a2" 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>
  </mets:fileGrp>
  [...]
```

2.7.2.1.9 Sección del mapa estructural (structMap)

Se creará un mapa estructural, que reflejará la estructura física de carpetas del PIA con objetos digitales del documento a preservar digitalmente, esto es, partiendo de la carpeta padre “objetos”. 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 PIA (ubicación, secuencia y estructura) y que facilitan la navegación dentro de este paquete y la realización de operaciones de procesado, transferencia y preservación digital. Este mapa estructural deberá llevar como valor “PIA_STRUCTMAP”.

Se creará además, en los casos en los que la SGCB proporcione archivos METS de los objetos digitales, otro mapa estructural (que llamaremos mapa estructural de la obra) que represente la estructura lógica o física de la obra digital a preservar. La información de esta sección de mapa estructural se extraerá de los archivos METS proporcionados por la SGCB, incluyéndose en cualquier caso información sobre el tipo de objeto, de la secuencia de elementos (páginas) y de los valores de visualización asociados a cada elemento, así como los atributos de enlace necesarios para relacionar los elementos del mapa estructural con las otras secciones del METS. .

Cualquier cambio en estos mapas estructurales deberá ser reflejado en los metadatos de evento.

Mapa estructural del PIA.

Se podrán incluir tantas divisiones estructurales se necesiten en el mapa estructural del PIA, pero siempre siguiendo la normativa de estructura de los PIA asentada por la normativa del Servicio de Repositorio.

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

En el momento actual, de acuerdo a esta normativa, el mapa estructural del PIA deberá ser de tipo físico referido a la estructura del paquete PIA, 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 "PIA_STRUCTMAP". Deberá llevar también un atributo ID con valor único, como podemos apreciar en el siguiente ejemplo:

```
<mets:structMap LABEL="PIA_STRUCTMAP" TYPE="PHYSICAL" ID="structMap001">
```

A continuación del elemento padre structMap se incluirá la estructura de carpetas del PIA mediante elementos DIV anidados.

Como podemos apreciar en el siguiente 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 PIA, tal y como vemos en el siguiente ejemplo:

```
<mets:structMap LABEL="PIA_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_Coleccion 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" GROUPLD="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>
```

[...]

Salvo en el caso de ficheros METS dedicados a representar a nivel de cabecera (título de publicación periódica) o libro presentado en varios volúmenes, los elementos DIV no deben contener en ningún caso elementos mptr. Estos elementos son punteros a contenido representado por un documento METS externo.

Mapa estructural de la obra

En el proyecto de la SGCB al mapa estructural del PIA a nivel de número de publicación periódica, de volumen de libro o de documento no librario (fotografía, mapa, grabado...) se añadirá otro mapa estructural de tipo físico (para la BVPB) o lógico (para la BVPH) que representará la estructura del documento digitalizado, relacionándose cada elemento estructural con los objetos digitales correspondientes que deben ser preservados: máster TIFF, derivado JPEG o archivos ALTO.

Este mapa se ubicará en el fichero METS antes que el mapa del PIA.

La SGCB proporcionará la información requerida para la reconstrucción de la estructura física y función de los elementos correspondientes a cada objeto digital en los ficheros METS de preservación y de carga que aporta dentro del conjunto de ficheros de metadatos para cada documento a preservar digitalmente.

Las características de este mapa estructural secundario serán obligatoriamente¹³:

1. Heredará la estructura de elementos DIV y los valores de los atributos type, order y label de los distintos elementos del mapa estructural del METS de preservación correspondiente.
2. De esta manera se conformará un mapa estructural con secciones DIV que anidarán tantos niveles de etiquetas <div> como niveles jerárquicos tenga la estructura definida para el documento en el METS de preservación. En cualquier caso, cada <structMap> anidará al menos una etiqueta <div> (div de primer orden) que corresponda a la obra en su conjunto.
3. El elemento <structMap> deberá llevar un atributo TYPE con valor "physical".
4. El elemento <structMap> deberá llevar un atributo LABEL cuyo valor sea el del correspondiente LABEL del METS de preservación aportado por la SGCB.

¹³ Para la elaboración de esta lista de criterios se ha seguido el perfil METS público de la SGCB. Algunos de estos criterios reproducen literalmente párrafos de este perfil. Este perfil está accesible en: <http://www.loc.gov/standards/mets/profiles/0000044.xml>

5. El primer <div> (div de primer orden) debe incluir un atributo DMDID que enlace con la 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.
6. Cada <div> del mapa estructural debe contener el atributo ORDER, cuyo valor sea el número de orden de dicho <div,> dentro del conjunto de etiquetas <div> que se encuentran a su mismo nivel de anidamiento, es decir, entre los <div> hermanos. El valor será heredado del correspondiente atributo ORDER en el METS de preservación aportado.
7. Cada <div> debe contener además el atributo TYPE. El valor de este atributo indicará el tipo o nivel de <div> (libro, capítulo, página...). En el primer nivel de <div> se indicará mediante este atributo el tipo de obra, por ejemplo "libro", "mapa", etc... Si se trata de una estructura de sólo dos niveles (lo que es el caso más habitual en la BVPB), por ejemplo una monografía dividida en páginas (una imagen por página), los valores de este atributo para el primer nivel de <div> indicarán el tipo de obra (por ejemplo TYPE="book"), mientras que los valores de este atributo para el segundo nivel de <div> indicarán que tal nivel corresponde a las imágenes individuales (normalmente de página -- TYPE="page"). Los valores de los atributos se darán en inglés, y serán los heredados de los atributos correspondientes del METS de preservación aportado.
8. Cada <div> debe contener además un atributo LABEL, cuyo valor servirá para identificar la parte de la estructura que representa dicha etiqueta. El valor será heredado del correspondiente atributo LABEL en el METS de preservación aportado.
9. Los <div> que correspondan a las imágenes de página (con atributo TYPE="page") indicarán el número de página como valor del atributo LABEL a través de un número, o, en su defecto, el tipo de página mediante una palabra representativa, tal como "[guarda]", "[cubierta]", etc.
10. Los elementos fptr de los DIV de página referirán mediante el atributo FILEID a los valores de los atributos ID de los elementos FILE correspondientes en la sección FileSec del METS del PIA.
11. Cada elemento DIV de página tendrá tantos elementos fptr hijos anidados como tipos de ficheros hayan sido declarados en la FileSec para esa página y de los que pueda establecerse la correspondencia automáticamente: TIFF, JPEG, ALTO (en su caso). Esto es así porque podemos considerarlos como versiones alternativas del mismo contenido (máster, derivado de visualización, OCR...). Podemos apreciar un caso donde se refieren las versiones TIFF y JPEG en el siguiente ejemplo:

```
<structMap LABEL="Contenido de la obra" TYPE="physical">
```

```
  <div LABEL="2531: Feliu y Perez, Bartolome (. Curso elemental de fisica experimental y aplicada y nociones de química inorgánica: para uso de los establecimientos de 2º enseñanza, seminarios y escuelas normales." DMDID="dmdSec-00100001-0000-4c3d-8d8f-6fd2d9996e4a" TYPE="work">
```

```
    <div ORDER="1" LABEL="[Cubierta]" TYPE="page">
```

```
      <fptr FILEID="TIFF_00100001-0004-4be6-b897-6290287c171e"/>
```

```
      <fptr FILEID="JPEG_00100001-0005-4b3f-ac64-817dc247e072"/>
```

```
    </div>
```

```
    <div ORDER="2" LABEL="[Guarda]" TYPE="page">
```

```
      <fptr FILEID="TIFF_00100001-0006-488b-8a91-9c95b819f752"/>
```

```
      <fptr FILEID="JPEG_00100001-0007-4873-818e-c788ccf568e0"/>
```

```
    </div>
```

```
  <div ORDER="3" TYPE="page" >
```

```

    <fptr FILEID="TIFF_00100001-0009-4ee4-8f28-5bab9f002c96"/>
    <fptr FILEID="JPEG_00100001-000a-4ebd-949b-696618abf59b"/>
  </div>
  [...]
<div ORDER="13" LABEL="1" TYPE="page">
  <fptr FILEID="TIFF_00100001-0026-36Aaa-9f28-5bab9f002c96"/>
  <fptr FILEID="JPEG_00100001-0041-6789-0091-6899911cc59f"/>
</div>
[...]
```

2.7.2.1.10 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 version="2.2" xsi:type="premis:file">
```

```

  <premis:objectIdentifier><premis:objectIdentifierType>UUID</premis:objectIdentifierType><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 los elementos 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>

```

2.7.2.1.11 Cómo especificar la función de los ficheros: 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,:

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

- master image. Para un fichero de imagen raster o vectorial con la función de máster.
- 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.
- Alto ocr. Para ficheros resultado de un procesado OCR con salida en formato ALTO.

Si hay otros tipos de derivados como "TIFF Editado" o "PDF_OCR", al ser derivados sus valores, para el atributo USE serán imagen referencia y texto referencia según corresponda a su tipo de medio.

En este tipo de casos, se pueden generar subgrupos dentro del elemento fileGrp, anidando varios fileGroup, de manera que quede reflejada esta tipología mediante una estructura jerárquica.

2.7.2.1.12 Cómo relacionar cada fichero máster con su correspondiente o sus 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 GROUPLID del elemento FILE. El valor de este atributo debe ser idéntico entre el máster y sus derivados para que se pueda representar sin ambigüedad esta relación. Veamos un ejemplo de relación entre los originales (serían los másteres no de preservación) y los derivados correspondientes de preservación:

Los originales:

`<mets:fileSec>`

```
<mets:fileGrp ADMID="amdSec_rights-0012db14-0000-44cf-940a-315da9137612" ID="fileGrp-Master_Image-0012db14-0000-44cf-940a-315da9137612" USE="master image">
```

```
  <mets:file ADMID="amdSec-0012db14-0016-47f3-97cf-c43238691962" DMDID="mods_descriptive-0012db14-0000-44cf-940a-315da9137612 datosCustodia-0012db14-0000-44cf-940a-315da9137612 holdings-0012db14-0000-44cf-940a-315da9137612" GROUPLID="page_3" ID="_0012db14-0016-47f3-97cf-c43238691962" MIMETYPE="image/tiff" SEQ="3">
```


Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

```
<mets:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="objetos/masteres/prensa_0003-0012db14-0016-47f3-97cf-c43238691962.tif" xlink:type="simple"/>
```

```
</mets:file>
```

```
<mets:file ADMID="amdSec-0012db14-0017-433c-969e-2b959a064af6" DMDID="mods_descriptive-0012db14-0000-44cf-940a-315da9137612 datosCustodia-0012db14-0000-44cf-940a-315da9137612 holdings-0012db14-0000-44cf-940a-315da9137612" GROUPID="page_4" ID="_0012db14-0017-433c-969e-2b959a064af6" MIMETYPE="image/tiff" SEQ="4">
```

```
<mets:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="objetos/masteres/prensa_0004-0012db14-0017-433c-969e-2b959a064af6.tif" xlink:type="simple"/>
```

```
</mets:file>
```

```
<mets:file ADMID="amdSec-0012db14-0018-41de-9972-934ce6590ccf" DMDID="mods_descriptive-0012db14-0000-44cf-940a-315da9137612 datosCustodia-0012db14-0000-44cf-940a-315da9137612 holdings-0012db14-0000-44cf-940a-315da9137612" GROUPID="page_5" ID="_0012db14-0018-41de-9972-934ce6590ccf" MIMETYPE="image/tiff" SEQ="5">
```

```
<mets:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="objetos/masteres/prensa_0005-0012db14-0018-41de-9972-934ce6590ccf.tif" xlink:type="simple"/>
```

```
</mets:file>
```

```
</mets:fileGrp>
```

.....

```
</mets:fileSec>
```

Y los derivados para preservación. Si observamos el atributo Groupid, coinciden los valores entre los ficheros originales y los de preservación correspondientes:

```
<mets:fileSec>
```

....

```
<mets:fileGrp ADMID="amdSec_rights-0012db14-0000-44cf-940a-315da9137612" ID="fileGrp-Reference_Image-0012db14-0000-44cf-940a-315da9137612" USE="reference image">
```

```
<mets:file ADMID="amdSec-0012db14-000c-4854-af74-4f29a388981c" DMDID="mods_descriptive-0012db14-0000-44cf-940a-315da9137612 datosCustodia-0012db14-0000-44cf-940a-315da9137612 holdings-0012db14-0000-44cf-940a-315da9137612" GROUPID="page_3" ID="_0012db14-000c-4854-af74-4f29a388981c" MIMETYPE="image/jpeg" SEQ="3">
```

```
<mets:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="objetos/derivados/jpeg/prensa_0003-0012db14-000c-4854-af74-4f29a388981c.jpg" xlink:type="simple"/>
```

```
</mets:file>
```

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

```

<mets:file ADMID="amdSec-0012db14-000d-494c-8453-48cdea9c9b13" DMDID="mods_descriptive-
0012db14-0000-44cf-940a-315da9137612 datosCustodia-0012db14-0000-44cf-940a-315da9137612 holdings-
0012db14-0000-44cf-940a-315da9137612" GROUPID="page_4" ID="_0012db14-000d-494c-8453-
48cdea9c9b13" MIMETYPE="image/jpeg" SEQ="4">

  <mets:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM"
xlink:href="objetos/derivados/jpeg/prensa_0004-0012db14-000d-494c-8453-48cdea9c9b13.jpg"
xlink:type="simple"/>

</mets:file>

<mets:file ADMID="amdSec-0012db14-000e-4277-8cd0-a855641f3240" DMDID="mods_descriptive-
0012db14-0000-44cf-940a-315da9137612 datosCustodia-0012db14-0000-44cf-940a-315da9137612 holdings-
0012db14-0000-44cf-940a-315da9137612" GROUPID="page_5" ID="_0012db14-000e-4277-8cd0-
a855641f3240" MIMETYPE="image/jpeg" SEQ="5">

  <mets:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM"
xlink:href="objetos/derivados/jpeg/prensa_0005-0012db14-000e-4277-8cd0-a855641f3240.jpg"
xlink:type="simple"/>

</mets:file>

</mets:fileGrp>
</mets:fileSec>

```

2.7.2.1.13 Metadatos PREMIS heredados

Los metadatos PREMIS heredados en XML serán incorporados al fichero METS del PIA en su lugar correspondiente dentro de la sección amdSec, subsecciones techMD, rightsMD, o digiprovMD.

En el caso en que los metadatos se refieran a la totalidad del documento digital se creará una sección amdSec específica para estos metadatos, que tendrán que ser conectados al documento completo mediante el atributo ID del elemento amdSec y el atributo AMDID de la división principal (DIV) del mapa estructural (structMap) o el atributo AMDID de un elemento fileGrp de la sección fileSec que abarque todos los objetos digitales que componen el documento.

2.7.2.1.14 Ficheros METS heredados

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

2.7.2.1.15 Ausencia de los elementos structLink y behaviorSec

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

2.7.2.1.16 Tratamiento de los ficheros de cartas de color

En algunos números de publicación periódica o libros se presentan ficheros de página repetidos que registran también la imagen de una carta de color. El tratamiento a dar a estos ficheros repetidos con carta de color en el código METS se atenderá a los siguientes criterios, que toman en consideración el que un fichero de carta de color sirve para controlar la calidad en la captura y procesado no sólo para la página a la que se ha adjuntado sino para todas las capturas de página del documento:

1. El fichero de la carta de color se ubicará en una carpeta específica que colgará dentro de la carpeta padre de los másteres, cuyo nombre será "carta_color".
2. El fichero será referido en la fileSec dentro de un elemento fileGrp propio que tendrá como valor del atributo USE "color card".

Seguidamente vemos un ejemplo de implementación de estos criterios:

<mets:fileSec>

```
<mets:fileGrp ADMID="amdSec_rights-0012db14-0000-44cf-940a-315da9137612" ID="fileGrp-Color_Card-0012db14-0000-44cf-940a-315da9137612" USE="color card">
```

```
  <mets:file ADMID="amdSec-0012db14-0015-4dec-808c-7bb67666f1f0" DMDID="mods_descriptive-0012db14-0000-44cf-940a-315da9137612 datosCustodia-0012db14-0000-44cf-940a-315da9137612 holdings-0012db14-0000-44cf-940a-315da9137612" GROUPLID="prensa_0003 - cartaColor.tif" ID="_0012db14-0015-4dec-808c-7bb67666f1f0" MIMETYPE="image/tiff">
```

```
    <mets:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM"
      xlink:href="objetos/masteres/carta_color/prensa_0003_-_cartaColor-0012db14-0015-4dec-808c-7bb67666f1f0.tif" xlink:type="simple"/>
```

```
  </mets:file>
```

```
</mets:fileGrp>
```

2.7.2.1.17 Tratamiento de los ficheros de partición de un fichero multipágina voluminoso

En algunos documentos seriados o libros se da la circunstancia de que en lugar de un único fichero PDF multipágina aparecen varios ficheros PDF que representan fracciones de ese fichero. Esto se hace así para evitar archivos PDF de gran tamaño. La codificación METS en estos casos se atenderá a estos criterios:

1. En la división del mapa estructural de la obra correspondiente a la versión PDF de la obra se deberán crear tantas subdivisiones como archivos PDF de fracción se presenten mediante elementos <div>. Cada una de esas subdivisiones referirá a través de un elemento <fptr> al fichero PDF de fracción que le corresponde, que evidentemente tendrá su propia representación en la fileSec.
2. Tanto la división principal como sus subdivisiones deberán llevar un elemento LABEL, cuyo valor será "multipage file".

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

3. El atributo ID de los elementos div correspondientes a las subdivisiones deberá llevar una cadena de caracteres indicativa de que se trata de una fracción, y preferiblemente el número de orden que le corresponde a esa fracción de acuerdo a la secuencia natural de las páginas que contiene su fichero PDF correspondiente.
4. En la sección de ficheros (fileSec) se deberá seguir respetando el principio de usar un elemento <fileGrp>, con un atributo USE con valor "multipage file", para contener al PDF multipágina, pero en este caso, este elemento tendrá tantos elementos <file> como ficheros de fracción PDF se presenten. El orden de estos elementos <file> seguirá el orden natural de las páginas representadas en sus ficheros PDF correspondientes

Seguidamente vemos un ejemplo de implementación de estos criterios:

```
<mets:fileGrp ADMID="amdSec_rights-0012db14-0000-44cf-940a-315da9137612" ID="fileGrp-Pdf-0012db14-0000-44cf-940a-315da9137612" USE="multipage file">
```

```
  <mets:file ADMID="amdSec-0012db14-0012-430e-919c-0d99a7520d05" DMDID="mods_descriptive-0012db14-0000-44cf-940a-315da9137612 datosCustodia-0012db14-0000-44cf-940a-315da9137612 holdings-0012db14-0000-44cf-940a-315da9137612" GROUPID="1989-11-25_numero_183-184-185_id120000390.001.pdf" ID="_0012db14-0012-430e-919c-0d99a7520d05" MIMETYPE="application/pdf">
```

```
    <mets:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="objetos/derivados/pdf/1989-11-25_numero_183-184-185_id120000390_001-0012db14-0012-430e-919c-0d99a7520d05.pdf" xlink:type="simple"/>
```

```
</mets:file>
```

```
  <mets:file ADMID="amdSec-0012db14-0013-493b-98af-b63243142c2a" DMDID="mods_descriptive-0012db14-0000-44cf-940a-315da9137612 datosCustodia-0012db14-0000-44cf-940a-315da9137612 holdings-0012db14-0000-44cf-940a-315da9137612" GROUPID="1989-11-25_numero_183-184-185_id120000390.002.pdf" ID="_0012db14-0013-493b-98af-b63243142c2a" MIMETYPE="application/pdf">
```

```
    <mets:FLocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href="objetos/derivados/pdf/1989-11-25_numero_183-184-185_id120000390_002-0012db14-0013-493b-98af-b63243142c2a.pdf" xlink:type="simple"/>
```

```
</mets:file>
```

```
</mets:fileGrp>
```

[...]

```
<mets:structMap ID="structMapLogical-0012db14-0000-44cf-940a-315da9137612" LABEL="Contenido de la obra" TYPE="physical">
```

```
  <mets:div ADMID="amdSec_rights-0012db14-0000-44cf-940a-315da9137612" DMDID="mods_descriptive-0012db14-0000-44cf-940a-315da9137612 datosCustodia-0012db14-0000-44cf-940a-315da9137612 holdings-0012db14-0000-44cf-940a-315da9137612" ID="divLogica_0012db14-0000-44cf-940a-315da9137612" LABEL="Las Antillas (Barcelona. 1866) | Las Antillas: Año I Número 1 - 1866 diciembre 10" TYPE="news:issue">
```

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

```
<mets:div ID="divFraccion_pdf_1" LABEL="multipage file" TYPE="news:number:part"
ORDER="1">
    <mets:fptr FILEID="_0012db14-0012-430e-919c-0d99a7520d05"/>
</mets:div>
<mets:div ID="divFraccion_pdf_2" LABEL="multipage file" TYPE="news:number:part"
ORDER="2">
    <mets:fptr FILEID="_0012db14-0013-493b-98af-b63243142c2a"/>
</mets:div>
```

2.7.2.1.18 Tratamiento de los ficheros en formato epub

En el proyecto BVPH hay publicaciones que presentan ficheros multipágina en formato epub, algunos de ellos, los que indique la SGCB, hay que conservarlos. La codificación METS para estos ficheros seguirá los siguientes criterios:

1. En la sección fileSec se creará un elemento <fileGrp> específico para el fichero epub, que contendrá un elemento <file> para este fichero. El valor de su atributo USE será "epub".
2. En el mapa estructural de obra correspondiente al epub se creará una nueva división específica para este fichero, cuyo atributo LABEL tendrá como valor "epub". Esta división llevará un elemento <fptr> que referirá, a través de su atributo FILEID, al fichero epub.

Seguidamente vemos un ejemplo de implementación de estos criterios:

```
<mets:fileGrp ADMID="amdSec_rights-0012db14-0000-44cf-940a-315da9137612" ID="fileGrp-Epub-0012db14-0000-44cf-940a-315da9137612" USE="epub">
    <mets:file ADMID="amdSec-0012db14-000f-478b-9528-6f1195each52" DMDID="mods_descriptive-0012db14-0000-44cf-940a-315da9137612 datosCustodia-0012db14-0000-44cf-940a-315da9137612 holdings-0012db14-0000-44cf-940a-315da9137612" GROUPID="las_antillas_ano_i_numero_1-1866_diciembre_10_10_12_1866_15080594.epub" ID="_0012db14-000f-478b-9528-6f1195each52" MIMETYPE="application/epub+zip">
        <mets:Flocat LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM"
        xlink:href="objetos/derivados/epub/las_antillas_ano_i_numero_1-1866_diciembre_10_10_12_1866_15080594-0012db14-000f-478b-9528-6f1195each52.epub"
        xlink:type="simple"/>
    </mets:file>
</mets:fileGrp>
```

[...]

```
<mets:structMap ID="structMapLogical-0012db14-0000-44cf-940a-315da9137612" LABEL="Contenido de la obra"
TYPE="physical">
  <mets:div ADMID="amdSec_rights-0012db14-0000-44cf-940a-315da9137612" DMDID="mods_descriptive-
0012db14-0000-44cf-940a-315da9137612 datosCustodia-0012db14-0000-44cf-940a-315da9137612 holdings-
0012db14-0000-44cf-940a-315da9137612" ID="divLogical_0012db14-0000-44cf-940a-315da9137612"
LABEL="Las Antillas (Barcelona. 1866)| Las Antillas: Año I Número 1 - 1866 diciembre 10" TYPE="news:issue">

  <mets:div ID="divNumeroCompleto_epub" LABEL="epub" TYPE="news:number">
    <mets:fptr FILEID="_0012db14-000f-478b-9528-6f1195eacb52"/>
  </mets:div>
```

2.7.2.2 Perfil METS específico del fichero METS de preservación digital del paquete PIA a nivel de cabecera y de obra multivolumen

2.7.2.2.1 A nivel de publicación periódica

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" LABEL="La Libertad (Madrid. 1919)|La Libertad"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:mets="http://www.loc.gov/METS/">
```

El atributo LABEL se usará para representar el título de la publicación.

Los ficheros METS llevarán dentro de la sección de metadatos descriptivos los metadatos bibliográficos que aparecen en la carpeta ISO2709 incluyendo sus dos tipos de registros: bibliográficos y de fondos (también llamados de ejemplar o ejemplares). Estos ficheros aparecen en esta carpeta en formato .mrc o .txt. Se precisará convertirlos automáticamente al formato MARCXML. La transformación se puede hacer con herramientas de solvencia, como MarcEdit. En la transformación a MARCXML deberá cuidarse que los caracteres acentuados no se transforman o se pierden, cosa que puede ocurrir con facilidad. Se deberá configurar la transformación para que la salida sea UTF-8. Si la transformación a MARCXML no derivará en buenos resultados podrán usarse en su lugar los ficheros en formato MARCXML ubicados en la carpeta MARC21, que tiene subcarpetas para, respectivamente, los registros bibliográficos, y los de ejemplares. Previamente al uso de estos ficheros deberá comprobarse que realmente están en formato MARCXML, ya que en algunas muestras pueden encontrarse formatos de visualización MARC en lugar de MARCXML.

Los dos tipos de registros preferiblemente se integrarán dentro del elemento MARC collection. En primer lugar se ubicará el bibliográfico y a continuación el o los de fondos (ejemplares). Por tanto, en esta configuración, el elemento collection llevará dos o más elementos <record>. También se permite que estos dos registros vayan insertados en dos

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

o más secciones dmdSEC independientes, debiendo ser una consecutiva de la otra y yendo en primer lugar el registro bibliográfico.

En el caso de que una publicación haya sido digitalizada por varias bibliotecas diferentes, se encontrarán varios registros de fondos (tantos como bibliotecas) que habrá que conservar dentro de la sección de metadatos descriptivos. Por ejemplo, en el caso de que se haya digitalizado en dos bibliotecas tendremos dos registros de fondos y tres elementos record dentro del elemento collection, si se optara por la primera configuración; o tres dmdSEC, si se optará por la segunda.

La estructura del fichero METS responderá al ejemplo que incluimos a continuación, en el que podemos apreciar cómo el mapa estructural se basa en crear la estructura lógica que organiza los números de la publicación periódica mediante elementos DIV anidados, partiendo de un elemento DIV padre para la publicación, y después anidando los elementos DIV para el año, mes, y número. El contenido del DIV dedicado a cada número llevará un elemento puntero que dirija al fichero METS externo que representa el número. Tal y como vemos en el siguiente ejemplo simplificado:

```
<mets:structMap ID="structMapLogical-0010000e-0000-41d8-869e-30d3dba619a2" LABEL="Contenido de la publicación" TYPE="LOGICAL">
```

```
<mets:div ADMID="amdSec-rights-0010000e-0000-41d8-869e-30d3dba619a2" DMDID="dmdSec-0010000e-0000-41d8-869e-30d3dba619a2" ID="0010000e-0000-41d8-869e-30d3dba619a2" LABEL="La Voz de Asturias" TYPE="periodical">
```

```
<mets:div ID="divLogical_2" LABEL="1923" TYPE="YEAR">
```

```
<mets:div ID="divLogical_3" LABEL="05" TYPE="MONTH">
```

```
<mets:div ID="divLogical_341" LABEL="01_id82290" TYPE="NUMBER">
```

```
<mets:mptr LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href=" [RUTA A METS CORRESPONDIENTE EN PIA DEL NUMERO]" />
```

```
</mets:div>
```

```
<mets:div ID="divLogical_4" LABEL="01_id82291" TYPE="NUMBER">
```

```
<mets:mptr LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href=" [RUTA A METS CORRESPONDIENTE EN PIA DEL NUMERO]" />
```

```
</mets:div>
```

```
<mets:div ID="divLogical_341" LABEL="06" TYPE="MONTH">
```

```
<mets:div ID="divLogical_561" LABEL="01_id82294" TYPE="NUMBER">
```

```
<mets:mptr LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href=" [RUTA A METS CORRESPONDIENTE EN PIA DEL NUMERO]" />
```

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

```

</mets:div>
<mets:div ID="divLogical_12221" LABEL="01_id82295" TYPE="NUMBER">

    <mets:mptr LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href=" [RUTA A
    METS CORRESPONDIENTE EN PIA DEL NUMERO]" />

</mets:div>
</mets:div>
<mets:div ID="divLogical_1888" LABEL="1924" TYPE="YEAR">

```

[.. ETC...]

```

    </mets:div>
  </mets:div>
</mets:structMap>

```

2.7.2.2.2 A nivel de libro con varios volúmenes

El Sistema de este METS será igual que el de publicación periódica, salvo que la sección de mapa estructural tendrá como divisiones lógicas mediante elementos DIV la obra multivolumen (el primer DIV) y sus volúmenes (el siguiente nivel DIV anidado). Los DIV de volúmenes estarán al mismo nivel jerárquico dentro del DIV padre correspondiente a la obra.

```

<mets:structMap ID="structMapLogical-0010000e-0000-41d8-869e-30d3dba619a2" LABEL="Contenido de la obra"
TYPE="LOGICAL">

```

```

  <mets:div ADMID="amdSec-rights-0010000e-0000-41d8-869e-30d3dba619a2" DMDID="dmdSec-0010000e-0000-
  41d8-869e-30d3dba619a2 " ID="0010000e-0000-41d8-869e-30d3dba619a2" LABEL=" El ingenioso hidalgo Don Quijote
  de la Mancha / compuesto por Miguel de Cervantes Saavedra" TYPE="book">

```

```

  <mets:div ID="divLogical_1" LABEL="El ingenioso hidalgo Don Quijote de la Mancha . Volumen I"
  TYPE="volume">

```

```

    <mets:mptr LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href=" " [RUTA A METS
    CORRESPONDIENTE EN PIA DEL VOLUMEN I]" />

```

```

  </mets:div>

```

```

  <mets:div ID="divLogical_2" LABEL=" El ingenioso hidalgo Don Quijote de la Mancha . Volumen II " TYPE="
  volume " ">

```

```

    <mets:mptr LOCTYPE="OTHER" OTHERLOCTYPE="SYSTEM" xlink:href=" " [RUTA A METS
    CORRESPONDIENTE EN PIA DEL VOLUMEN II]" />

```



```
</mets:div>
```

```
</mets:div>
```

```
</mets:structMap>
```

Los metadatos descriptivos, de la misma manera, estarán en una única sección de metadatos descriptivos que aglutinará dentro del elemento *collection* dos registros en MARCXML: uno para el bibliográfico, que irá en primer lugar; y otro para el de fondos, que irá justo a continuación. También se admitirá la configuración de que los dos registros estén en dos *dmdSec* separados, aunque consecutivos y en el orden que acabamos de señalar. Estos registros se derivarán de la conversión a MARCML de los ficheros correspondientes en formato ISO 2709 (extensión .mrc) que están en las carpetas ISO2709→Bibliográficos e ISO2709→Ejemplares. Si la transformación a MARCXML no derivará en buenos resultados podrán usarse en su lugar los ficheros en formato MARCXML ubicados en la carpeta MARC21, que tiene subcarpetas para, respectivamente, los registros bibliográficos, y los de ejemplares. Previamente al uso de estos ficheros deberá comprobarse que realmente están en formato MARCXML, ya que en algunas muestras se han encontrado formatos de visualización MARC en lugar de MARCXML.

3 Tareas de preservación digital a realizar en SEPREDI_SGCB

3.1 Procedimientos de sincronización de copias con separación geográfica

Cada vez que se añaden contenidos y paquetes nuevos de información o se efectúan procesos de migración en los ficheros preservados dentro de la unidad custodiada en la empresa, se sincronizan las dos unidades de almacenamiento (en la sede de empresa contratista y la SGCB), verificándose durante el proceso que todas las copias son idénticas.

3.2 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 PIA a partir de los códigos hash. Este chequeo detecta cualquier problema de corrupción o daño de datos en el sistema de almacenamiento. El propio servicio de repositorio 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 son registrados en el sistema de gestión del repositorio y en los metadatos PREMIS empaquetados junto a los contenidos de cada paquete de información.

3.3 Informes periódicos de actividad y estado del sistema de preservación

Periódicamente se hará un envío de informes de actividad, donde se reflejarán todos los cambios y circunstancias habidas con en SEPREDI_SGCB en un plazo determinado: ingresos, chequeos, copias de seguridad, incidencias, actualizaciones de contenidos, procesos de migración, informes de seguimiento tecnológico, mejoras del servicio, etc.

3.4 Alertas de preservación digital

El servicio de seguimiento de riesgos de obsolescencia detectará situaciones problemáticas en este sentido, alertando a la SGCB de este hecho y haciéndole llegar una propuestas de aplicación de estrategias de preservación digital.

3.5 Actualización del plan de preservación digital

El administrador externo hará un seguimiento del entorno tecnológico con el fin de detectar cambios de criterios por parte de la comunidad de expertos en preservación digital con respecto a los estándares de facto de cara a la preservación digital de 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. Con el resultado de este seguimiento propondrá actualizaciones del plan de preservación, que será aprobado por la SGCB a través de su administrador interno.

En ningún caso la empresa encargada de la administración del sistema realizará cambios en los flujos de trabajo especificados en la documentación de SIPREDI_SGCB o en el plan de preservación de contenidos digitales para las bibliotecas digitales de la SGCB sin autorización de la SGCB.

3.6 Realización de procesos de migración

La migración, junto con la encapsulación, el refresco y el replicado de soporte, son las únicas estrategias de preservación digital activas que consideramos en este momento en el modelo de repositorio de preservación. La única estrategia que requiere la transformación de los ficheros de objeto digital es la migración. El cambio de soporte o sistema de almacenamiento de los PIA no implica un proceso de migración.

En el modelo de repositorio aquí planteado un proceso de migración supone una transformación del paquete PIA 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 implicará necesariamente la actualización de los metadatos técnicos y de procedencia digital.

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

Los técnicos del servicio de repositorio, responsables de la vigilancia de la evolución tecnológica, avisarán a la SGCB de la conveniencia o necesidad de realizar procesos de migración. Estos procesos se realizarán si la SGCB los autoriza o solicita que se lleven a cabo. Los técnicos del servicio de repositorio informarán también a la SGCB de la necesidad o conveniencia 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, y la SGCB autorizará o solicitará la realización de estas operaciones. En las futuras versiones de los PIA que se tengan que crear debido la aplicación de procesos de migración se irán creando los ficheros siguientes en la carpeta del paquete PIA denominada “logs_datos_sip”, que se añadirán a los ficheros de control ya heredados desde la primera versión del PIA, 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 PIA obsoleto que ha sido tomado como origen de la migración al PIA 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 PIA origen de la migración correspondiente con los nombres de ficheros y carpetas del PIA migrado, con el siguiente formato: cada fila tendrá los datos de una correspondencia de ficheros o carpetas en la forma *nombre en PIA origen, nombre PIA resultante*. Las filas se separarán por un carácter Intro. Si a una sola carpeta de origen en PIA corresponden varias en el PIA resultante se repetirá la fila tantas veces como carpetas correspondan en el PIA resultante, teniendo la columna para el PIA 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 PIA 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 PIA vigente en el momento de la migración aplicada. Este dato deberá ser

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

registrado asimismo en el registro de datos correspondiente para el PIA 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 PIA 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 PIA 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 PIA que haya sufrido distintos procesos de migración cualquiera de las versiones obsoletas del PIA previamente custodiadas en el repositorio.

4 Utilidades de creación de PIC

4.1 Buscador de contenidos preservados

El buscador de los contenidos preservados debe estar instalado en cada unidad de almacenamiento de SIPREDI_SGCB. Debe emplear la Información Descriptiva extraída de los metadatos MARC21 incluidos en los PIA. El anexo III determina los campos que se utilizarán para la búsqueda.

Los resultados de la búsqueda deben proporcionar acceso a todos los elementos de los PIA que respondan a los criterios de búsqueda utilizados (sus carpetas, imágenes digitales y metadatos).

El buscador ha de permitir realizar una selección sobre los resultados de búsqueda y la descarga de dicha selección. De esta manera se han de poder descargar todos o parte de los resultados de búsqueda.

Asimismo ha de haber opciones que limiten la descarga según el tipo de formato: imágenes JPEG, imágenes TIFF, PDFs, Epubs, Archivos OCR, metadatos del PIA, metadatos del PIT, etc.

De esta manera se podrán descargar todos los archivos de los PIA seleccionados, así como sólo los archivos de los PIA seleccionados que tengan uno o varios formatos determinados.

4.2 Restauración de paquetes prePIT

Esta herramienta restaura los paquetes prePIT, esto es restaurar los datos tal como estaba antes del inicio del proceso de normalización transformación y carga en SIPREDI_SGCB. Esta herramienta permitirá a la SGCB recuperar la información tal como estaba en los medios de origen.

5 Equipamiento informático que soporta SIPREDI_SGCB

5.1 Sistema de copias con separación geográfica

El equipamiento informático que soporta SIPREDI_SGCB debe concebirse para mantener la confidencialidad y la seguridad informática, presente y futura. La SGCB debe disponer de una copia íntegra de sus paquetes de información para la recuperación de contenidos en caso de necesidad o como respaldo.

De los paquetes de información preservados se harán al menos cuatro copias con separación geográfica en sistemas de almacenamiento. Dos de estas copias serán conservadas en las instalaciones de la SGCB y custodiadas en sus instalaciones y las otras dos serán conservadas por la empresa contratista de los servicios de preservación (Administrador Externo) en un armario ignífugo, con clave de acceso, depositado en sus instalaciones con todas las medidas necesarias para su seguridad: alarma, lugar seguro y aislamiento del resto de las instalaciones de la empresa.

Todas las copias del sistema (los aparatos electrónicos y los contenidos) son propiedad de la SGCB, la empresa responsable del sistema deberá entregar las copias y los contenidos que conserve de SIPREDI_SGCB a la SGCB si ésta lo solicita.

5.2 Características del sistema

El sistema de almacenamiento debe ser RAID (*Redundant Array of Independent Disks*) o tecnologías más avanzadas. El nivel del tipo de RAID (mínimo RAID 10) debe garantizar:

- Protección de datos.
- La distribución de datos entre los discos.
- La tolerancia a fallos, la recuperación y el funcionamiento del sistema debe ser posibles aunque falle o deje de funcionar un disco.
- Alta tasa de rendimiento de transferencia de datos.
- Ofrecer la posibilidad de tener al menos cuatro copias completas de los contenidos a preservar.

5.3 Acceso al sistema

La empresa encargada del sistema transferirá los contenidos de los soportes que contienen los contenidos a preservar de la SGCB a los dispositivos de almacenamiento mencionados en los puntos anteriores y entregará, una vez termine el trabajo, los dispositivos que contengan una copia del sistema a la SGCB. Dicha copia de almacenamiento contendrá los discos duros necesarios para los contenidos preservados y ampliables para las futuras entregas de contenidos. Al sistema de almacenamiento lo acompaña un ordenador portátil en el que se ha instalado la aplicación y las herramientas necesarias para acceder a los contenidos, entre ellos un buscador de los contenidos preservados. La empresa encargada del sistema de preservación es la encargada de proporcionar este ordenador y mantenerlo actualizado y operativo.

El acceso al sistema de almacenamiento debe ser posible, aparte de con las herramientas propias del sistema, directamente desde el sistema operativo del portátil que se acompaña.

Las tareas habituales de consulta o recuperación de documentos se harán con un usuario con permiso solo de lectura.

El usuario con permiso de escritura estará en posesión del Administrador Externo y de la SGCB. Este usuario es necesario para realizar las tareas continuas y periódicas de preservación.

Descripción del Sistema de Preservación de las Bibliotecas Digitales de la Subdirección General de Coordinación Bibliotecaria

Por seguridad, las estaciones de discos no están conectadas a Internet. Además, el sistema de preservación se mantiene la mayor parte del tiempo desconectado de la red eléctrica. Cuando se desea utilizarlo se conecta temporalmente el portátil a la estación de discos.

ANEXO I Ejemplo del sistema de organización del depósito de preservación de SIPREDI_SGCB

Las mayúsculas para los nombres de las carpetas son para destacar pero no tienen por qué ser reales, entre paréntesis figura la explicación de la función de la carpeta o fichero.

CHECK (carpeta con el fichero general para comprobar todo el repositorio)

DATA

check_aip.txt (Fichero resumen de todos los paquetes BagIt. Explicado más arriba)

bag-info.txt

bagit.txt

manifest-md5.txt

tagmanifest-md5.txt

ADMIN_EXTERNO (carpeta con ficheros volcados por la empresa adjudicataria del Servicio: base datos, software propio, buscador de contenidos preservados, etc.)

SGCB

CONTENIDO_DISCOS_ORIGEN (carpeta que contiene exclusivamente ficheros de listados PDF y MDB sobre los contenidos de los discos externos USB donde se custodian los documentos y que están en las carpetas BASE)

2012-BVPH (carpeta identificativa de año captura y proyecto)

NombreDisco-660e8400-e29b-41d4-a716-446655440055 (contiene los ficheros PDF y Access que estaban en la carpeta base del disco. Esta carpeta inicia el empaquetamiento PIA de los ficheros base del primer disco. Es la carpeta padre del paquete PIA por lo que lleva un código UUID).

DATA (a partir de aquí se sigue la norma de empaquetamiento del servicio de repositorio y BagIt)

LOGS-DATOS-SIP

METADATOS_RECIBIDOS (como no hay metadatos sobre estos ficheros de listados se elimina)

OBJETOS

FicheroListado-970e8400-e29b-41d4-a716-446655440068.pdf

Base-609b0501-a38b-00f4-a716-990053642174.mdb

ETC.

**Descripción del Sistema de Preservación de las Bibliotecas Digitales de
la Subdirección General de Coordinación Bibliotecaria**

mets-*nombreDiscoEtiqueta1-660e8400-e29b-41d4-a716-446655440055.xml*

bag-info.txt

bagit.txt

manifest-md5.txt

tagmanifest-md5.txt

NombreDisco-660e8400-h38a-52f4-b816-436754040079 (IDEM)

[Etc...]

[ETC....]

2013-BVPH (IDEM)

2013-BVPB

(En las carpetas de BVPB queda todo igual a las carpetas de BVPH, excepto el directorio "OBJETOS" en el que se incluyen carpetas por biblioteca).

OBJETOS

Biblioteca1

FicheroListado-970e8400-e29b-41d4-a716-446655440068.pdf

Base-609b0501-a38b-00f4-a716-990053642174.mdb

ETC.

Biblioteca2

FicheroListado-88af9934-aa9f-51d4-a716-446655440068.pdf

Base-777b0501-a38b-00f4-a716-990053642174.mdb

ETC.

BVPB (esta carpeta contiene los ficheros de documentos y de sus metadatos propiamente dichos, ya no contiene los ficheros de listados de BASE. Esta carpeta contiene todos los documentos del proyecto BVPB)¹⁴.

¹⁴ Se podrá crear una carpeta hija, justo a continuación de la carpeta BVPB, que represente al centro que ha aportado los documentos siguiendo el mismo sistema de códigos de denominación creado para el proyecto BVPB.

**Descripción del Sistema de Preservación de las Bibliotecas Digitales de
la Subdirección General de Coordinación Bibliotecaria**

HD-000-PTJ-550e8400-e29b-41d4-a716-446655440000 (signatura de libro+UUID)
(empaquetamiento a nivel de libro. Si una obra libraria se presenta en un único volumen,
este nivel de libro se obvia. Este nivel se usa para poder aglutinar todos los volúmenes de
una obra con varios volúmenes).

bag-info.txt

bagit.txt

manifest-md5.txt // Contiene una línea sólo por cada fichero contenido en la carpeta
DATA a su mismo nivel.

tagmanifest-md5.txt

DATA

LOGS-DATOS-SIP

listado.txt // Sólo alude a los ficheros conservados en la carpeta
METADATOS_RECIBIDOS.

sip_estr_cr.txt // Representación de la estructura de carpetas del
SIP a nivel de obra.

tab_corp // Sólo alude a los ficheros conservados en la carpeta
METADATOS_RECIBIDOS.

METADATOS_RECIBIDOS (contiene los ficheros de metadatos a conservar
recibidos a nivel de libro con varios volúmenes)

METS- HD-000-PTJ-550e8400-e29b-41d4-a716-446655440000.xml // Es el
fichero METS a este nivel que referencia los METS propios de los volúmenes.

VOL_I-781e8400-a39b-41d4-a716-446655440013 (carpeta para el volumen I)
(empaquetamiento a nivel de volumen de libro)

DATA

LOGS-DATOS-SIP

id_form_fich.txt

listado.txt

sip_estr_cr.txt

tab_corp.txt

METADATOS_RECIBIDOS

MARC21

EDM

METS

METS-CARGA

OBJETOS

MÁSTERES

FICHERO-970e8400-e29b-41d4-a716-446655440068.TIF

ETC.

DERIVADOS

ALTO

JPEG

PDF

EPUB

mets-VOL_I-781e8400-a39b-41d4-a716-446655440013

bag-info.txt

bagit.txt

manifest-md5.txt

tagmanifest-md5.txt

VOL_II (carpeta para el volumen II) (empaquetamiento a nivel de volumen libro)

BVPH (carpeta para los documentos de la BVPH)¹⁵

LA_VOZ_DE_ALMERIA-440e8400-e29b-52f3-0716-566655447819 (ejemplo de diario. Esta es la carpeta padre por publicación periódica, para que sea más ilustrativo el ejemplo se ha usado el título del diario en lugar de un identificador)

bag-info.txt

bagit.txt

manifest-md5.txt // Contiene una línea sólo por cada fichero contenido en la carpeta DATA a su mismo nivel.

tagmanifest-md5.txt

DATA

LOGS-DATOS-SIP

¹⁵ Se podrá crear una carpeta hija, justo a continuación de la carpeta BVPH, que represente al centro que ha aportado los documentos siguiendo el mismo sistema de códigos de denominación creado para el proyecto BVPH.

listado.txt // Sólo alude a los ficheros conservados en la carpeta METADATOS_RECIBIDOS.

sip_estr_cr.txt // Representación de la estructura de carpetas del SIP a nivel de publicación periódica.

tab_corp // Sólo alude a los ficheros conservados en la carpeta METADATOS_RECIBIDOS.

METADATOS_RECIBIDOS (contiene los ficheros de metadatos a conservar recibidos a nivel de publicación periódica)

METS- LA_VOZ_DE_ALMERIA-440e8400-e29b-52f3-0716-566655447819.xml // Es el fichero METS a este nivel que referencia los METS propios de los números de la publicación.

1901 (carpeta para año)

01 (carpeta para mes)

NÚMERO DE PUBLICACIÓN-440e8400-e29b-41d4-a716-446655440110 (esta es la carpeta padre del PIA por documento físico que conforma el número de publicación periódica. Lleva su propio UUID)

DATA

LOGS-DATOS-SIP

id_form_fich.txt

listado.txt

sip_estr_cr.txt

tab_corp.txt

METADATOS_RECIBIDOS

MARC21

EDM

METS

METS-CARGA

OBJETOS

MÁSTERES

FICHERO-970e8400-e29b-
41d4-a716-
446655440068.TIF

ETC.

DERIVADOS

ALTO

JPEG

PDF

EPUB

mets-NUMERO DE PUBLICACIÓN -440e8400-
e29b-41d4-a716-446655440110

bag-info.txt

bagit.txt

manifest-md5.txt

tagmanifest-md5.txt

**NÚMERO DE PUBLICACIÓN-678a9400-e29b-41d4-a716-
442355422210**

ADELANTADO_ALMERIA-640e8400-e29b-52f3-0716-566655007595

PROGRESO_ALMERIENSE-940e8400-e29b-52f3-0716-566655447770

.....

ANEXO II Perfil de preservación digital de SEPREDI_SGCB

El perfil de preservación ajustado a los requerimientos y preferencias de los contenidos actuales es el siguiente:

OBJETOS DIGITALES

Tipo de medio	Formatos de fichero de entrada	Sistema de compresión admitido	Codificación de caracteres en textos	Formatos de preservación	Herramientas para la normalización aceptadas
Imágenes raster	TIFF	Ninguno preferible para color y grises, admisible sistema de amplio uso, pero sin pérdida	-	<p>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.</p> <p>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).</p> <p>No se admitirá TIFF con compresión JPEG o ZIP.</p> <p>No se admitirán imágenes en paleta de colores.</p>	<p>ImageMagick (http://www.imagemagick.org/)</p> <p>Adobe Photoshop CC o CS</p>
	JPG (JFIF)	JPG	-	Se mantiene como formato de preservación para los derivados, pero no para los másteres.	<p>JPGtoTIFF (GraphicsMagick)</p> <p>JPGtoTIFF (ImageMagick)</p> <p>ImageIO</p> <p>Adobe Photoshop CC o CS</p>
Portable Document Format	PDF	Cualquier sistema admitido por	Cualquier sistema admitido por	PDF/A en sus versiones 1, 2 y 3, y sus diferentes	<p>Ghostscript (http://www.ghostscript.com/)</p> <p>PDF/A Converter</p>

**Descripción del Sistema de Preservación de las Bibliotecas Digitales de
la Subdirección General de Coordinación Bibliotecaria**

Tipo de medio	Formatos de fichero de entrada	Sistema de compresión admitido	Codificación de caracteres en textos	Formatos de preservación	Herramientas para la normalización aceptadas
		el estándar aplicado.	el estándar aplicado.	niveles de conformidad PDF 1.4 o superior.	Adobe preflight Adobe Acrobat 9 o superior
Base de datos	MDB (versiones de Access anteriores a Access 2007)	-	-	accdb.	Microsoft Access en versiones que generan accdb y aún abren ficheros en MDB, pero en versiones posteriores a Access 2007 (2010, 2013, 2016). Se usará el sistema de conversión recomendando por el propio Microsoft ¹⁶ . La conversión y la validación se harán manualmente. Reflejándose en los metadatos PREMIS este hecho.
OCR	ALTO	--	Cualquier sistema admitido por el estándar XML. Preferible UTF-8.	ALTO	--
	EPUB	--	Cualquier sistema admitido por el estándar aplicado.	EPUB	--

¹⁶ Accesible en la página "Convert a database to the .accdb file format". Disponible en: <https://support.office.com/en-US/Article/Convert-a-database-to-the-accdb-file-format-098ddd31-5f84-4e89-8f44-db0cf7c11acd>

**Descripción del Sistema de Preservación de las Bibliotecas Digitales de
la Subdirección General de Coordinación Bibliotecaria**

METADATOS

Estándar	Versiones admitidas	Codificación de caracteres	Formatos de preservación	Herramienta para la normalización
METS	Versiones 1.3 y posteriores ¹⁷ .	Cualquier sistema admitido por el estándar XML. Preferible UTF-8	METS en versiones posteriores a 1.3.	--
PREMIS	Versiones 2.* y 3.*. Codificación en XML dentro de los ficheros METS o en fichero XML externo.	Cualquier sistema admitido por el estándar XML. Preferible UTF-8	PREMIS en versiones posteriores a 2.*.	--
MIX	Cualquier versión en codificación en XML dentro de los ficheros METS o en fichero XML externo.	Cualquier versión en codificación en XML dentro de los ficheros METS o en fichero XML externo.	MIX en su última versión.	--
MARCXML	Cualquier versión en codificación en XML dentro de los ficheros METS o en fichero XML externo.	Cualquier versión en codificación en XML dentro de los ficheros METS o en fichero XML externo.	Marc21 en codificación MARCXML válido	--
MODS	Cualquier versión en codificación en XML dentro de los ficheros METS o en fichero XML externo.	Cualquier versión en codificación en XML dentro de los ficheros METS o en	MODS	--

¹⁷ La versión 1.3 y las anteriores pueden tener problemas de compatibilidad con las últimas versiones en elementos de amplio uso. La versión 1.3 es de mayo de 2003.

**Descripción del Sistema de Preservación de las Bibliotecas Digitales de
la Subdirección General de Coordinación Bibliotecaria**

Estándar	Versiones admitidas	Codificación de caracteres	Formatos de preservación	Herramienta para la normalización
		fichero XML externo.		
EDM / ESE	Cualquier versión en codificación en XML dentro de los ficheros METS o en fichero XML externo.	Cualquier versión en codificación en XML dentro de los ficheros METS o en fichero XML externo.	EDM / ESE	--
Marc21	ISO 2709, MRC, MRK, TXT correctamente estructurado y convertible a MARCXML automáticamente.	Cualquier sistema admitido por el estándar aplicado.	ISO 2709 (.MRC) o MARCXML (.XML). No se admitirá MRK.	--

ANEXO III Buscador de contenidos preservados

En este anexo presentamos la tabla de mapeados de campos desde el formato MARC a la estructura de campos de la herramienta de búsqueda.

Los campos que están en negrita son indispensables, por lo que deben ser incluidos necesariamente en el sistema de búsqueda, siendo opcionales el resto.

Es preciso permitir el cruce de las búsquedas de los diferentes campos mediante operadores booleanos (Y/O/NOT).

Es preciso permitir realizar las búsquedas en todos los campos desde una única consulta.

Campo MARC	Campo de buscador	Notas
001, 017\$a, 020\$a, 022\$a, 024\$a, 028\$a, 030\$a, 035\$a	Número/código de control	
130\$a, 210\$a, 222\$a, 240\$a, 245\$a\$b, 246\$a\$b, 247\$a, 501\$a, 505 \$a\$t, 533\$a\$t. 534\$a\$t, 730\$a\$t, 740\$a\$p	Título	
100\$a, 110\$a\$b, 111\$a, 245\$c, 700\$a, 710\$a, 711\$a, 720\$a	Autor/colaborador	
260\$a\$e, 264\$a, 751\$a, 752\$a\$b\$c\$d 008/15-17	Lugar de publicación	Los valores del 008/15-17 son códigos normalizados, la búsqueda en este campo tendría que realizarse sobre una tabla con la correspondencia entre los valores normalizados y valores en lenguaje natural en español.
008/35-37	Lengua	Los valores del 008/35-37 son códigos normalizados, la búsqueda en este campo tendría que realizarse sobre una tabla con la correspondencia entre los valores normalizados y valores en lenguaje natural en español.
260\$a\$f, 264\$b	Editor	
008/07-10, 260\$c\$g, 264\$c	Fecha de publicación	
080, 082, 600\$a\$t, 610\$a\$b\$t, 611\$a, 630\$a, 648\$a, 650\$a, 651\$a, 655\$a, 662\$a\$b\$c\$d	Materia	
852\$a del registro de fondos o bibliográfico	Biblioteca	En los ficheros METS de BVPB y BVPH los códigos de biblioteca están en el registro de fondos (852\$a), que se encuentra a continuación del bibliográfico. Se deberán asociar (mediante una tabla formulario) los códigos a los nombres de las bibliotecas, para poder buscar mediante este campo nombres de biblioteca.
852\$j del registro de fondos o bibliográfico	Signatura	
Cabecera/06, 655\$a	Tipo de material	En cuanto a los valores de la cabecera, se hará un mapeo interno para poder buscar por los

**Descripción del Sistema de Preservación de las Bibliotecas Digitales de
la Subdirección General de Coordinación Bibliotecaria**

		valores de esta posición traducidos al español, pudiendo ser seleccionados los valores de un menú desplegable: texto impreso-a, música-c/d, mapas-e/f, medios proyectables-g, grabaciones sonoras-i/j, material gráfico--k, manuscritos-t
500\$a, 502\$a, 504\$a, 510\$a, 520\$a, 522\$a, 545\$a, 547\$a, 561\$a, 563\$a, 585\$a, 586\$a	Descripción	
[Nombre de carpeta]	Año de digitalización	
	Fecha de alta	Fecha de entrada en el sistema de preservación