



CENTRO ARGENTINO DE INFORMACIÓN CIENTÍFICA Y TECNOLÓGICA

# **Buenas Prácticas para la Preservación Digital de Archivos Históricos**

Mariano Collado

Natalia Efrón

Versión 2017

## Contenido

1. Normalización de nombres de archivo	pág. 2
2. Normalización de formatos	pág. 2
3. Identificación de formatos	pág. 3
4. Conversión de formatos	pág. 3
5. Metadatos de preservación	pág. 4
6. Software de preservación	pág. 5
7. Estrategias de preservación	pág. 5

Entendemos la preservación digital como todas aquellas acciones destinadas a garantizar el acceso a largo plazo a los documentos digitales. En este informe sistematizamos los resultados del trabajo realizado en los últimos años (2011-2016), que consideramos será una herramienta de utilidad para abordajes futuros en términos metodológicos.

## 1. Normalización de nombres de archivo

Todos los archivos digitales deben tener un número o código identificador único, que debe seguir un criterio lógico de conformación basado en el Cuadro de clasificación del fondo archivístico.

Se recomienda estandarizar con prefijos que refieran a la institución, colección y/o serie (de mayor a menor jerarquía), más una codificación de números sucesivos.

En el caso de que los archivos estén nombrados previamente por el productor es preferible consignar el nombre dentro de los metadatos del documento y asignarle un nuevo nombre con un criterio lógico.

Ejemplo 1: Fondo Rex González

Documento: <b>REX-FN1000.jpg</b>		
Prefijo de colección: REX	Sufijo de sección: FN	Numeración consecutiva del 1 al infinito
Productor Rex González	FN: Foto Negativos	Número de negativo de la Sección Fotográfica: 1000

Ejemplo 2: Colección Cordeu

Documento: <b>CORDEU-TCAMPO-1-1.mp3</b>		
Prefijo de colección: CORDEU	Sufijo de sección: TACMPO	Numeración consecutiva del 1 al infinito
Edgardo Cordeu	TCAMPO: Trabajo de Campo	Numero de audio del trabajo de campo: 1

## 2. Normalización de formatos

La elección del formato digital en el que se representarán los contenidos dependerá de la tipología documental de los mismos (texto, imagen, audio), y de la finalidad de uso asignada (preservación a largo plazo o exposición en web). Para cada tipología documental existen determinados formatos estándar a nivel internacional que es recomendable adoptar.

Los formatos de preservación, no comprimen la información, no generan modificaciones en los archivos y son independientes del software de lectura, sin embargo, para facilitar la exploración y uso de los archivos digitales se recomienda realizar un copia de menor calidad (menor tamaño) para facilitar su circulación y transferencia.

A continuación se consignan los formatos de preservación (archivo permanente) y acceso (uso) recomendados:

Tipología documental	Formatos de preservación	Formatos de acceso
Texto	PDF/A (ISO 19005) (*.pdf)	PDF PDF/A
Imagen	TIFF (sin comprimir) (*.tif)	JPEG PNG
Audio	WAV (96kHz 24bit PCM) (*.wav)	MP3

### 3. Identificación de formatos

A veces el formato real de un archivo digital no coincide con el que se nos informa. Todos los archivos destinados a un repositorio digital deben ser analizados mediante una herramienta de identificación para comprobar la identidad de su formato. Si se detectasen archivos que tienen un formato distinto al esperado, deben volver a convertirse mediante algún otro método fiable que produzca el formato deseado.

Existe un software llamado DROID, que utiliza el registro de formatos PRONOM (U.K. Archives), que es capaz de verificar y caracterizar el formato de los archivos.

### 4. Conversión de formatos

Para el caso en el que los documentos de origen sean digitales (*born digital*), pero no se ajusten a los formatos de archivo estándar se debe hacer una conversión de formato: migrar el contenido del documento de un formato a otro. Existen diversas aplicaciones para realizar esta operación, pero no todas son recomendables.

Hay varias maneras de convertir un documento de texto a formato PDF, pero no todas ellas son válidas a los efectos de la preservación. No todas disponen de la opción de conversión a PDF/A e incluso muchas de las que sí la tienen no producen un auténtico PDF/A, aunque así lo afirmen. Para el software de identificación de formatos DROID, que escanea los documentos e informa sobre su verdadera conformación, el único conversor que crea PDF/A auténticos es *PDF Creator 2.X* (es decir, versiones igual o superiores a la 2.0).

Un dato muy importante a tener en cuenta para convertir un documento en PDF/A es que se debe partir del documento original (Word u otro) y nunca desde un PDF intermedio, ya que esto genera un enorme tamaño de archivo final.

En el caso de partir de un archivo de imagen JPEG, o de un archivo de audio MP3, se recomienda no convertir a TIFF o a WAV, respectivamente, ya que se estaría modificando el nombre de la extensión pero no la estructura interna del documento.

Se recomienda trabajar siempre sobre copias y no sobre los archivos de resguardo, aún cuando los mismos no tengan las características recomendadas para su preservación.

## 5. Metadatos de preservación

Existe un conjunto de metadatos (información del archivo digital) que es necesario adjuntar al documento de archivo con el fin de facilitar las futuras migraciones que se necesiten para poder seguir accediendo a su contenido. Los metadatos de preservación se introducen mediante el software archivístico que se esté utilizando (Ejemplo: AtoM), y forman parte del conjunto de metadatos que conforman el registro documental del objeto digital.

A continuación se reseñan los metadatos básicos de preservación:

- Código de formato PUID de PRONOM

Mediante el análisis efectuado por DROID se puede conocer el formato del archivo digital según la clasificación codificada PUID, elaborada por los Archivos Nacionales del Reino Unido. La base de datos PRONOM tiene identificados y caracterizados la enorme mayoría de los formatos digitales existentes y sus diferentes versiones. Esta información de nuestro archivo permitirá optimizar los futuros procesos de migración de formato necesarios para alargar su vida.

- Suma de verificación del objeto digital basado en algoritmos como MD5 o SHA-1

Cada archivo digital puede ser representado por una única secuencia alfanumérica, distinta de la de cualquier otro archivo. Esta relación unívoca entre un objeto digital y su suma de verificación ó *checksum*, permite comprobar en cualquier momento la integridad del documento simplemente comparando su *checksum* actual con el *checksum* asociado al momento de su ingesta en el repositorio. La alteración de un *checksum* en el futuro puede ser la alarma que nos indica la necesidad de migrar el formato afectado.

- Modificaciones de archivo: tipo de evento, fecha, formato origen, formato destino, etc.

Con el paso del tiempo la obsolescencia de los formatos de archivo nos obliga a actualizarlos para seguir permitiéndonos el acceso al documento. Cada migración de formato, cada cambio operado sobre el archivo digital debe ser debidamente documentado para conocer su historial y saber gestionar otros cambios futuros.

- Relaciones del objeto digital con otros objetos digitales del repositorio

Conocer los vínculos que algunos archivos digitales tienen entre sí dentro del repositorio nos brinda una información extra para interpretar y resolver futuras posibles pérdidas de datos.

## 6. Software de preservación

La preparación del paquete de información (documento + metadatos) a archivar es fundamental, ya que de las características y la calidad del mismo depende la futura disponibilidad de la información que contiene. Existen una serie de actuaciones (micro servicios) que se realizan sobre los archivos a almacenar (verificación, conversión, generación de *checksum*, etc) que se llevan a cabo mediante distintos softwares.

El software de preservación Archivemática realiza de forma automatizada o semi-automatizada todos los tratamientos requeridos para la generación del paquete de información de archivo.

A continuación se reseñan algunas aplicaciones de preservación:

- DROID

caracteriza el archivo brindando una serie de parámetros. El formato se presenta bajo la codificación PUID, perteneciente al registro de formatos PRONOM. También genera el checksum de cada archivo analizado.

<http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>

- SendTo MD5

Software libre de generación de *checksum* de fácil aplicación a través del menú desplegable de cada archivo digital. Existen muchos similares.

<http://vaultec.mbnet.fi/software.php>

- Archivemática

Software libre que procesa de forma automática o semi-automática los archivos provenientes del productor y los deja en las condiciones óptimas para su ingreso en el repositorio digital.

<https://www.archivematica.org/es/>

## 7. Estrategias de preservación

La preservación es una actuación continua que requiere de una actitud vigilante y activa por parte de los responsables de un archivo o repositorio digital. La aplicación de estrategias para mitigar o contrarrestar la obsolescencia informática (hardware y software) debe ser de cumplimiento obligado y estricto si se desea evitar la pérdida de información.

A continuación se consignan las estrategias más habituales:

- Migración

Migrar los datos de un formato a otro con el objetivo de usar siempre formatos vigentes. Almacenar información en formatos o softwares obsoletos no permite garantizar la preservación.

- Emulación

Emular softwares en desuso permite acceder a la información de archivos en formatos obsoletos. Se recomienda realizar todos los pasos anteriores para no llegar a tener que hacer uso de esta estrategia.

- Actualización

Actualizar periódicamente la información de los soportes de almacenamiento (cd/dvd/pen drive/discos externos) antes que se vuelvan obsoletos.

- Replicación

Replicar la información en dos espacios de almacenamiento (discos rígidos y o servidores). Se busca con esta estrategia garantizar la existencia de una copia de seguridad frente a una falla impredecible.