

Estándares de digitalización: requerimientos mínimos (actualizado en octubre de 2010)

1. Antecedentes

En 1999 se creó un grupo de trabajo del CBUC con el objetivo que éste elaborara unas recomendaciones sobre estándares básicos para la digitalización de diferentes tipos de documentos (libros, revistas, material gráfico, etc.) con diferentes finalidades (preservación y consulta). Este grupo elaboró el documento “Estándares de digitalización: elementos mínimos” que se difundió en febrero de 2000. En julio de 2005 la Biblioteca de Catalunya hizo una revisión de vigencia basándose en proyectos de digitalización de referencia.

Han pasado cinco años más y de nuevo se ha procedido a una revisión para mantener actualizados los estándares. De los referentes anteriores algunos no han revisado sus estándares, motivo por el cual se han consultado documentos y pautas más recientes de proyectos e instituciones de prestigio. En el punto 4 encontraréis las referencias consultadas.

Como resultado de este estudio comparativo se desprende que los estándares de digitalización elaborados en el año 2000 y revisados en 2005 son básicamente vigentes.

En los últimos años ha surgido algún formato nuevo como el JPEG2000 que se está valorando como alternativa al TIFF a efectos de preservación para cuestiones de sostenibilidad y viabilidad (menos ocupación en disco), con todo, no todas las fuentes lo han aceptado abiertamente aún.

Se han incorporado al documento los estándares mínimos para documentos sonoros y vídeos a partir de las recomendaciones de la IASA, que son aceptadas internacionalmente.

Encontraréis las modificaciones y adiciones marcadas en verde.

2. Estándares de digitalización: requerimientos mínimos

Esta propuesta se limita a los dos usos más frecuentes que los documentos digitalizados tienen en las bibliotecas: preservación y acceso por Internet. Por lo tanto, los datos mínimos recomendados en cada caso se han establecido según las diferentes necesidades: conservar la copia exacta del documento original (preservación) o bien ofrecer un acceso rápido y de buena calidad a ficheros de texto y/o imagen alojados en servidores (acceso por Internet).

No se han analizado en este documento otros usos que podríamos considerar subproductos de los dichos anteriormente (uso comercial o reproducción, proyección en pantalla grande, consulta en red local, etc.). En cualquier caso se recomienda producir siempre un fichero master inicial que se debe guardar siempre para preservar una copia fiel del original, y de este extraer el resto de copias y subproductos.

A continuación se detallan los requerimientos mínimos propuestos para la digitalización de documentos con los objetivos de preservación y acceso.

2.1. Preservación

TIPO DE DOCUMENTO	FORMATO	RESOLUCIÓN	DEFINICIÓN	COMPRESIÓN
Texto impreso (libros, publicaciones periódicas, partituras...)	TIFF preferente JPEG2000 como 2ª opción	300 ppp Recomendado 400 ppp si se tiene que aplicar OCR	Original b/n: 8 bits escala de grises Original color: 24 bits color	Sin compresión Comprimido sin pérdida
Manuscritos	TIFF preferente JPEG2000 como 2ª opción	300 ppp	Original b/n: 8 bits escala de grises Original color: 24 bits color	Sin compresión Comprimido sin pérdida
Mapas	TIFF preferente JPEG2000 como 2ª opción	500 ppp	Original b/n: 8 bits escala de grises Original color: 24 bits color	Sin compresión Comprimido sin pérdida
Fotografías	TIFF preferente JPEG2000 como 2ª opción	500 ppp	Original b/n: 8 bits escala de grises Original color: 24 bits color	Sin compresión Comprimido sin pérdida
Diapositivas	TIFF preferente JPEG2000 como 2ª opción	500 ppp	Original b/n: 8 bits escala de grises Original color: 24 bits color	Sin compresión Comprimido sin pérdida
Grabación sonora	WAVE or BWF	96 Khz	24 bits	Sin compresión
Audiovisual	MPEG2 Vídeo MPEG1 Audio preferente AVI como 2ª opción vídeo	720x576 (4:3) 48 Khz	9400Kbps 25.00fps 16 bits	Sin compresión

2.2. Acceso

TIPO DE DOCUMENTO	FORMATO	RESOLUCIÓN	DEFINICIÓN	COMPRESIÓN
Texto impreso (libros, publicaciones periódicas, partituras...)	JPEG, GIF, PDF	JPEG: 150 ppp GIF: 72 ppp	JPEG: 24 bits GIF: 8 bits	JPEG: baja GIF: estándar
Manuscritos	JPEG, GIF, PDF	JPEG: 150 ppp GIF: 72 ppp	JPEG: 24 bits GIF: 8 bits	JPEG: baja GIF: estándar
Mapas	JPEG, PDF	300 ppp	24 bits o 8 bits	Compresión baja
Fotografías	JPEG, PDF	300 ppp	24 bits o 8 bits	Compresión baja
Diapositivas	JPEG, PDF	300 ppp	24 bits o 8 bits	Compresión baja
Grabación sonora	MP3	44.1 khz	128Kbps ; 192 Kbps [vel. transferencia]	Compresión baja (calidad CD)
Audiovisual	MPEG2 Vídeo MPEG1 Audio	720x576 (4:3) 48 Khz	9400Kbps 25.00fps 16 bits	

NOTA:

PDF es un formato de visualización que adopta las características (resolución, definición, compresión) del formato original (JPEG, GIF, etc.)

2.3. Recomendaciones y consideraciones

CONSIDERACIONES SOBRE LO QUE IMPLICA DIGITALIZAR UN DOCUMENTO:

- Sería deseable que los documentos se digitalizasen sólo una vez (sobretudo en el caso de material antiguo o valioso). Por tanto, se debe tener cuidado en la creación del fichero master (el de preservación) y crear las copias necesarias a partir de él.
- Es importante establecer un procedimiento interno para hacer el control de calidad del material recibido ya digitalizado.
- Se debe tener en cuenta que a menudo es necesaria una manipulación de la imagen posterior a la digitalización (de cara a su consulta) y que esta labor se debe prever.
- En el caso de documentos (antiguos, principalmente) que se deban restaurar, y dependiendo de la actuación a realizar, se debe valorar si se digitaliza antes o después de la restauración (por ejemplo, la laminación hace que se pierda contraste). En cambio, en el caso de sonoros y vídeos siempre se debe hacer la restauración a priori.

RECOMENDACIONES TÉCNICAS:

- En la medida de lo posible, se recomienda alejarse de soluciones propietarias y cerradas; es mejor usar formatos estándares y sistemas abiertos.
- Se recomienda un grado de compresión bajo para mantener la calidad de la imagen (con una compresión alta se puede perder información). También se puede usar compresión binaria.
- Se recomienda que el proceso de digitalización se ajuste a la medida del original al 100%.
- Se recomienda que se especifiquen las características de la digitalización para cada tipo de documento o parte del documento, por ejemplo un documento que contenga texto y láminas.
- Se recomienda valorar para cada tipo de documento textual la realización de OCR para disponer de un índice de palabras que se pueda buscar.
- Los documentos en los que la dificultad de lectura del texto o bien la calidad de la parte gráfica lo haga recomendable, se deberían escanear preferentemente en escala de grises.

RECOMENDACIONES SOBRE EL ENCARGO A EMPRESAS EXTERNAS:

- Se recomienda muy especialmente hacer un listado de especificaciones técnicas para entregar a la empresa o persona que lleve a cabo la digitalización. Si se debe determinar el plazo de entrega, la responsabilidad sobre el material entregado, si la digitalización se hará en la biblioteca o en la empresa (incide normalmente en el coste), etc.

- Se recomienda incluir en el listado como requerimiento la entrega por parte de la empresa de los metadatos técnicos y de preservación de acuerdo con las pautas internacionales.
- Es recomendable especificar si el documento a digitalizar se puede o no desencuadernar para obtener hojas sueltas (si se puede desencuadernar es más fácil y más económico de digitalizar).
- Es recomendable hacer una prueba inicial que consista en escanear bajo diferentes niveles de resolución, de tonos y de compresión una muestra de los documentos tanto en blanco y negro como en color, para acabar de ajustar parámetros.
- El proveedor debe garantizar que utilizará sistemas de recuperación y visualización de la información estándares, compatibles con los que usa la biblioteca.
- La biblioteca debe documentar y comunicar al proveedor cómo desea que se nombren los ficheros que se generarán, así como la estructura de los directorios. Estos nombres deben ser unívocos, pues deben identificar cada documento, y se debe evitar crear conflictos con otros ficheros cuando se vuelquen en el servidor, por ejemplo.

3. Glosario

Formato: Norma que describe cómo estructurar los datos para generar un fichero gráfico.

Resolución: Indica el nombre de píxeles que forman una imagen. Se mide en ppp (puntos por pulgada), en inglés dpi (dots per inch).

Definición: Permite establecer el nombre de colores (incluyendo el blanco/negro y la escala de grises) que queremos que aparezcan en el documento. Se mide en bits por píxel.

Compresión: Reducción de la medida del fichero de la imagen mediante algoritmos matemáticos que eliminan información redundante en la definición del gráfico. Esta técnica permite almacenar o transmitir estos ficheros con más rapidez y ocupando menos espacio.

4. Fuentes consultadas

- The Library of Congress. *American Memory. Technical Standards for Digital Conversion Of Text and Graphic Materials* (2006, rev. 2007). <http://memory.loc.gov/ammem/about/techStandards.pdf>
- National Library of Australia. *Image capture standards*. <http://www.nla.gov.au/digital/capture.html>
- British Library *Endangered Archives Programme. Guidelines for copying archival material* (2009). http://www.bl.uk/about/policies/endangeredarch/pdf/copying_guidelines.pdf
- Harvard University. *Recommended File Formats for Deposit into the DRS* (last revised 2009). <http://hul.harvard.edu/ois/digpres/guidance.html>
- National Library of the Netherlands. *Alternative File Formats for Storing Master Images of Digitisation Projects* (2008). [http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/Alternative File Formats for Storing Masters 2 1.pdf](http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/Alternative_File_Formats_for_Storing_Masters_2_1.pdf)
- International Association of Sound and Audiovisual Archives. *Guidelines on the Production and Preservation of Digital Audio Objects. IASA TC04* (2009). <http://www.iasa-web.org/tc04/audio-preservation>