



Estàndards de digitalització

Elements mínims. Proposta del CBUC

(actualitzat a juliol 2005)

Aquest document actualitza el que es va elaborar al febrer del 2000. Les modificacions que s'han realitzat s'assenyalen amb una línia vertical en el marge dret.

1. Estàndards de digitalització: elements mínims

Aquesta proposta es limita als dos usos més freqüents que els documents digitalitzats tenen a les biblioteques: preservació i consulta per Internet. Per tant, les dades mínimes recomanades en cada cas s'han establert segons les diferents necessitats: conservar la còpia exacta del document original (preservació) o bé donar un accés ràpid i de bona qualitat a fitxers de text i/o imatge allotjats en servidors (consulta per Internet).

No s'han analitzat en aquest document altres usos que podríem considerar subproductes dels esmentats anteriorment (ús comercial o reproducció, projecció en pantalla gran, consulta en xarxa local, etc.). Aquests subproductes es poden obtenir del fitxer que l'escàner dona per defecte (amb extensió .tiff) i que es recomana guardar sempre per tal de preservar una còpia fidel de l'original.

Els tipus de documents que es contemplen en aquest document són els que bàsicament té qualsevol biblioteca. S'han deixat fora de la proposta els documents en format microfitxa, vídeo, àudio i similars.

A continuació es detallen els elements mínims proposats per a la digitalització de documents amb els objectius de preservació i consulta per Internet:

1.1. Preservació

TIPUS DE DOCUMENT	FORMAT	RESOLUCIÓ	DEFINICIÓ	COMPRESSIÓ
Llibre antic (fins a 1940)	TIFF	300 dpi	Original b/n: 8 bits Original color: 24 bits	Sense compressió
Llibre modern (a partir de 1940)	TIFF	300 dpi	Original b/n: 8 bits Original color: 24 bits	Sense compressió
Publicacions periòdiques	TIFF	300 dpi	Original b/n: 8 bits Original color: 24 bits	Sense compressió
Material d'arxiu (manuscrit)	TIFF	300 dpi	Original b/n: 8 bits Original color: 24 bits	Sense compressió
Mapes	TIFF	500 dpi	Original b/n: 8 bits Original color: 24 bits	Sense compressió
Fotografies	TIFF	500 dpi	Original b/n: 8 bits Original color: 24 bits	Sense compressió
Diapositives	TIFF	500 dpi	Original b/n: 8 bits Original color: 24 bits	Sense compressió

NOTA:

El Material docent no s'ha considerat en aquest punt, doncs està format per materials que tenen un interès temporal (semestre, any acadèmic).

1.2. Consulta per Internet

TIPUS DE DOCUMENT	FORMAT	RESOLUCIÓ	DEFINICIÓ	COMPRESSIÓ
Llibre antic (fins a 1940)	JPEG, GIF, PDF	JPEG: 150 dpi GIF: 72 dpi	JPEG: 24 bits GIF: 8 bits	JPEG: baixa GIF: estàndard
Llibre modern (a partir de 1940)	JPEG, GIF, PDF	JPEG: 150 dpi GIF: 72 dpi	JPEG: 24 bits GIF: 8 bits	JPEG: baixa GIF: estàndard
Publicacions periòdiques	JPEG, GIF, PDF	JPEG: 150 dpi GIF: 72 dpi	JPEG: 24 bits GIF: 8 bits	JPEG: baixa GIF: estàndard
Material d'arxiu (manuscrit)	JPEG, GIF, PDF	JPEG: 150 dpi GIF: 72 dpi	JPEG: 24 bits GIF: 8 bits	JPEG: baixa GIF: estàndard
Mapes	JPEG, PDF	300 dpi	24 bits	Compressió baixa
Fotografies	JPEG, PDF	300 dpi	24 bits	Compressió baixa
Diapositives	JPEG, PDF	300 dpi	24 bits	Compressió baixa
Material docent	JPEG, GIF, PDF, HTML, WORD, ASCII	JPEG: 50 dpi (text) i 300 dpi (imatges) GIF: 72 dpi	JPEG: 24 bits GIF: 8 bits	JPEG: baixa GIF: estàndard

NOTA:

PDF és un format de visualització que adopta les característiques (resolució, definició, compressió) del format original (JPEG, GIF, etc.)

1.3. Recomanacions i consideracions

CONSIDERACIONS SOBRE EL QUE IMPLICA DIGITALITZAR UN DOCUMENT:

- Seria desitjable que els documents es digitalitzéssin un cop només (sobretot en el cas de material antic o valuós). Per tant, cal tenir cura en la creació del fitxer màster (el de preservació) i crear les còpies necessàries a partir d'ell.
- És important establir un procediment intern per a fer el control de qualitat del material rebut ja digitalitzat.
- Cal tenir en compte que sovint cal una manipulació de la imatge posterior a la digitalització (de cara a la seva consulta) i que aquesta tasca s'ha de preveure.
- En el cas de documents (antics, principalment) que s'hagin de restaurar, i depenent de l'actuació a fer, cal valorar si s'ha de digitalitzar abans o després de la restauració (per exemple, la laminació fa que es perdi contrast).

RECOMANACIONS TÈCNIQUES:

- En la mesura del possible, es recomana fugir de solucions propietàries i tancades; és millor usar formats estàndards i sistemes oberts.
- Es recomana un grau de compressió baix per a mantenir la qualitat de la imatge (amb una compressió alta es pot perdre informació). També es pot usar compressió binària.
- Es recomana que el procés de digitalització s'ajusti a la mida de l'original al 100%.
- Es recomana que s'especifiquin les característiques de la digitalització per a cada tipus de document o part del document, per exemple un document que contingui text i làmines.
- Es recomana valorar per a cada tipus de document la realització d'OCR per disposar d'un índex de paraules cercable.
- Els documents en què la dificultat de lectura del text o bé la qualitat de la part gràfica ho faci recomanable, s'hauria d'escanejar preferentment en escala de grisos.

RECOMANACIONS SOBRE L'ENCÀRREC A EMPRESES EXTERNES:

- Es recomana molt especialment fer una fitxa tècnica de cada sèrie digitalitzada, deixant per escrit informació relativa al procés (format, resolució, qui ha digitalitzat la sèrie, quan, etc.).
- També es considera molt recomanable redactar un plec d'especificacions tècniques per a lliurar a l'empresa o persona que dugui a terme la digitalització. S'hi ha de determinar el termini d'entrega, la responsabilitat sobre el material lliurat, si la digitalització es farà a la biblioteca o a l'empresa (incideix normalment en el cost), etc.

- Es recomanable especificar si el document a digitalitzar es pot o no desenquadrarnar per obtenir fulls solts (si es pot desenquadrarnar és més fàcil i més econòmic de digitalitzar)
- Es recomanable fer una prova inicial que consisteixi en escanejar sota diversos nivells de resolució, de tons i de compressió una mostra dels documents tant en blanc i negre com en color, per acabar d'ajustar paràmetres.
- El proveïdor ha de garantir que emprerà sistemes de recuperació i visualització de la informació estàndards, compatibles amb els que usa la biblioteca.
- La biblioteca ha de documentar i comunicar al proveïdor com vol que s'anomenin els fitxers que es generaran, així com l'estructura dels directoris. Aquests noms han de ser únics, doncs han d'identificar cada document, i s'ha d'evitar crear conflictes amb altres fitxers quan es bolquin al servidor, per exemple.

2. Glossari

Format: Norma que descriu com estructurar les dades per a generar un fitxer gràfic.

Resolució: Indica el nombre de píxels que formen una imatge. Es mesura en dpi (dots per inch), que en català tradueix per ppp (punts per polzada).

Definició: Permet establir el nombre de colors (incloent el blanc/negre i l'escala de grisos) que volem que apareguin al document. Es mesura en bits per píxel.

Compressió: Reducció de la mida del fitxer de la imatge mitjançant algorismes matemàtics que eliminen informació redundant en la definició del gràfic. Aquesta tècnica permet emmagatzemar o transmetre aquests fitxers amb més rapidesa i ocupant menys espai.