



Consorci de
Biblioteques Universitàries
de Catalunya

Estàndards de digitalització: requeriments mínims (actualitzat a octubre 2010)

1. Antecedents

Al 1999 es va crear un grup de treball del CBUC amb l'objectiu que aquest elaborés unes recomanacions sobre estàndards bàsics per a la digitalització de diferents tipus de documents (llibres, revistes, material gràfic, etc.) amb diferents finalitats (preservació i consulta). Aquest grup va elaborar el document "Estàndards de digitalització: elements mínims" que es va difondre al febrer del 2000. El juliol del 2005 la Biblioteca de Catalunya en va fer una revisió de vigència basant-se en projectes de digitalització de referència.

Han passat cinc anys més i de nou s'ha procedit a una revisió per tal de mantenir actualitzats els estàndards. Dels referents anteriors alguns no han revisat els seus estàndards, motiu pel qual s'han consultat documents i pautes més recents de projectes i institucions de prestigi. Al punt 4 trobareu les referències consultades.

Com a resultat d'aquest estudi comparatiu se'n desprèn que els estàndards de digitalització elaborats l'any 2000 i revisats el 2005 són bàsicament vigents.

Els darrers anys ha sorgit algun format nou com és el JPEG2000 que s'està valorant com alternativa al TIFF a efectes de preservació per qüestions de sostenibilitat i viabilitat (menys ocupació en disc), amb tot, no totes les fonts l'han acceptat obertament encara.

S'han afegit al document els estàndards mínims per a documents sonors i audiovisuals a partir de les recomanacions de la IASA, que són acceptades internacionalment.

Trobareu les modificacions i addicions assenyalades en verd.

2. Estàndards de digitalització: requeriments mínims

Aquesta proposta es limita als dos usos més freqüents que els documents digitalitzats tenen a les biblioteques: preservació i accés 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 (accés 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.). En qualsevol cas es recomana produir sempre un fitxer màster inicial que cal guardar sempre per tal de preservar una còpia fidel de l'original, i d'aquest extreure'n la resta de còpies i subproductes.

A continuació es detallen els requeriments mínims proposats per a la digitalització de documents amb els objectius de preservació i accés.

2.1. Preservació

TIPUS DE DOCUMENT	FORMAT	RESOLUCIÓ	DEFINICIÓ	COMPRESSIÓ
Text imprès (llibres, publicacions periòdiques, partitures...)	TIFF preferent JPEG2000 com a 2a opció	300 ppp Recomanat 400 ppp si s'hi ha d'aplicar OCR	Original b/n: 8 bits escala de grisos Original color: 24 bits color	Sense compressió Comprimat sense pèrdua
Manuscrits	TIFF preferent JPEG2000 com a 2a opció	300 ppp	Original b/n: 8 bits escala de grisos Original color: 24 bits color	Sense compressió Comprimat sense pèrdua
Mapes	TIFF preferent JPEG2000 com a 2a opció	500 ppp	Original b/n: 8 bits escala de grisos Original color: 24 bits color	Sense compressió Comprimat sense pèrdua
Fotografies	TIFF preferent JPEG2000 com a 2a opció	500 ppp	Original b/n: 8 bits escala de grisos Original color: 24 bits color	Sense compressió Comprimat sense pèrdua
Diapositives	TIFF preferent JPEG2000 com a 2a opció	500 ppp	Original b/n: 8 bits escala de grisos Original color: 24 bits color	Sense compressió Comprimat sense pèrdua
Enregistrament sonor	WAVE or BWF	96 Khz	24 bits	Sense compressió
Audiovisual	MPEG2 Vídeo MPEG1 Àudio preferent AVI com a 2a opció vídeo	720x576 (4:3) 48 Khz	9400Kbps 25.00fps 16 bits	Sense compressió

2.2. Accés

TIPUS DE DOCUMENT	FORMAT	RESOLUCIÓ	DEFINICIÓ	COMPRESSIÓ
Text imprès (llibres, publicacions periòdiques, partitures...)	JPEG, GIF, PDF	JPEG: 150 ppp GIF: 72 ppp	JPEG: 24 bits GIF: 8 bits	JPEG: baixa GIF: estàndard
Manuscrits	JPEG, GIF, PDF	JPEG: 150 ppp GIF: 72 ppp	JPEG: 24 bits GIF: 8 bits	JPEG: baixa GIF: estàndard
Mapes	JPEG, PDF	300 ppp	24 bits o 8 bits	Compressió baixa
Fotografies	JPEG, PDF	300 ppp	24 bits o 8 bits	Compressió baixa
Diapositives	JPEG, PDF	300 ppp	24 bits o 8 bits	Compressió baixa
Enregistrament sonor	MP3	44.1 khz	128Kbps ; 192 Kbps [vel. transferència]	Compressió baixa (qualitat CD)
Audiovisual	MPEG2 Vídeo MPEG1 Àudio	720x576 (4:3) 48 Khz	9400Kbps 25.00fps 16 bits	

NOTA:

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

2.3. 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), en canvi en el cas de sons i vídeo sempre s'ha de fer la restauració a priori.

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 textual 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'haurien d'escanejar preferentment en escala de grisos.

RECOMANACIONS SOBRE L'ENCÀRREC A EMPRESES EXTERNES:

- Es considera molt recomanable redactar un plec d'especificacions tècniques per a lliurar a l'empresa o persona que dugui a terme la digitalització. Si s'ha de determinar el termini de lliurament, 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 recomana incloure en el plec com a requeriment el lliurament per part de l'empresa de les metadades tècniques i de preservació d'acord amb les pautes internacionals.

- 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 emprará 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.

3. 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 ppp (punts per polzada), en anglès dpi (dots per inch).

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.

4. Fonts consultades

- 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%20links%20en%20publicaties/publicaties/Alternative%20File%20Formats%20for%20Storing%20Masters%202%201.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>