

## RECOMANACIONS TÈCNIQUES PER A LA DIGITALITZACIÓ DE DOCUMENTS

Els dos motius més freqüents per digitalitzar una col·lecció són la difusió i la preservació. I aquests són els objectius del dipòsit digital de la Diputació de Barcelona, tant difondre les col·leccions locals com preservar-les (originals en paper).

### 1. Punts a tenir en compte per unes bones pràctiques en la digitalització de documents

La digitalització s'ha de fer com a mínim a 300 ppp (punts per polzada), 400 ppp si s'aplica OCR

a) Eines necessàries:

- Escàner.
- Programari OCR per reconeixement de text. Programari recomanable, ABBYY FineReader 11 Professional, eina fàcil d'utilitzar, el seu reconeixement de caràcters és molt òptim i és econòmic, o bé, Omnipage 17 eina amb moltes prestacions, és ràpid i molt fiable.

b) Formats:

- Textos, revistes, llibres, escanejar-los en format PDF (Adobe Reader v7 o superior o Adobe Acrobat Standard) o epub.
- Imatges en format gif, jpeg, png o tiff.
- Àudio en format mp3, wav, vma.

Formats per mantenir la preservació dels documents (conservar la còpia exacta dels documents originals):

TIPUS DE DOCUMENT	FORMAT	RESOLUCIÓ	DEFINICIÓ	COMPRESSIÓ
Text imprès, (llibres, publicacions periòdiques, partitures...)	TIFF preferentment JPEG2000 com a 2a opció	300ppp, 400ppp quan s'hagi de passar l'OCR	Original b/n: 8 bits escala de grisos Original color: 24 bits color	Sense compressió Comprimit sense pèrdua
Manuscrit	TIFF preferentment JPEG2000 com a 2a opció	300ppp	Original b/n: 8 bits escala de grisos Original color: 24 bits color	Sense compressió Comprimit sense pèrdua
Mapa	TIFF preferentment JPEG2000 com a 2a opció	500ppp	Original b/n: 8 bits escala de grisos Original color: 24 bits color	Sense compressió Comprimit sense pèrdua

Fotografia	TIFF preferentment JPEG2000 com a 2a opció	500ppp	Original b/n: 8 bits escala de grisos Original color: 24 bits color	Sense compressió Comprimat sense pèrdua
Diapositiva	TIFF preferentment JPEG2000 com a 2a opció	500ppp	Original b/n: 8 bits escala de grisos Original color: 24 bits color	Sense compressió Comprimat sense pèrdua
Enregistrament sonor	TIFF preferentment JPEG2000 com a 2a opció	96 Khz	24 bits	Sense compressió
Audiovisual	H264 Vídeo MPEG1 Àudio preferentment	1904x796	8435 kb/s 25 fps	Sense compressió

Formats d'accés a la informació, com es visualitzaran el documents en el repositori donant un accés ràpid i de bona qualitat:

TIPUS DE DOCUMENT	FORMAT	RESOLUCIÓ	DEFINICIÓ	COMPRESSIÓ
Text imprès, (llibres, publicacions periòdiques, partitures...)	JPEG, GIF, PDF	JPEG: 150ppp GIF: 72ppp	JPEG: 24 bits GIF: 8 bits	JPEG: baixa GIF: estàndard
Manuscrit	JPEG, GIF, PDF	JPEG: 150ppp GIF: 72ppp	JPEG: 24 bits GIF: 8 bits	JPEG: baixa GIF: estàndard
Mapa	JPEG, PDF	300ppp	24 bits o 8 bits	Compressió baixa
Fotografia	JPEG, PDF	300ppp	24 bits o 8 bits	Compressió baixa
Diapositiva	JPEG, PDF	300ppp	24 bits o 8 bits	Compressió baixa
Enregistrament sonor	MP3	44.1 Khz	128kbps:192 kbps (vel. transferència)	Compressió baixa (qualitat CD)
Audiovisual	H264 Vídeo MPEG1 Àudio	1904x796	8435 kb/s 25 fps	

### c) Com escanejar una revista o un llibre:

- Primer extreure els elements aliens al suport original: folre, cinta adhesiva, grapes i el codi de barres d'identificació al catàleg, sempre que sigui possible.
- Comprovar si el llibre es pot escanejar a un pàgina doble sense que sobresurtin els marges, sinó és farà plana a plana.
- El llibre ha de enquadrar-se perfectament dins del camp de l'escàner (marge esquerra i marge dret).
- Incorrecte si es veuen franges negres o és superposa el text.
- El llom del llibre ha d'estar sempre el més baix possible, per evitar franges negres que es produeix per deixar espais.
- Un cop escanejat, emmagatzemar-lo amb els formats anteriors. El nom del fitxer ha de seguir les recomanacions del punt 4 d'aquest document.
- En el cas de les revistes, cada número ha d'anar en un fitxer diferent.

## 2. A tenir en compte per l'empresa digitalitzadora o per la biblioteca

Recomanem que proporcioni 4 còpies de cada fitxer a), b), c) d) e), f) en tot cas el nombre mínim serà de 2 còpies:

a) El document sencer en format PDF (molt important que no estigui protegit i que porti OCR incorporat). Si la mida del fitxer supera els 150 Mb, s'haurà de dividir el fitxer en 2 parts o reduir la resolució del PDF sempre que això no comporti una pèrdua significativa de qualitat i l'OCR funcioni correctament.

b) El document en format TIFF en la màxima qualitat possible (600 ppp per exemple). Serà la còpia màster de preservació.

c) El document en format JPEG2000 a menys resolució (300 ppp per exemple). Serà la còpia de visualització.

d) El text extret amb OCR en format text pla (en un fitxer .txt o .pdf per exemple). PDF és un format de visualització que adopta les característiques (resolució, definició i compressió) del format original (JPEG, GIF, etc.).

e) Recomanem que l'empresa proporcioni un fitxer en format METS XML per cada document. METS és un format contenidor que permet descriure objectes digitals. El METS ha de contenir:

- Descripció bibliogràfica del document en format MARC XML, DUBLIN CORE XML.
- Metadades tècniques referents al format de digitalització. Màquina que s'ha fet servir per digitalitzar, versió que s'utilitza d'ADOBE PDF, versió de JPG, versió de TIFF, resolució de les imatges, etc.
- Llista de fitxers relacionats amb el document. Per exemple, si el document és un llibre i l'empresa proporciona el PDF i els TIFF, al METS ha d'aparèixer el nom dels fitxers en la secció corresponent.
- Si el document conté diverses imatges o fitxers organitzats de forma jeràrquica, el METS ha de descriure la jerarquia dels fitxers.

Tots els documents han de tenir un ID unívoc que ha d'aparèixer al METS (ex. Nom del fitxer). El nom dels fitxers ha de tenir sempre la mateixa estructura i ha de contenir aquest ID per a facilitar la seva identificació.

Exemple :

```
<directori - nom de la col.lecció>  
<Subdirectori - nom de subcol.lecció>  
<id - nom del fitxer>
```

- f) Es recomana, per garantir la preservació de les dades, generar metadades PREMIS, que acompanyaran cada objecte digitalitzat.
- g) Si el fitxer

El diccionari de dades PREMIS defineix les metadades de preservació com la informació que un repositori utilitza per portar a terme el procés de preservació digital, metadades destinades al manteniment de la viabilitat, la disponibilitat, la claredat, l'autenticitat i la identitat en el context de la preservació.