

ARCA. REQUERIMENTS TÈCNICS

(Actualització Març 2019)

Aquest document estableix els nous requeriments tècnics obligatoris i aquells que són molt recomanables per a participar en el projecte ARCA.

1. REQUERIMENTS TÈCNICS OBLIGATORIS

OBTENCIÓ D'IMATGES

- Cada número d'una publicació periòdica es considerarà una unitat documental. El conjunt de les unitats documentals conformen la capçalera d'una revista.
- Es digitalitzaran totes les pàgines incloent-hi les portades i contraportades.
- Digitalització preferentment a pàgina simple.
- El procés de digitalització s'ajustarà a la mida de l'original al 100%.
- La digitalització dels documents es realitzarà en color, qualitat RGB a 24 bits per píxel, o escala de grisos, 8 bits per píxel, en funció de les característiques de cada títol.
- La resolució recomanable serà de 400 ppp i la mínima de 300 ppp.
- El maquinari utilitzat en la captació d'imatges consistirà preferentment en escàners verticals amb suport basculant que proporcionin la resolució de preservació exigida.
- El format de sortida de la còpia màster serà TIFF sense compressió o amb compressió ZIP.
- A partir del màster es crearan dues còpies en format JPG:
 - Còpia de difusió a una resolució mínima de 150 ppp i compressió baixa
 - Còpia d'accés en miniatura de 480 píxels d'alçada o 480 píxels d'amplada en imatges apaïssades

RECONeixEMENT ÒPTIC DE CARÀCTERS (OCR)

- Es realitzarà l'OCR de cada imatge TIFF dels documents. A partir d'aquest procés es crearan els següents fitxers:
 - Fitxer codificat en format ALTO ([Analyzed Layout and Text Object](#)), versió 3.1 o superior, per a cada imatge
 - Fitxer PDF/A amb text ocult per a cada unitat documental definida (número, volum, ...)

INDEXACIÓ I ESTRUCTURA DE LA INFORMACIÓ DIRECTORIS

- Cada document serà identificat pel nom i identificatiu de fitxer únic.
- Els fitxers s'organitzaran en els següents nivells:

- **Primer Nivell: Directori principal per capçalera**

Estarà format pel títol de la publicació o bé les paraules clau d'aquest. Opcionalment pot incloure un número identificatiu (topogràfic, número de registre, etc.) si la institució el considera necessari.

- **Segon Nivell: Directoris per formats**

- Directori `\jpg`
- Directori `\miniatures`
- Directori `\ALTO`
- Directori `\PDF`
- Directori `\mets`

Sota els directoris `\PDF` i `\METS` es trobaran els arxius que faran referència a les unitats documentals. El nom s'ha de correspondre amb el nom del directori de la unitat documental.

- **Tercer nivell: Directoris per unitats documentals**

Es crearà un subdirector per a cada un dels números de la publicació. Aquests directoris es poden indexar segons la data de publicació de cada número en la forma AAAA_MM_DD.

- Capçalera\jpg `\directori unitat documental`
- Capçalera \miniatures `\directori unitat documental`
- Capçalera \ALTO `\directori unitat documental`

- **Quart nivell: Directoris per fitxers**

Els fitxers `JPG` de la còpia de difusió, els fitxers `JPG` de les miniatures i els fitxers `ALTO` tindran el mateix nom. El nom ha de ser únic dins una mateixa publicació.

- Capçalera \jpg\directori unitat documental `\imatges.jpg`
- Capçalera \miniatures\ directori unitat documental `\imatges miniatures.jpg`
- Capçalera \ALTO\ directori unitat documental `\fitxers.xml`

SUPORT FÍSIC DE LLIURAMENT

- El suport físic de lliurament final s'acordarà entre la institució i la Biblioteca de Catalunya. Seran preferentment discos durs o DVDs.
- Cada suport tindrà una etiqueta externa i una interna que identificaran el seu contingut.
- L'estructuració de l'espai intern de cada suport serà d'acord amb el descrit a l'apartat de *Indexació i estructura de directoris*.

2. REQUERIMENTS TÈCNICS MOLT RECOMANABLES

FITXERS METS

Per facilitar la publicació a ARCA és molt recomanable lliurar un fitxer METS per cada unitat documental. Aquest fitxer s'ajustarà al perfil [METS Profile for Historical Newspapers \[Draft\]](#) i incorporarà:

- secció capçalera <metsHdr>
- secció metadades descriptives <dmdSec></dmdSec>: metadades descriptives en format MODS que ha d'incloure l'element <mods:enumerationAndChronology> i <mods:number>
- secció arxiu <fileSec></fileSec>:
- secció mapa estructural <structMap></structMap>