

Formats per a la digitalització de documents

Miquel Térmens Graells

RESUM: Es descriuen els principals conceptes que s'han de tenir en compte a l'hora d'utilitzar els fitxers gràfics en activitats de digitalització retrospectiva de documents impresos. A partir de la bibliografia internacional s'exposa quins són els principals problemes que el procés de captació de les imatges pot generar, com utilitzar els formats de gràfics digitals i com mantenir un control de qualitat del procés. Com a apèndix s'adjunten els estàndards que diverses institucions reconegudes estan aplicant en les seves experiències de digitalització.

Arreu del món diverses institucions estan portant a terme experiències per a posar en marxa la biblioteca digital. Aquest concepte —la biblioteca digital— té com a element clau l'ofertament als usuaris de documents digitals de text complet. Partint d'aquesta definició, queda clar que la biblioteca digital es caracteritza bàsicament per l'accés i no tant per la preservació, característica que l'allunya més dels arxius que no pas de les biblioteques tradicionals. Més endavant veurem com aquest decantament per les necessitats de l'accés afecta i arriba a condicionar les solucions tècniques que s'aplicaran a la digitalització dels documents.

Abans d'entrar en matèria cal fer un important aclariment: la diferència entre documents digitals originals i documents digitals resultat de reconversió digital o *scanning*.

Un document digital és un fitxer informàtic basat en algun tipus de codificació que li permet representar informació. Ara bé, els serveis que podrà donar aquest document estan en bona part predeterminats des de l'origen, des del moment de la seva creació. Així, els documents creats de manera original ja en format digital han estat confeccionats amb la previsió que seran visualitzats

per mitjà de pantalles d'ordinador o d'impressores, amb les limitacions tècniques que tant les unes com les altres imposen: dimensions, colors, resolució, contrast, etc. Aquesta previsió farà que les diferències entre allò que l'autor va crear i el que l'usuari percep siguin mínimes o, en tot cas, atribuïbles a les diferències de programari i maquinari utilitzats per cada un. En aquest context no s'haurien de donar situacions de textos no llegibles, colors distorsionats o desplaçaments de pantalla (*scrolls*) innecessaris.

D'altra banda, tenim els documents digitals que són el resultat de la digitalització retrospectiva (*scanning*) d'originals en un altre suport, normalment el paper. En aquests hi ha hagut, per tant, un canvi de format, una conversió que, com totes les conversions, significa un canvi en la natura de les coses i, per tant, la creació d'una de nova. La nova versió digital és diferent quant al suport i a les possibilitats de maneig que aquest ofereix, però també ho pot ser —ho és gairebé sempre— pel que fa a les propietats del contingut informatiu. Si en el cas dels documents gràfics continent i contingut van estretament lligats —pensem en la situació extrema dels documents pictòrics, en els quals les dimensions i la qualitat de la superfície del suport condicionen la percepció dels pigments— en els textuals la qualitat de la impressió o el traç de la cal·ligrafia, per exemple, poden condicionar la llegibilitat o la interpretació del contingut intel·lectual.

Clarificats aquests dos àmbits, cal dir que els estàndards tècnics a aplicar-hi han de ser necessàriament diferents: en els cas dels documents digitals originals, aquests vindran marcats sobretot per les possibilitats que vagi oferint la indústria informàtica, però en el cas dels documents digitals procedents de reconversió, els estàndards vindran condicionats per l'intent d'acostar-se a l'original no digital i de conservar tan bé com es pugui totes les característiques del seu contingut. En aquest darrer àmbit, la casuística és àmplia, tant com ho són les diferències entre els diferents processos de creació intel·lectual i material dels documents tradicionals; d'aquí que no hi hagi solucions universals i que davant d'un mateix cas hom estigui proposant estàndards diferents.

La utilització de diferents formats digitals segons els diferents usos tendeix a desconcertar els professionals de la informació i, en general, tothom que no estigui avesat a la informàtica. Hom tendeix a confondre digitalització de les dades, estandardització i uniformització, quan, en realitat, es tracta de conceptes no unívocs. En el món de la informàtica els estàndards —siguin de *hm* (sorgits d'acords) o *defacto* (imposats per la indústria i l'ús)— són legiò i no paren de fer-se i desfer-se segons l'evolució d'aquesta indústria. Però també són nombrosos els estàndards —en aquest cas més de facto que de *iure*— que tant abans com ara s'utilitzen en l'edició dels documents tradicionals: quants tipus de paper s'utilitzen (gramatge, color, mides, acabat...)?, quantes tipografies?,

quants mètodes d'impressió?, quants formats de compaginació?, quants formats documentals (llibre, revista, carta, gràfic...)?, quants suports?, etc. Ara bé, la posada en marxa de projectes de digitalització retrospectiva no es pot fer sense un important grau d'agrupament tipològic que faciliti, primer, un elevat nivell de mecanització del procés d'escaneig i, després, una gestió assumible dels documents digitals resultants; també s'ha d'admetre que aquest agrupament podrà comportar pèrdues en la fidelitat dels documents reconver-tits.

Aquest article vol ajudar a clarificar els processos tècnics de digitalització retrospectiva de documents textuais i gràfics a partir d'originals en un altre suport —bàsicament el paper—, deixant de banda els documents sonors i d'imatge en moviment i també sense entrar a discutir l'àmbit dels nous continguts creats directament en suport digital.¹ La primera exclusió ve justificada per la, malauradament, baixa presència d'aquest tipus de materials en els nostres centres i es veu reforçada per un panorama tècnic encara molt confús; la segona exclusió es produeix arran de la constatació que, ara com ara, l'elecció dels formats dels documents digitals nous està bàsicament en mans dels seus productors, ja siguin els editors o els distribuïdors.

CONCEPTES GRÀFICS

Abans d'entrar en qualsevol procés de digitalització cal conèixer —i fins i tot dominar— els principals conceptes que es relacionen amb els fitxers gràfics. Només a partir d'aquest coneixement podrem interpretar correctament les diferents estratègies de digitalització.²

Primerament cal saber que en aquests processos només es generen **gràfics de mapa de bits**, (*bitmap*), mentre que els gràfics vectorials poden tenir algun paper, segurament creixent, en els documents digitals de nova creació, com poden ser les revistes electròniques. La principal característica dels gràfics de mapa de bits és la d'estar formats per un conjunt de punts o **pixels** amb

1. Una guia introductòria és la de: Howard Besser; Jennifer Trant, *Introduction to imaging: issues in constructing an image database* (Santa Monica, CA, EUA: The Getty Art History Information Program, 1995), 48 p. És més especialitzada l'obra de: Peter Robinson, *The digitization of primary textual sources* (Oxford: Oxford University Computing Services, 1993), VIII, 104 p. Més actual: Anne R. Kenney; Stephen Chapman, *Digital imaging for libraries and archives* (Ithaca NY: Cornell University Library, 1996), 200 p.

2. Un diccionari bàsic és el de Joe Farrace, *The digital imaging dictionary* (New York: Allworth, 1996), 223 p.

diferents característiques del color, cadascun dels quals es correspondrà després amb un punt de la impressora o de la pantalla de l'ordinador; aquest gran trencaclosques, a diferència dels gràfics vectorials, s'adapta malament als canvis de mida que se li vulguin aplicar, per la qual cosa és gairebé obligatori generar-los en la mida final que es desitgi.

Ara bé, dins d'aquesta gran família hi ha nombrosos **formats**; aquests s'han d'entendre com una norma que descriu com estructurar les dades per a formar un fitxer gràfic.³ Aquestes regles internes orienten cada format en el sentit de ser més eficients en alguna categoria concreta de gràfics: de pocs colors, de molts colors, generats per programes de dibuix, com a resultat de fotografies, amb poca complexitat, etc. Dins d'aquest àmbit, l'eficiència es mesura bàsicament per la capacitat d'oferir més qualitat en el gràfic que es genera emprant el menor nombre possible de *bytes*.

Una d'aquestes regles internes dels formats és **l'algoritme de compressió**. Com tothom ja sap, la quantitat d'informació que es requereix per a definir un gràfic és d'una magnitud tal, que si no és utilitzant tècniques matemàtiques de compressió, consistents en l'eliminació d'informació redundant, el fitxer que es generaria seria massa gran com per oferir un maneig pràctic. Ara bé, alguns d'aquests algorismes aconsegueixen altes cotes de reducció a costa de perdre informació, amb la qual cosa creen gràfics digitals que són de menor qualitat que els seus originals. Un exemple d'algoritme amb pèrdua (*lossy*) és el JPEG, utilitzat dins del format JFIF, però que permet escollir fins a nivells molt elevats el grau de compressió que hom desitja. En canvi, són algorismes sense pèrdua d'informació (*lossless*) el LZW dels gràfics GIF i l'ITU (abans conegut amb les sigles CCITT) dels gràfics TIFF. Les escales de reducció són fixes en alguns casos —aproximadament 2:1 en GIF— o variables en d'altres —fins a 99:1 del JPEG (normalment entre 10:1 i 20:1) o fins a 150:1 del MrSID (normalment entre 15:1 i 100:1).

Una altra característica important és el **nombre de colors** que es poden generar, factor que ve determinat pel **nombre de bits** (no confongueu amb bytes: 8 bits=1 *byte*) que s'utilitzen per a definir el color de cada pixel; per aquest motiu, nombre de colors, nombre de bits i **definició de bits** (*bit depth*) sovint s'utilitzen de manera indistinta. El següent quadre ens mostra les definicions actualment disponibles:

3. A qui ho vulgui saber tot sobre aquest tema li recomanem la monumental obra de James D. Murray; William vanRyper, *Encyclopedia of graphics/deformats*, 2nd ed. (Sebastopol, CA: O'Reilly & Associates, 1996), XXXVI, 1116 p. Més accessible és l'article de José A. Senso, "Gestión de información gráfica: los formatos", *Information World en Español*, vol. 6, no. 9 (septiembre 1997), p. 9-14.

1 bit	2 colors (blanc i negre)
4 bits	16 colors o gammes de gris
8 bits	256 colors o gammes de gris
16 bits	32.768 o 65.536 colors
24 bits	16.777.216 colors
32 bits	16.777.216 colors

Aquests colors es generen a partir d'una **paleta de colors estàndard**, que inclou tota la gamma cromàtica, o d'una *d'adaptada*, amb només els colors que s'estan utilitzant, cosa que pot ajudar a reduir la mida final del fitxer.

La **resolució gràfica** és un altre dels conceptes clau perquè ens indica el nombre de píxels que formen una imatge; per exemple 300 (en sentit horitzontal) x 200 (en sentit vertical). Cal no confondre la resolució gràfica utilitzada per a mesurar l'amplada i la llargada d'un gràfic o d'una pantalla amb els **DPI** o **resolució lineal** d'una impressora o escàner. Els DPI (*dots per inch*) o PPP (punts per polzada) mesuren, com el seu nom indica, el nombre de punts o píxels que en una polzada una impressora és capaç d'imprimir o un escàner de reconèixer i generar; es tracta, per tant, d'una mesura lineal i no de superfície com l'altra.⁴ Malgrat aquestes precisions, cal advertir que constantment hom troba barrejats tots dos conceptes, fins i tot en manuals tècnics de marques de prestigi. Un nombre alt de DPI, per exemple 400 DPI, identifica un gràfic de qualitat, però una resolució gràfica alta, per exemple 1.200 x 800 píxels només indica que es visualitza molt gran.⁵

Podríem seguir revisant altres conceptes propis dels fitxers gràfics, però els que hem vist aquí són suficients per a comprendre'n el funcionament bàsic i sobretot l'aplicació. Així, podem veure que la mida dels fitxers que es generin —tema clau per a la seva gestió— dependrà bàsicament de la resolució (si volem tenir el gràfic més gran o més petit) i del nombre de bits/colors utilitzats (com

4. Cal anar en compte a l'hora d'interpretar quina és la resolució màxima que permet un escàner. Molts escàners poden augmentar la *resolució òptica* de l'equip mitjançant programes que interpolen o generen més punts a partir de dos de preexistents; aquesta major *resolució interpolada*, per tant, no és real i no és desitjable per a aconseguir productes d'alta qualitat.
5. Coneixent la resolució (gràfica o lineal) i la definició de color d'una imatge, podem calcular l'espai que ocupa de diferents maneres: si sabem la resolució gràfica, com 300x200, i la definició del color, com 8 bits (256 colors), el càlcul serà $300 \times 200 \times 8 = 480.000 \text{ bits} / 8 = 60.000 \text{ bytes} = 60 \text{ Kb}$; si tenim les mides de l'original, com 15x10 cm (6x4"), la resolució lineal que es vol obtenir, com 300 DPI, i la definició de color, com 16 bits (32K colors), ara el resultat serà $(6 \times 300) \times (4 \times 300) \times 16 = 1.800 \times 1.200$ (que serà la resolució gràfica) $\times 16 = 34.560.000 \text{ bits} / 8 = 4.320.000 \text{ bytes} = 4,32 \text{ Mb}$. Aquestes seran les mides del fitxer sense comprimir, perquè quan els gravem amb qualsevol dels formats gràfics es reduiran gràcies a les tècniques de compressió.

més colors, més qualitat, però també calen més bits). En canvi, la qualitat resultant —important per a l'usuari— anirà lligada als colors i a la resolució, però t'untbé a la pèrdua o no d'informació que provoqui l'algoritme de compressió. Aquí ja podem entreveure un dels principals reptes de la digitalització: l'equilibri entre les necessitats de qualitat, amb la demanda de més colors, més resolució i sense pèrdua (per tant, amb menys compressió); i la necessitat de reduir la mida dels fitxers per a agilitar la transmissió, amb menys colors, menys resolució i més compressió de dades (per tant, amb pèrdua d'informació).

ELS FORMATS GRÀFICS

La quantitat de formats gràfics existents actualment és enorme,⁶ però només una minoria és àmpliament utilitzada en processos de digitalització; vegem-los a continuació:

GIF

El *Graphics Interchange Format* va ser desenvolupat per CompuServe amb la intenció que fos utilitzat dins la seva xarxa telemàtica. Només admet de 2, 4 o 8 bits per píxel i, per tant, només pot mostrar 2, 16 o 256 colors, que fins fa poc era la resolució estàndard dels monitors dels ordinadors personals. A nivell de compressió utilitza l'algoritme LZW, que no té pèrdua i actua de manera eficient davant de colors plans, franges i, en general, dibuixos generats per ordinador, però que comprimeix menys davant de fotografies o imatges complexes. Cal assenyalar que n'existeixen dues versions, 87a i 89a, la segona de les quals permet generar gràfics transparents o en moviment. Per les seves possibilitats, quan va aparèixer la World Wide Web, el GIF es va convertir en el seu format per excel·lència i ara és reconegut per qualsevol navegador.

JFIF/JPEG⁷

L'acrònim JPEG (*Joint Photographic Experts Group*) fa referència a l'organització que va crear aquest format i també a l'algoritme de compressió que s'hi utilitza, però si hom vol ser purista ha de constatar que JPEG només és el nom del format de compressió, mentre que el format gràfic que el conté és JFIF (JPEG

6. Murray i vanRyper en mostren un centenar, molts d'ells amb variants i diferents versions.

7. *JPEG-JBIG Homepage*, <<http://vvrww.jpeg.org>> [Consulta: 11 gener 1999].

File Interchange Format). Aquesta mateixa confusió ja ens indica fins a quin punt la compressió n'és un element determinant: és una compressió amb pèrdua, però que l'usuari pot graduar fins a un nivell de 100:1 (és a dir, que el fitxer resultant serà cent vegades menor que l'original), mentre que el nombre de bits ha de ser 8 o 24 (256 o 16 milions de colors). El JPEG és un dels pocs algoritmes que actuen bé amb grans imatges de 24 bits i, per tant, està molt indicat per a casos complexos com ara les fotografies i s'ha convertit en un complement del GIF pel fet de ser també força conegut pels navegadors de la WWW.

TIFF

El *Tagged Image File Format* és considerat el format més versàtil per a guardar imatges. Aquesta qualitat permet que s'hi utilitzin algoritmes de compressió diferents (sense compressió, RLE, LZW, ITU Grup III, ITU Grup IV i JPEG), i el TIFF queda com un format d'encapsulament i d'unificació externes. La versatilitat també es demostra pel fet que permet generar gràfics amb una qualitat de 1 a 24 bits (de blanc i negre a 16 milions de colors). Actualment TIFF es presenta en tres versions diferents (4.0, 5.0 i 6.0), cadascuna de les quals ha anat afegint noves capacitats a l'anterior. També cal destacar que la popularitat del TIFF és deguda al fet que incorpora els algoritmes ITU Grups III i IV, creats específicament per a emmagatzemar amb alta compressió i sense pèrdua imatges en blanc i negre; actualment ITU Grup III és l'algoritme propi de les transmissions per fax, mentre que ITU Grup IV n'és una millora implementada en altres contextos. Tot plegat ha convertit el TIFF en el format base del programari de la majoria d'escàners.

Photo-CD°

h'Image Pac, conegut comercialment com a Kodak Photo-CD, és un format propietari creat per Eastman Kodak amb l'objectiu inicial que servís com a suport en cd-rom dels negatius fotogràfics, sense perdre qualitat. El seu algoritme de compressió està, per tant, plenament orientat a les imatges fotogràfiques per a les quals obté molt bons resultats, amb menys distorsions que el JPEG i poca pèrdua d'informació, però aquests decreixen davant d'altres

8. [Kodak]: Frequently Asked Questions: Kodak Photo CD, [Rochester, NY]: Kodak, [1998], <<http://www.kodak.com/global/en/service/faqs/faq1001a.shtml>>, [Consulta: 11 gener 1999].

tipologies gràfiques.⁹ En tot cas, té molta acceptació en arxius gràfics, sobretot de biblioteques d'art i de centres de documentació lligats a museus d'art. Els fitxers Photo-CD són multiimatge; això vol dir que cada un d'aquests incorpora la imatge en 5 o 6 resolucions diferents predeterminades:

	<i>Photo CD Master Disc</i>	<i>Photo CD Pro</i>
Base/16 (128 x 192 pixels)	X	X
Base/4 (256 x 348 pixels)	X	X
Base (512 x 768 pixels)	X	X
Base*4 (1.024 x 1.536 pixels)	X	X
Base*16 (2.048 x 3.072 pixels)	X	X
Base*64 (4.096 x 6.14 pixels)		X

El més sorprenent i atractiu des del punt de vista de la biblioteca és que aquests fitxers multiimatge es generen automàticament amb el programari pertinent (aquí i en el maquinari es basa el negoci de Kodak), i que el programari per a visualitzar-los és gratuït. Es tracta, per tant, de gairebé una solució "mans en clau", ben provada. Per contra, Photo-CD, alhora que ens lliga a una marca comercial, també presenta alguns problemes tècnics: té uns alts requeriments de procés de màquina que el fan actuar lentament en ordinadors personals corrents i hi ha la impossibilitat d'accedir des de l'entorn web a una resolució determinada d'una imatge, tot i que sembla que darrerament s'ha intentat trobar-hi algunes solucions.

Adobe Acrobat

El *Portable Document Format (PDF)*,¹⁰ conegut a nivell comercial amb el nom *d'Acrobat*, és un format creat per Adobe i propietat d'aquesta empresa líder en programaris de tractament digital professional. En general no és considerat com un format pur d'emmagatzemament, sinó només com un format de visualització, perquè en realitat parteix de documents digitals previs —un o més— i no d'originals impresos, que després empaqueta en un sol fitxer. La

9. Una excel·lent avaluació d'aquest format a: Anne R. Kènney; Oya Y. Rieger, *Using Kodak Photo CD technology for preservation anil access: a guide for librarians, archivists and curators* ([Ithaca, NY]: Cornell University, [May 1998]), 21 p. Versió Acrobat a: <<http://www.library.cornell.edu/preservation/kodak/kodak-pdf.pdf>>, [Consulta: 11 gener 1999]. Una comparació entre Photo-CD i JPEG a: Jef Bone, *PhotoCD vs. JPEG*, 16 juliol 1997, <<http://www.infomedia.net/scan/TSF-PhotoCD.html>>, [Consulta: 11 gener 1999].

10. <<http://www.adobe.com/acrobat>>, [Consulta: 11 gener 1999].

compressió, parcialment manipulable, es fa a partir dels algorismes JPEG, LZW i ITU Grups III i IV, segons si es tracta de text o de gràfics. Els fitxers Acrobat presenten l'avantatge d'agrupar els documents multipàgina en un únic fitxer, facilitant-ne la seva gestió tant als usuaris com als administradors del sistema. Gràcies a la capacitat d'incorporar diferents claus i nivells d'accés, també ofereixen més protecció contra el seu ús indegut, però a costa de reduir el nivell de qualitat i de lligar-se a un format propietari.

ÚS DEL FUTUR DOCUMENT DIGITAL

Cal tenir clar amb quin objectiu es posarà en marxa la digitalització d'impresos o de documents històrics. Del resultat d'aquesta pregunta dependran en bona part la resta de solucions adoptades. Alguns usos, a vegades combinables entre si, poden ser:

- la consulta via web;
- la consulta especialitzada per mitjà d'ordinadors específics;
- la projecció en pantalla gegant per a usos docents;
- la reproducció d'alta qualitat amb finalitats de recerca o difusió;
- servir de banc d'imatges per a la seva reproducció dins de publicacions impreses;
- arxiu amb finalitats de preservació.

Es pot plantejar oferir l'accés a aquestes opcions des de tres localitzacions diferents, amb les implicacions d'amplada de banda de xarxa i d'equipament necessari que això implica:

- des de punts específics de les biblioteques;
- des de qualsevol punt de la xarxa local de la institució;
- des de qualsevol punt d'Internet, incloent-hi els domicilis particulars.

Finalment, també cal considerar que s'hi poden donar diferents nivells d'accés:

- només als usuaris autoritzats;
- per a qualsevol usuari de la biblioteca o del centre;
- de lliure accés.

L'accés també comporta la disponibilitat de diferents drets sobre els documents digitals, com poden ser la recuperació i visualització dels documents

.iiiiib unes resolucions gràfiques o altres (diferents centres es plantegen permetre als investigadors l'accés als fitxers de més qualitat, però no als usuaris i'omuns), la utilització per a usos propis o la impressió amb diferents qualitats. Tots aquests paràmetres d'utilització s'haurien almenys de plantejar abans d'emprendre accions de digitalització retrospectiva, perquè de la contestació d'alguna d'aquestes preguntes pot dependre la tria d'uns formats i d'unes qualitats o altres.

L'ENFOCAMENT ACTUAL

Hi ha uns elements mínims de consens dins la comunitat especialitzada en l'enfocament tècnic que cal aplicar a la digitalització retrospectiva.

Primerament, tothom està d'acord que cal fer una distinció clara entre fitxers d'arxiu i fitxers de visualització. Actualment hi ha limitacions tècniques per a la transmissió i visualització de fitxers gràfics que aconsellen reduir al màxim la seva mida. Ara bé, la reducció del nombre de bytes dels fitxers només es pot aconseguir, com ja em dit abans, amb l'aplicació de tècniques de compressió, processos que sovint també comporten una reducció en la qualitat de les imatges, que potser no és apreciable en usos normals, però que pot ser crítica en usos més professionals. Per aquesta raó, cal establir estàndards de fitxers d'arxiu, que preservin la màxima qualitat de les imatges, i estàndards de fitxers de visualització que en facilitin l'ús per part dels usuaris comuns.

Per cada imatge original normalment es genera més d'un fitxer digital. Això es fa per a adequar la disponibilitat de fitxers a les diferents necessitats d'ús; aquestes són bàsicament tres:

- Tenir un fitxer *mestre o d'arxiu* (el *master*) amb la màxima qualitat possible, que actua de còpia de seguretat i, alhora, serveix per a generar les altres còpies o per a futurs reprocessaments si milloren els estàndards; és, per tant, una imatge no comprimida o amb baixa compressió sense pèrdua, que a vegades també pot ser utilitzada pels usuaris per a reproduccions de qualitat.
- Oferir un fitxer *intermig* o de referència, per a ser utilitzat normalment pels usuaris un cop l'han recuperat en una cerca; l'objectiu és tenir una imatge d'alta qualitat i de mida intermitja que pugui ser visualitzada per les pantalles de la majoria d'ordinadors i que, si cal, se li pugui aplicar un *zoom* per a estudiar-la amb més detall.
- Incloure un fitxer *miniatura* (*thumbnail*) per a presentar juntament amb el registre bibliogràfic en catàlegs de darrera generació i ajudar els usuaris a

decidir si volen recuperar la imatge en més alta qualitat; el fitxer ha de ser molt petit i adaptable a qualsevol entorn informàtic (el web és imprescindible), encara que sigui oferint una baixa qualitat.

En el procés d'escaneig només s'han de crear els fitxers d'arxiu, i a partir d'aquests, seguint processos més o menys automatitzats es generen totes les altres còpies destinades a la visualització. Els estàndards que s'apliquin a aquestes podran variar segons avanci la tecnologia i, per tant, s'ha d'estar preparat per a aplicar processos de renovació i substitució massiva de fitxers. Ara bé, tot això no serà possible en el futur si la digitalització inicial no s'ha fet amb el màxim de rigor possible i la informació ha estat recollida en els fitxers d'arxiu.

Malauradament també s'està d'acord en el fet que en aquest moment resulta impossible assegurar a nivell professional la consistència del color a través de les diferents plataformes d'escaneig, visualització i impressió, i de totes elles amb l'original. Mentre no es despleguin nous estàndards com el de *Color Management Systems* (CMS), la millor recomanació és la de digitalitzar amb la màxima resolució possible les fotografies i il·lustracions en què calgui millor qualitat." També és imprescindible tenir ben calibrats els escàners i les pantalles, tema que destaca tothom que explica la seva experiència en projectes operatius.

En aquells casos en què no hi hagi presència de color, la decisió de gravar els fitxers en blanc i negre purs (1 bit per pixel) o bé en escala de grisos (8 bits per pixel) és una de les més difícils de prendre, com ho demostra el fet que les institucions nord-americanes amb més experiència a vegades hagin canviat els seus enfocaments.¹² Tothom que tingui una mínima experiència en el maneig de gràfics sap que les imatges bitonals representen molt malament les línies corbes i les diagonals, oferint uns resultats molt llejós gairebé visibles a simple vista, llevat que es treballi en un nombre alt de DPI; les impressores matricials són un bon exemple d'aquests resultats. Per contra, les escales de grisos milloren la visualització gràcies als programaris que utilitzen tècniques com l'"antiàlies" (*antialiasing*), però a costa d'augmentar enormement la mida dels fitxers resultants. Però el dilema encara té un altre vessant: els gràfics en escala de grisos són millors per a les pantalles d'ordinadors (enganyen millor els nostres ulls), però els bitonals s'imprimeixen amb més facilitat (requereixen menys temps de procés).

11. Franziska Frey, "Digital imaging for photographic collections: foundations for technical standards", *RLG DigiNews*, vol. 1, no. 3 (15 December 1997), <<http://www.rlg.org/preserv/diginews/diginews3.html>>, [Consulta: 11 gener 1999].

12. Compareu les versions d'agost de 1996 i de juliol de 1998 del document Carl Fleishhauer, *Digitalformats for content reproductions*.

La utilització de compressions amb pèrdua provoca, en alguns casos, l'aparició d'"artefactes" (*artifacts*) a la imatge visibles a simple vista. Aquests artefactes són errors de l'algoritme de compressió que generen línies, punts i colors sense sentit i inexistents en l'original. És prioritari evitar-los, la qual cosa s'aconsegueix bàsicament utilitzant en cada format gràfics només en aquelles aplicacions per a les quals va ser creat; així, per exemple, el JPEG provoca "artefactes" si és utilitzat en dibuixos fets amb ordinador que tenen grans blocs de colors plans i línies o canvis sobtats.

Per acabar, també cal destacar que la digitalització no s'acaba amb l'operació d'escaneig, perquè els resultats es poden millorar molt —és obligatori fer-ho— amb el retoc posterior de les imatges amb programes d'edició digital com Photoshop o Corel Draw; amb aquests podrem, per exemple, millorar el contrast, corregir la gamma de colors i optimitzar la paleta. En alguns casos, es poden automatitzar aquestes operacions.

Les característiques i estat de conservació dels suports i, encara més, les tècniques emprades en la reproducció del contingut, aconsellen dividir els documents originals en diferents grups, als quals cal donar els tractaments tècnics més apropiats; vegem-los a continuació.¹³

Materials gràfics

Es tracta de documents gràfics sobre paper, com ara gravats i fotografies. En alguns casos la il·lustració va acompanyada d'algun tipus de text en forma de títol, resum del contingut o dades d'impressió; s'ha de procurar que aquesta informació textual no es perdi i continuï sent llegible. Un altre problema d'aquests materials són les seves dimensions, que en molts casos superen els formats admesos pels escàners corrents i, alhora, dificulten la seva visualització per pantalla, impedit, sovint, la seva exposició sencera al 100% de zoom.

El cas de les fotografies és especialment sensible perquè es tracta d'originals de gran qualitat i que, per tant, requereixen ser digitalitzats amb elevades resolucions. Un segon problema és el de la representació del color, perquè l'entorn digital encara no l'ha normalitzat, almenys amb els nivells de fidelitat que es requereixen en usos professionals.¹⁴

13. *Technical notes on formats for digital reproductions*, [Washington DC: Library of Congress, 1998], <http://memory.loc.gov/ammem/award/html/technical_notes1.html>, [Consulta: 11 gener 1999].

14. Francisca Frey, "Digital imaging for photographic collections: foundations for technical standards", *RLG DigiNews*, vol. 1, no. 3 (15 december 1997), <<http://www.rlg.org/preserv/diginews/diginews3.html>>, [Consulta: 11 gener 1999].

Materials textuais reproduïts com a imatges

Quan els textos són la part més important d'un document, els processos de digitalització també han de canviar per tal de poder permetre els majors nivells possibles d'utilització del seu contingut intel·lectual; això s'aconsegueix amb la seva reproducció com a text i imatges cercables. Aquesta línia de treball és la que ofereix millors rendiments futurs, però també la que presenta més requeriments tècnics, actualment només en mans d'uns pocs centres especialitzats: digitalització amb OCR (*Optical Character Recognition*), creació de documents codificats amb SGML (amb el desenvolupament o l'adaptació d'especificacions DTD -*Document Type Definition*-) i, finalment, la creació de sistemes de recuperació i visualització per part dels usuaris.¹⁵

Com a alternativa més assequible, tècnicament i econòmicament, hi ha una altra línia de treball: es poden reproduir els textos com a imatges i permetre la seva consulta directa com a tals o en associació amb textos cercables. Aquesta és la metodologia que segueixen la majoria dels projectes de biblioteca digital.

Com ja hem dit abans, és possible la digitalització dels documents en forma tonal (color o escala de grisos) o bitonal (blanc i negre). En la majoria de casos, el sistema tonal ofereix millors resultats que el simple blanc i negre. Les imatges bitonals s'han de generar quan estiguem davant de treballs tipogràfics o amb grafisme lineal (per exemple, il·lustracions fetes amb tremp) i normalment faciliten una reproducció més senzilla a causa d'uns menors requeriments d'impressora.¹⁶

No es pot oblidar la possibilitat de crear imatges tonals per a "arxiu" i bitonals per a "referència".

15. Com a exemple del rigor necessari per a digitalitzar i després aplicar l'OCR, pensem que la *Library of Congress* actualment està exigint als seus proveïdors externs una taxa d'encerts del 99,95%; en la mateixa línia, un estudi finançat per la *Deutsche Forschungsgemeinschaft* (Associació Alemanya de Recerca) assenyalava directament que si no s'arriba al 99,95% —o sigui, només 4 o 5 errors per 1.000 caràcters— surt més econòmic entrar a mà el text; vegeu: Hartmut Weber; Marianne Dörr, *Digitisation as a method of preservation?* Amsterdam, July 1997 (edició original en alemany el 1996), <<http://www.knaw.nl/ecpa/publ/weber.html>> [Consulta: 11 gener 1999]. Es pot trobar un estudi pilot sobre els requeriments d'escanejat de l'OCR a l'article: Stephen Chapman; Anne R. Kenney, "Digital conversion of research library materials: a case for full informational capture", *D-Lib Magazine*, October 1996, <<http://www.dlib.org/dlib/october96/Cornell/10chapman.html>> [Consulta: 11 gener 1999].
16. El millor estudi tècnic sobre els problemes que presenten les imatges tonals és el de James M. Reilly; Franziska S. Frey, *Recommendations for the evaluation of digital images produced from photographic, microphotographic, and various paper formats* (Rochester, NY: Rochester Institute of Technology. Image Permanence Institute, May 1996), 40 p. Report to the Library of Congress National Digital Library Project. Reproducció Acrobat disponible a: <<http://lcweb2.loc.gov/-ammem/lcireprt.pdf>> [Consulta: 11 gener 1999].

Mapes

Ils documents cartogràfics són especialment problemàtics a l'hora de digitalitzar a causa de les seves grans dimensions i també per les necessitats de qualitat que tenen els seus usuaris: visualització amb diferents graus d'aproximació, diferenciació entre tons de color, claredat en la reproducció de les parts textuals, etc. Aquests dos requeriments fan que els equips necessaris per al seu cscanneig hagin de tenir característiques especials. També és la raó que explica que poques institucions hagin volgut iniciar processos seriosos de digitalització de mapes, més enllà de les proves o de la creació de miniatures i mapes de situació lligats a la recuperació per catàlegs.

Des del punt de vista dels formats digitals, els mapes són la tipologia documental que més s'allunya de la utilització dels formats genèrics com GIF i JPEG en les funcions de visualització i distribució. La necessitat d'intentar lligar elevada compressió, alta qualitat i possibilitats de *zoom* o de descàrrega parcial dels fitxers està portant a experimentar amb altres solucions, encara que siguin propietàries, ben adaptades al treball en xarxa.¹⁷ Exemples destacats són les compressions wavelet i fractal^{18*} i l'emergent format FlashPix.¹⁹

CONTROL DE QUALITAT

La digitalització de documents requereix un acurat control de la qualitat al llarg de tot el procés ja que, no ho oblidem, aquest té un alt cost i qualsevol digitalització retrospectiva, fins i tot si s'acaba considerant com a defectuosa, és, a la pràctica, única i irrepetible. Els majors perills que cal vigilar són tres:

- inadequació entre la tipologia del document original i el format gràfic elegit per a la seva digitalització;
- no-seguiment durant el procés de digitalització dels estàndards preestablerts;
- errors materials en la digitalització produïts pel treball en lots.

17. David Yehling Allen, "Creating and distributing high resolution cartographic images", *RLG DigiNews*, vol. 2, no. 4 (15 august 1998), <<http://www.rlg.org/preserv/diginews/diginews24.html>> [Consulta: 11 gener 1999].

18. Steven Puglia, "Fractal and wavelet compression", *RLG DigiNews*, vol. 2, no. 3 (15 june 1998), <<http://www.rlg.org/preserv/diginews/diginews23.html>> [Consulta: 11 gener 1999].

19. Flashpix té el suport del consorci industrial Digital Imaging Group <<http://www.digitalimaging.org>> [Consulta: 11 gener 1999].

La prevenció i/o detecció d'aquests errors requereix l'establiment de mecanismes de control en les tres fases del procés —classificació dels documents, encàrrec de la feina i recepció dels documents digitalitzats—, segons veurem a continuació.

Classificació dels documents

Abans d'encarregar qualsevol digitalització, s'han d'haver revisat i classificat els documents que hi participaran, seguint les següents indicacions:

- comprovar que l'interès dels documents en justifiqui la digitalització;
- comprovar que el seu estat de conservació i, si s'escau, el de la seva enquadernació pot suportar el procés de digitalització;
- detectar i, si s'escau separar, l'existència de notes, punts de llibre i altres afegits aliens als documents;
- classificar els documents segons el seu format, tipologia (publicació periòdica, llibre, fulletó...) i contingut tipogràfic (text, dibuixos, fotografies...);
- analitzar per a cada grup anterior quin destí es donarà al document digital i, en conseqüència, quin tipus de format digital s'ha d'utilitzar;
- inventariar per triplicat (centre d'origen, centre coordinador de la digitalització i empresa concessionària) tot el material fent constar el format i característiques de digitalització de cadascun;
- preparar degudament el material per al seu trasllat al local de digitalització.

Encàrrec de la feina

L'encàrrec de la feina a empreses especialitzades externes és una pràctica habitual que s'ha de regir per criteris de preu i de qualitat. Aquests haurien d'estar especificats de manera precisa dins les clàusules tècniques del contracte de digitalització. De manera ideal, caldria incloure-hi:

- garantia del proveïdor de manipular els materials originals segons els bons usos acceptats i de garantir-ne la integritat;
- garantia del proveïdor de no realitzar per a ús propi o per a tercers còpies dels originals o dels fitxers ja digitalitzats, en qualsevol sistema de reproducció;
- especificació detallada dels processos de digitalització a aplicar, en especial:
 - format gràfic i, si cal, subformat;

- estàndard de compressió i ratio a aplicar;
- nombre de bits a utilitzar en cada imatge i, quan calgui, si aquests han de ser de color o d'escala de grisos;
- nombre de DPI utilitzats en l'escaneig;
- taula de colors que cal utilitzar;
- i, si cal, mida en píxels de la imatge a obtenir;
- suport en què s'hauran d'entregar els fitxers digitalitzats;
- acceptació dels criteris aplicats per la institució contractant per a considerar un fitxer com a correcte;
- compromís de repetir els fitxers que siguin detectats com a defectuosos dins del procés de recepció.

Recepció dels documents digitalitzats¹⁰

Quan es reben documents digitalitzats procedents d'un subministrador, ja sigui extern o intern, s'hauria de seguir un procediment de revisió del material entregat i de la seva qualitat; els punts essencials serien els següents:

- comparació de les remeses d'originals enviats a digitalitzar i de les retornades, i comprovació del seu estat de conservació;
- comparació de les llistes de fitxers entregats amb els efectivament rebuts i reclamació de qualsevol absència;
- revisió dels fitxers rebuts per comprovar:
 - que el nom del fitxer segueixi la convenció preestablerta;
 - que la imatge s'atengui als paràmetres de qualitat designats²¹ i que no presenti problemes, com ara la pèrdua de línies fines causada per un escaneig amb poca resolució, la pèrdua dels detalls més petits de les il·lustracions, la imatge torta, el text tallat en algun dels laterals (incloent-hi els números de pàgina) i els originals gràfics escanejats sense el peu de text o sense el títol;
 - que no hi falti cap pàgina per digitalitzar en el cas de documents multipàgina.

20. Basat en: Library of Congress. National Digital Library Program, *Quality review of document images. Internal training guide* (Washington DC: Library of Congress. National Digital Library Program, September 1996), <<http://lcweb2.loc.gov/ammem/award/docs/docimqr.html>> [Consulta: 11 gener 1999]. Vegeu també l'apartat corresponent de les *NARA guidelines...*, p. 16-17.

21. L'objectivació dels estàndards de qualitat és un dels aspectes més difícils que presenta la digitalització. Una aproximació és l'aplicació de tècniques de *benchmarking*, com proposa l'especialista de la Universitat de Cornell, Anne Kenney, *Benchmarking image quality: front conservation to presentation*. <<http://www.uky.edu/~kiernan/DL/kenney.html>> (Consulta: 11 gener 1999).

Tant la Library of Congress com la NARA obliguen el seu personal a revisar de manera aleatòria el contingut del 10 % de les imatges i retornen al proveïdor les incorrectes. A més, la NARA només accepta fins a un 1% d'errors detectats en la inspecció d'un lot de fitxers i si se supera aquesta xifra, tot el lot és retornat a la unitat o al proveïdor que el va generar.²²

DIGITALITZACIÓ I MICROFILMACIÓ

Durant molts anys la microfilmació ha estat una aposta segura per a la preservació de documents, especialment d'arxiu i hemerogràfics. El desenvolupament de les tècniques de digitalització ha obert la possibilitat de que aquestes puguin substituir la microfilmació. Els elements que cal valorar són de gran importància i a tall d'inventari podem esmentar: el cost de creació dels respectius *masters*, el cost de conservació (ben elevat en el cas dels documents digitals), el cost de recòpia del suport (gairebé innecessari en els microfilms), la perdurabilitat tècnica del format (dubtosa en els fitxers digitals) i la facilitat de recuperació (molt més elevada en els documents digitals).

No és aquí el Hoc per a resseguir aquest debat, però sí que ho és per a comentar una tècnica que relaciona els dos procediments: es tracta de la digitalització a partir d'una microfilmació prèvia.²³ La majoria d'empreses que treballen dins d'aquests camps han desenvolupat eficients sistemes automatitzats que permeten la digitalització a baix cost de microfilms preexistents; aquest, per tant, no és el problema. La qüestió és què fer amb els documents que encara no s'havien microfilmats: cal seguir microfilmant-los i després, si cal, digitalitzar-los? o cal destriar categories documentals a les quals d'ara endavant només s'aplicarà la digitalització? Les respostes no són fàcils i, per acabar-ho de complicar, hi ha empreses privades que recomanen microfilmant no ja com a destí final, sinó com a mitjà per a facilitar la digitalització a partir dels rotlles de microfilm; l'aposta és suggerent: "per un preu no gaire superior podeu tenir les dues coses";²⁴ però aquestes empreses no fan sinó repetir arguments que grans especialistes de la matèria ja havien proposat a inicis dels noranta.²⁵

22. *NARA guidelines...*, p. 17.

23. L'escañeig de microfilms de diaris és revisat a partir de tres projectes per Alan Howell, "Film scanning of newspaper collections: international initiatives", *RLG DigiNews*, vol. 1, no. 2 (15 agost 1997), <<http://www.rlg.org/preserv/diginews/diginews2.html>> [Consulta: 11 gener 1999].

24. Una mostra representativa la podeu trobar en el fabricant d'escàners de microfilms i microfites Sunrise: <<http://www.sunriseimg.com>> [Consulta: 11 gener 1999].

25. Don Willis, *A hybrid systems approach to preservation of printed materials* (Washington DC: Commission on Preservation and Access, 1992), 111, 45 p.

El debat continua a nivell internacional, però s'ha clarificat força amb estudis comparatius com els realitzats per la Cornell University²⁶ i l'Associació Alemanya de Recerca.²⁷ Ambdues investigacions volien mesurar els resultats que s'hi podien obtenir en una aproximació "híbrida": digitalitzar a partir d'una microfilmació prèvia, o bé digitalitzar primer i després extreure'n un microfilm com a còpia de preservació. Les principals conclusions van ser, de manera gairebé coincident:

- en microfilmar primer s'obliga a fer dues conversions (paper - microfilm - digital) que afectaran la qualitat final;
- la microfilmació es tendeix a fer amb un alt contrast per tal de fer més lleugibles els textos i reduir les taques del paper; això fa que sigui poc recomanable per a imatges tonals;
- si es volen obtenir fitxers en color, la digitalització directa manté molt millor la qualitat que no pas si es fa a partir d'un microfilm; si el resultat ha de ser en escala de grisos, la diferència es redueix però continua estant a favor de la digitalització directa; la captura bitonal tendeix a ser millor amb els equips de microfilmació;
- els escàners de microfilms és possible que no tinguin prou resolució per a capturar tota la informació;
- els escàners de microfilm presenten problemes tècnics a l'hora d'adaptar la quadrícula a escanjar a les dimensions reals del document microfilm;
- si després es vol aplicar un OCR, s'aconsegueixen millors resultats si es fa una digitalització directa i no a partir d'un microfilm;
- la resolució digital obtinguda pels escàners és menor que la resolució fotogràfica dels aparells de microfilmació;
- actualment la disponibilitat d'empreses amb maquinària professional i experiència és molt superior en el camp de la microfilmació.

Com es pot veure, tots dos procediments tenen avantatges i desavantatges, i en bona part dependrà de les capacitats econòmiques i tècniques de cadascú, i dels objectius que es vulguin aconseguir el fet de triar-ne un o altre.

26. Anne R. Kenney, *Digital to microfilm conversion: a demonstration project 1994-1996: final report to the National endowment for the humanities* (Ithaca, NY: Cornell University Library. Department of Preservation and Conservation), <<http://www.library.cornell.edu/preservation/comfin.html>> [Consulta: 11 gener 1999].

27. Hartmut Weber; Marianne Dörr, *Digitisation...*

ESTÀNDARDS TÈCNICS APLICATS A LA DIGITALITZACIÓ

A continuació es presenten els estàndards que actualment estan aplicant algunes de les principals institucions implicades en processos de digitalització.²⁸

Library of Congress

La *Library of Congress* va endegar el projecte pilot *American Memory*²⁹ com a experiència prèvia al desenvolupament actual de la *National Digital Library*, en la qual s'ha integrat. Sens dubte és la institució que a nivell mundial ha arribat més enllà en els processos de digitalització retrospectiva, tant pel volum dels materials tractats com per l'exhaustiva tasca tècnica que s'hi està portant a terme.

Actualment s'hi generen tres fitxers per cada document: miniatura, referència i arxiu.³⁰

Materials gràfics

Miniatura

<i>Definició:</i>	<i>8 bits per pixel</i>
<i>Format:</i>	<i>GIF</i>
<i>Compressió:</i>	<i>Pròpia del GIF</i>
<i>Resolució:</i>	<i>D'uns 150x100 a uns 200x200 pixels</i>

Referència

<i>Definició:</i>	<i>Escala de grisos: 8 bits per pixel; color: 24 bits per pixel</i>
<i>Format:</i>	<i>JFIF (JPEG File Interchange Format)</i>
<i>Compressió:</i>	<i>JPEG (normalment una compressió 10:1 en escala de grisos i de 20:1 en color)</i>

28. Hom pot trobar les opinions d'altres institucions nord-americanes a The Library of Congress, *The Library of Congress/Amerilcdi National Digital Library Competition: lessons learned*, [Washington DC]: Library of Congress, 31/8/98, <<http://memory.loc.gov/ammem/award/lessons.html>> [Consulta: 11 gener 1999]. Notícies sobre altres experiències a diferents països europeus es poden trobar en el número monogràfic "Digital images in libraries", *VINE*, no. 107 (1997), 83 p.
29. <<http://lcweb2.loc.gov/ammem/>> [Consulta: 11 gener 1999].
30. Carl Fleishhauer, *Digital formats for content reproductions*, Washington DC: Library of Congress. American Memory, 13 July 1998, <<http://memory.loc.gov/ammem/formats.html>> [Consulta: 11 gener 1999]. Fleishhauer és coordinador tècnic del National Digital Library Program de la Library of Congress.

Resolució: De 500 x 400 a 1000x700 pixels³¹

Arxiu

Definició: Escala de grisos: 8 bits per pixel; color: 24 bits per pixel

Format: TIFF

Compressió: no comprimit

Resolució: de 3000x2000 a 5000x4000 pixels.

Format alternatiu

Ja hem indicat que en l'àmbit dels arxius d'imatges s'utilitza força el Kodak Photo-CD (Image Pac) com a format alternatiu, sobretot en funcions d'arxiu i referència. La *Library of Congress*, en canvi, no l'utilitza mai, per tal de no lliurar-se a un format propietari.

Materials textuais reproduïts com a imatges

Imatges tonals de manuscrits o documents impresos

Miniatura

En segons quins casos (per exemple, si els fitxers que es generen són petits) es podria obviar i utilitzar directament la imatge de "referència".

Definició: Escala de grisos: 4 bits per pixel; color: 8 bits per pixel

Format: GIF

Compressió: LZW

Resolució: Fins a 500 pixels horitzontals

Referència

Admeten que una alternativa a la presentada és la utilització de gràfics GIF.

Definició: Escala de grisos: 8 bits per pixel; color: 24 bits per pixel

Format: JFIF (JPEG)

Compressió: 5:1

Resolució: 150 dpi

Arxiu

Definició: Escala de grisos: 8 bits per pixel; color: 24 bits per pixel

Format: TIFF

Compressió: No comprimida

Resolució: 300 dpi

31. La *Library of Congress* actualment està fent els seus contractes de digitalització a una resolució de 640 x 480 (Carl Fleishhauer: *Digital formats...*).

Format alternatiu

La *Library of Congress* no ho fa, però altres institucions generen fitxers Adobe Acrobat (PDF) a partir d'alguna de les opcions anteriors i aquests són els que ofereixen als usuaris per a visualitzar i reproduir.

Imatges bitonals de manuscrits o documents impresos

Miniatura

Definició: Escala de grisos: 4 bits per pixel; color: 8 bits per pixel

Format: GIF

Compressió: LZW

Resolució: Fins a 500 pixels horitzontals

Referència i arxiu

Definició: Blanc i negre, 1 bit per pixel

Format: TIFF

Compressió: ITU Grup IV

Resolució: 300 dpi

Formats alternatius:

Adobe Acrobat (PDF) a partir d'alguna de les opcions anteriors.

*Mapes*³²

Miniatura

Definició: Color de 8 bits per pixel, amb separació RGB

Format: GIF

Compressió: LZW

Resolució: 72 dpi

Referència

Amb el programari adient es poden oferir les imatges en diferents resolucions segons les necessitats dels usuaris.

32. Library of Congress. National Geography and Map Division, *Scanning cartographic materials*, [Washington DC]: [Library of Congress. National Digital Library Program], <<http://memory.loc.gov/ammem/gmdhtml/gmddigit.html>> [Consulta: 11 gener 1999].

Definició: Color de 24 bits per pixel, amb separació RGB
l'ormat: MrSID (Multi-resolution Seamless Image Database)³³
Compressió: Wavelet
Resolució: 300 dpi

Arxiu

Definició: Color de 24 bits per pixel, amb separació RGB
Format: TIFF
Compressió: No comprimit
Resolució: 300 dpi

Electronic Text Center

VElectronic Text Center, de la *University of Virginia Library*, sembla que està aplicant un procés molt més simple a l'hora de digitalitzar tant documents històrics com d'altres més corrents: sense fer distinció de casos, només generen dos fitxers.³⁴

Referència

Definició: 24 bits per pixel (també quan no hi hagi color)
Format: JFIF (JPEG)
Compressió: A determinar en cada cas
Resolució: No determinada

Arxiu

Definició: 24 bits per pixel (també quan no hi hagi color)
Format: TIFF
Compressió: No comprimida
Resolució: 400 dpi

33. MrSID és un format propietari i d'utilització poc estesa, però que malgrat tenir pèrdua permet compressions de 22:1 sense que aquesta es detecti; el gran avantatge que té és que permet recuperar la mateixa imatge en múltiples. La Library of Virginia, entre d'altres, també l'utilitza. Vegeu l'empresa que el distribueix: LizardTech <<http://www.lizardtech.com>> [Consulta: 11 gener 1999].
34. *Archival digital image creation*. (Charlottesville, VA: University of Virginia. Electronic Text Center, July 1996), <<http://etext.lib.virginia.edu/helpsheets/specscan.html>> [Consulta: 11 gener 1999].

NARA

La *National Archives and Records Administration* (NARA) dels Estats Units, encarregada de preservar la documentació de l'administració federal, ha desenvolupat unes directrius per a la digitalització dels seus materials d'arxiu.³⁵ De moment, aquestes encara no tenen caràcter normatiu perquè formen part d'una iniciativa pilot,³⁶ i la institució adverteix que també estan bàsicament orientades a facilitar l'accés, més que no pas a la preservació. Malgrat això, les seves orientacions són tan completes com les de la *Library of Congress*: es donen quatre tractaments diferents segons la tipologia formal dels documents originals i cadascun genera tres fitxers: miniatura, d'accés (referència) i mestre (d'arxiu).

Documents textuais

Miniatura

<i>Definició:</i>	<i>Escala de grisos: 4 bits per pixel; color: 8 bits per pixel</i>
<i>Format:</i>	<i>GIF</i>
<i>Compressió:</i>	<i>LZW</i>
<i>Resolució:</i>	<i>Uns 200 pixels en la direcció més llarga 72 dpi</i>

Accés

Documents de més de 8,5" x 14"

<i>Definició:</i>	<i>Escala de grisos: 4 bits per pixel; color: 8 bits per pixel</i>
<i>Format:</i>	<i>GIF</i>
<i>Compressió:</i>	<i>LZW</i>
<i>Resolució:</i>	<i>12 dpi</i>

Documents de menys de 8,5" x 14"

<i>Definició:</i>	<i>Escala de grisos: 8 bits per pixel; color: 24 bits per pixel</i>
-------------------	---

35. Steven Puglia; Barry Roginski, *NARA guidelines for digitizing archival materials for electronic access*, (College Park, MD: National Archives and Records Administration, January 1998), 19 p. Disponible en format Acrobat a <<http://www.nara.gov/nara/vision/eaps/digguide.pdf>> [Consulta: 11 gener 1999]. Existeix també un resum en un sol quadre sinòptic a: Steven Puglia, *Electronic Access Project scanning and file format matrix* (College Park, MD: National Archives and Records Administration, January 1998), 19 p. Disponible en format Acrobat a <<http://www.nara.gov/nara/vision/eap/digmatrix.pdf>> [Consulta: 11 gener 1999].
36. *Electronic Access Project*, <<http://www.nara.gov/nara/vision/eap/eap.html>> [Consulta: 11 gener 1999].

Format: JFIF (JPEG)
Compressió: 20:1 o major
Resolució: 90 dpi

Mestre

Definició: Escala de grisos: 8 bits per pixel; color: 24 hits per pixel
format: TIFF
Compressió: No comprimida
Resolució: 200-300 dpi

Fotografies

Miniatura

Definició: Escala de grisos: 4 bits per pixel; color: 8 bits per pixel
Format: GIF
Compressió: LZW
Resolució: Uns 200 pixels en la direcció més llarga
12 dpi

Accés

Definició: Escala de grisos: 4 bits per pixel; color: 8 bits per pixel
Format: GIF
Compressió: LZW
Resolució: 12 dpi

Mestre

Definició: Escala de grisos: 8 bits per pixel; color: 24 bits per pixel
Format: TIFF
Compressió: No comprimida
Resolució: 3.000 dpi en la direcció més llarga

Mapes i dibuixos

Miniatura

Definició: Escala de grisos: 4 bits per pixel; color: 8 bits per pixel
Format: GIF
Compressió: LZW
Resolució: Uns 200 pixels en la direcció més llarga
12 dpi

Accés

Definició: Escala de grisos: 8 bits per pixel; color: 24 bits per pixel
Format: JFIF (JPEG)
Compressió: 20:1 o major
Resolució: 1.200 pixels en la direcció més llarga
120 dpi

Mestre

Definició: Escala de grisos: 8 bits per pixel; color: 24 bits per pixel
Format: TIFF
Compressió: No comprimida
Resolució: 200-300 dpi

Gràfics

Miniatura

Definició: Escala de grisos: 4 bits per pixel; color: 8 bits per pixel
Format: GIF
Compressió: LZW
Resolució: Uns 200 pixels en la direcció més llarga
12 dpi

Accés

Documents de més de 11" x 17"

Definició: Escala de grisos: 4 bits per pixel; color: 8 bits per pixel
Format: GIF
Compressió: LZW
Resolució: Uns 600 pixels en la direcció més llarga
12 dpi

Documents de menys de 11" x 17"

Definició: Escala de grisos: 8 bits per pixel; color: 24 bits per pixel
Format: JFIF (JPEG)
Compressió: 20:1 o major
Resolució: Uns 1.200 pixels en la direcció més llarga
120 dpi

Mestre

Definició: Escala de grisos: 8 bits per pixel; color: 24 bits per pixel
Format: TIFF

Compressió: No comprimida

Resolució: 200-300 dpi

MEMÒRIA DEL MÓN, DE LA UNESCO

El projecte de la UNESCO *Memoria del món*,³⁷ iniciat el 1992, té l'objectiu de fomentar la preservació del patrimoni documental mundial a partir de l'endegament de projectes pilot en diferents països. Una de les seves línies fonamentals d'acció és la digitalització de documents històrics de gran valor cultural, i per tal de portar-ho a terme s'han elaborat unes recomanacions.³⁸ D'una manera força confosa, s'hi indica que els conservadors dels documents hauran d'escollir en el moment de la digitalització entre algun dels tres nivells de qualitat que es proposen —consulta de base, ús intermig o còpia facsímil—, a més de crear un segon fitxer destinat a exercir funcions de catàleg. En canvi, a les especificacions tècniques només es mostren les dades corresponents al *fitxer d'úsintermig*.

Documents textuais moderns

De tipus corrent

Definició: 1 Bit per pixel

Format: TIFF 6.0

Compressió: ITU Gmp IV

Resolució: 200 dpi

Incorporant fotografies en blanc i negre

Definició: Escala de grisos: 8 bits per pixel

Format: TIFF 6.0

Compressió: JPEG sense pèrdua

Resolució: 100 dpi

37. <<http://www.unesco.org/webworld>> [Consulta: 11 gener 1999].

38. *Memory of the World Programme: general guidelines to safeguard documentary heritage* [Paris: UNESCO, november 1995], <<http://www.unesco.org/webworld/mdm/administ/en/guidelines.html>> [Consulta: 11 gener 1999].

Documents textuais antics

Monocroms

Definició: Escala de grisos: 4 bits per pixel
Format: TIFF 6.0
Compressió: JPEG sense pèrdua
Resolució: 100 dpi

Amb color

Definició: Color: 24 bits per pixel
Format: TIFF 6.0
Compressió: JPEG amb o sense pèrdua segons els casos
Resolució: 100 dpi

Fotografies

Sobre paper

Definició: Escala de grisos: 8 bits per pixel; color: 24 bits per pixel
Format: TIFF 6.0
Compressió: JPEG amb o sense pèrdua segons els casos
Resolució: 100 dpi

Sobre diapositiva

Definició: Escala de grisos: 8 bits per pixel; color: 24 bits per pixel
Format: TIFF 6.0
Compressió: JPEG amb o sense pèrdua segons els casos
Resolució: 200 a 2.000 dpi segons la compressió aplicada

Mapes

Definició: Escala de grisos: 8 bits per pixel; color: 24 bits per pixel
Format: TIFF 6.0
Compressió: JPEG sense pèrdua
Resolució: 100 dpi

THE MAKING OF AMERICA II

The Making of America (MoA) II és un projecte coordinat per la *Digital Library federation (DLF)* dels Estats Units per a investigar la problemàtica sorgida de la creació d'una biblioteca digital de materials arxivístics que funcioni de manera distribuïda; l'experiència és coordinada per la University of California at Berkeley i hi participen la Cornell University, New York Public Library, l'ensylvania State University i Stanford University. Com que tot just va iniciar les seves activitats el 1998, encara no ha tingut temps d'elaborar els seus procediments tècnics, però sí que ha indicat quins seran els seus nivells mínims, que podem considerar com els propis d'un dels millors projectes de biblioteca digital.⁴⁰

Definició: Escala de grisos: 8 bits per pixel; color: 24 bits per pixel

Format: TIFF sembla un bon candidat

Compressió: Sense pèrdua

Resolució: Mínim 600 dpi

No accepten formats propietaris com el Photo-CD.

HISTORIC AMERICAN SHEET MUSIC PROJECT

L'Historic American Sheet Music Project de la Duke University (Durham, NC, Estats Units) és un projecte pilot de digitalització de partitures musicals que va ser un dels guanyadors, el 1996-1997, del concurs que anualment convoca la *Library of Congress* al millor projecte dins d'aquest camp; actualment forma part de *The Digital Scriptorium* i s'ha anunciat que properament també s'integrarà dins les col·leccions de l'American Memory. La digitalització s'ha fet amb els següents paràmetres:⁴²

39. <<http://sunsite.berkeley.edu/moa2>> [Consulta: 11 gener 1999].

40. *The Making of America II testbed project while paper*, versió 2.0, (Berkeley): 15 September 1998), 42 p. Versió Acrobat disponible a <<http://sunsite.berkeley.edu/MOA2/vvp-v2.pdf>> [Consulta: 11 gener 1999].

41. <<http://SCRIPTORIUM.LIB.DUKE.EDU/>> [Consulta: 11 gener 1999].

42. *Creation of the images and database*, Duke University, <<http://SCRIPTORIUM.LIB.DUKE.EDU/SHEETMUSIC/TECHINFO-SCANNING.HTML>> [Consulta: 11 gener 1999].

Miniatura

Definició: 8 bits per pixel
Format: JFIF (JPEG)
Resolució: 100 pixels d'amplada
12 dpi

Referència

Definició: 8 bits per pixel
Format: JFIF (JPEG)
Resolució: 300 pixels d'amplada
i 50 dpi

Arxiu

Definició: Color: 24 bits per pixel
Format: JFIF (JPEG)
Resolució: 150 dpi

ARCHIVO GENERAL DE INDIAS

La digitalització d'una part important dels fons de *YArchivo General de Indias*, de Sevilla, és coneguda com una de les iniciatives d'abast més ampli que s'han realitzat a nivell mundial.⁴³ El mèrit d'iniciar el projecte el 1988, quan pocs s'ho plantejaven, també va comportar problemes com el fet de no poder disposar encara de formats ara populars com GIF i JPEG. Això els va portar a crear el seu propi algoritme de compressió amb un factor de reducció 2:1 i sense pèrdua d'informació; actualment s'està fent la migració a JPEG.

Definició: Escala de grisos: 4 bits per pixel (16 colors)
Format: JFIF (JPEG)
Compressió: Ajustada fins a una pèrdua màxima del 15%
Resolució: 100 dpi

43. Pedro González, *Computerization of the Archivo General de Indias: strategies and results*. [Council on Library and Information Resources, September 1998], <<http://www.clir.org/pubs/reports/gonzalez/contents.html>> [Consulta: 11 gener 1999].

ALTRES REFERÈNCIES⁴⁴

L'Scottish Cultural Resources Access Network (SCRAN) és un consorci que està creant una biblioteca digital amb documentació sobre la història i la cultura escoceses a partir dels recursos obtinguts de la loteria. Ells han escollit el format mestre el Photo-CD i a partir d'aquest generen altres fitxers JPEG per a fer possible l'accés via web. En casos de materials especials amb necessitats més sofisticades, han optat per usar el format TIFF sense compressió.

El *Science Museum* i el *Victoria & Albert Museum* de Londres han començat a digitalitzar part de les seves col·leccions documentals. En ambdós casos també han escollit Photo-CD com el format mestre i el JPEG com el d'accés immediat.

44. Neil Deagrie; Daniel Greenstein, *A strategic policy framework for creating and preserving digital collections: a report to the Digital Archiving Working Group*. Version 4.0. [London]: British Library Research and Innovation Centre, 1998. <<http://www.ukoln.ac.uk/services/papers/bl/framework/framework.html>> [Consulta: 11 gener 1998].