

El entorno digital

Formatos y preservación digital

LUIS TORRES FREIXINET

Jefe de la unidad de sistemas de reproducción del
Excmo. Ayuntamiento de Zaragoza

RESUMEN:

La tipología de los formatos digitales es una demostración práctica del fenómeno de la obsolescencia digital pero remarcada por un elemento clave: no siempre la supervivencia de un formato se debe a su grado de eficacia o efectividad técnica. Nos centraremos en unos pocos: tiff, jpeg, pdf y Djvu.

Por otro lado la Preservación Digital se nos presenta como el gran reto del futuro. Actualmente hay pocas soluciones, dada la complejidad del reto, pero ya podemos observar varias iniciativas muy interesantes. Nos centraremos más en el modelo teórico OAIS.

PALABRAS CLAVES:

tiff, pdf, jpeg2000, djvu, MODS, METS, JHOVE, OAIS

ABSTRACT:

The digital environment: Formats and digital preservation.

The digital formats' tipology is a clear example of digital obsolescence with a remarkable component: the survival of an specific format does not depend on its efficiency. We will focus in a few of these formats: tiff, jpeg, pdf, Djvu. On the other hand, the Digital Preservation seems to be the great challenge towards the future. Nowadays there are few solutions, given the complex nature of the problem, but some very interesting initiatives can be used at this moment. We will focus in the OAIS theoretical model.

KEYWORDS:

tiff, pdf, jpeg2000, Djvu, MODS, METS, JHOVE, OAIS

FORMATOS

La tipología de los formatos digitales es una demostración práctica del fenómeno de la obsolescencia digital pero remarcada por un elemento clave: no siempre la supervivencia de un formato se debe a su grado de eficacia o efectividad técnica.

El mercado acaba mandando al rincón más olvidado a aquellos formatos que no han tenido la gracia de recibir la bendición del apoyo de los usuarios.

Aunque nos vamos a centrar mayoritariamente en unos cuantos formatos no podemos dejar de mostrar una sucinta recopilación de los formatos de imagen (formatos de archivo gráficos) más habituales.

PRINCIPALES FORMATOS DE ARCHIVOS GRÁFICOS RASTERIZADOS

Extensión	Tipo MIME	Detalles
.art		Formato propietario. Software cliente de AOL soportando normalmente una única imagen fija muy comprimida.
.bmp	image/bmp	Formato de Windows puede recibir compresión sin pérdidas pero no todos los programas pueden abrirlos.Windows Bitmap

.cin	image/cineon	Es desarrollado por ANSI (Instituto Nacional Estadounidense de Estándares) y contiene cabeceras fijas. Es un subconjunto del ANSI/ SMPTE DPX. CINEON
.cpt		Sería el equivalente al formato pds (Adobe Photoshop) pero con un menor peso. Poca compatibilidad. Corel Photo-Paint Image
.dng	Application/dng	DNG 1.1 Adobe Digital Negative
.dpx	image/dpx	Desarrollado por Kodak con cabeceras de imagen flexibles y variables. Digital Picutre eXchange file format
.djvu	Application/djvu	Formato de tecnología wavelet. es un estándar abierto. La especificación de formato de archivo, así como una de las implementaciones de código abierto el decodificador (y parte del codificador) están disponibles.
.exr	image/exr	OpenEXR es el formato en código libre que sirve para las imágenes de alto rango dinámico (HDR). Extended Dynamic Range Image File Format
.fpx	image/vnd.fpx	Desarrollado por Kodak y HP permite varias resoluciones de una imagen. Con y sin compresión de 8 a 24 bits de profundidad de color (o bits por pixel). Flaxhpix
.gif	image/gif	Su popularidad reside en su uso masivo en internet. Es un formato de 8 bits que además puede darnos animación mediante frames. Como compresión usa el LZW. Graphics Interchange Format
.jpg/.jpe	image/jpeg	Usado ampliamente en internet. Comprime con pérdida pero esta puede ser ajustada. Hablaremos de él más adelante. Joint Photographic Experts Group
.Jpg2/.jp2		Podemos hablar del formato del futuro. Entramos en los formatos basados en la tecnología wavelet o de segmentación. Joint Photographic Experts Group
.sid	Application/sid	El formato MrSID (pronunciado Mister SID) es un estándar abierto de compresión de imágenes raster desarrollado por Los Alamos National Laboratory, un centro público de I+D de los EE.UU., y comercializado por la empresa LizardTech. Su acrónimo proviene de Multi-resolution Seamless Image Database y permite mostrar archivos digitales de gran tamaño con un tiempo de carga mínimo gracias a la tecnología wavelet. Cumple los requisitos del OGC
.pbm		Es un formato simple para gráficos en blanco y negro. Puede contener texto modificable por un procesador de texto. Está relacionado con los formatos .PGM (grises) y .PPM (color). Portable Pixmap
.pcd	image/pcd	Propietario de KODAK (Photo Cd) compresión con perdidas y 24 bit color. ImagePac Photo CD
.pcx		Paintbrush. Compresión sin perdida. Picture eXchange
.pgm		Es un formato simple para escala de grises. Puede contener texto modificable por un procesador de texto. Está relacionado con los formatos .PBM (binario) y .PPM (color). Portable Pixmap
.png	image/png	Formato desarrollado para intentar sustituir al .GIF en internet
.ppm		Es un formato simple para color. Puede contener texto modificable por un procesador de texto. Está relacionado con los formatos .PBM (binario) y .PGM (gris). Portable Pixmap

.psd	Application/x-photoshop	Formato de Adobe Photoshop con capas y poco compatible fuera del programa original.
.tiff/.tif		TIFF es definido comúnmente como el formato ideal para guardar los originales digitales. El formato es propietario de Adobe. Actualmente va por la versión 6 (año 1992), aunque presenta problemas de compatibilidad en alguno de los subconjuntos. Tagged Image File Format
.CR2 / .CRW .DCR/.KDC .MRW .NEF .ORF .PEF .RAF .SRF .X3F		Formatos que incluiríamos en el grupo de los denominados CAMERA RAW. Perteneciendo a los diferentes fabricantes. En el mismo orden: CANON KODAK MINOLTA NIKON OLYMPUS PENTAX FUJI SONY SIGMA
OTROS FORMATOS		
.eps	image/eps	Como salida de los dispositivos PostScript (.ps). Es un formato de documentos, creado por la empresa Adobe, para describir documentos listos para imprimir, es decir, documentos que ya están maquetados y preparados para ser impresos, pero que ya no podemos editar. Hay impresoras que entienden directamente este formato (impresoras PostScript). El formato Postscript se ha convertido, junto con el PDF, en un estándar para el almacenamiento de documentos listos para imprimir. Encapsulated PostScript
.dxf	image/vnd.dwg	Formato normalizado de texto ASCII usado para almacenar datos de programas CAD
.pdf	application/pdf	Realmente no es un formato gráfico. En realidad es una versión que nace del PostScript pero que actualmente, con sus continuas mejoras y actualizaciones ha superado (por la difusión) a éste. Propietario de ADOBE. Portable Document Format
.svg/.svgz		Formato vectorial basado en XML definido por W3C para ser usado en navegadores web. Scalable Vector Graphics

Aspectos a tener en cuenta desde la óptica de la preservación digital

Siguiendo las recomendaciones de la Library of Congress (EE.UU.) los podríamos resumir en siete puntos:

- 1. Divulgación. (Disclosure)** El grado en el que las especificaciones y herramientas de validación de la integridad técnica existen y son accesibles. Un espectro de los niveles de divulgación puede ser observado para formatos digitales. No es más importante la aprobación por un organismo de normalización reconocido, lo es la existencia y accesibilidad ilimitada de toda la documentación referida.
- 2. Aprobación. (Adoption)** Grado en que el modelo ya es utilizado por los principales creadores, divulgadores, o

usuarios de los recursos de información. Esto incluye utilizar un formato maestro, para su entrega a los usuarios finales, y como medio de intercambio entre sistemas.

- 3. Transparencia (Transparency).** Grado en el que la representación digital está abierta a un análisis directo con herramientas básicas, como por un editor de texto.

- 4. Auto-documentación. (Self-documentation)** Auto-documentación de los objetos digitales. Contienen elementos descriptivos básicos, técnicos, administrativos y otros metadatos.

- 5. Dependencias externas. (External Dependencies)** Grado en el que un formato particular depende de hardware, sistema operativo, o software para la prestación o el uso previsto y la complejidad de tratar con esas dependencias

externas en el futuro desde entornos técnicos.

6. Impacto de las patentes. (Impact of Patents) Grado de las instituciones de archivo para mantener el contenido en un formato que se inhibió por patentes.

7. Técnicas de mecanismos de protección. (Technical Protection Mechanisms) Aplicación de mecanismos como la encriptación que dan los contenidos de confianza.

No obstante existe una línea de trabajo americana muy importante de cara a la preservación seguida a través del Global Digital Formato de Registro (GDFR), una actividad de colaboración inicialmente estimulada por el interés común de en Harvard y en el MIT. El GDFR pretende con este esfuerzo mantener un registro activo que apoyará la ejecución de las operaciones sobre los formatos de archivos, para identificar, validar, e incluso poder llegar a transformarlos. Pero a partir de 2006, el mayor peso del desarrollo en el GDFR se está llevando a cabo en OCLC, utilizando fondos de la Fundación Andrew Mellon. El GDFR está relacionado con el desarrollo de la JHOVE conjunto de herramientas para la caracterización y validación de los formatos, y la propia Library of Congress no está exenta del mismo proyecto junto con las principales bibliotecas del mundo.

FORMATOS RAW, TIFF Y JPEG

Habitualmente en mis cursos cuando hablo del formato RAW lo suelo denominar el "FORMATO NO FORMATO".

En realidad estamos accediendo a los datos en bruto del sensor de la cámara. Y digo cámara porque actualmente este formato con sus diferentes extensiones (en función del fabricante/modelo) está restringido al uso de las cámaras fotográficas digitales, es decir, no podemos obtener un verdadero fichero RAW de un escáner.

Si disparamos dos fotos del mismo motivo, una en jpg con baja compresión (alta calidad) y otra en raw, seguramente se verá mejor la tomada en jpg; mejor nitidez/enfoque, mejor contraste, mejor iluminación, mejores colores,...

Esto es debido a que una cámara digital suele aplicar distintos filtros digitales para mejorar la imagen. Sin embargo, el formato raw nos muestra la foto tal y como el sensor la capturó, sin ningún filtro de mejora. Se verán unos colores más neutros, menos saturados, un enfoque más blando y una iluminación que dependerá más de la exposición que hicimos, más visiblemente sobre o subexpuesta si fuera el caso.

Sin embargo, una foto en jpg tiene 24 bits/píxel frente a 36 bits/píxel que suele contener el modo raw.

24 bits serán suficientes para ver toda la gama de colores posible, pero serán claramente insuficientes cuando queramos realizar ciertos ajustes (iluminación, corrección de tonalidades,...) a la imagen. Por otro lado, una imagen en formato raw, aunque en apariencia parezca más pobre, contiene muchísima más información y será posible ajustar todos los parámetros sin perder la información inicial dado que una vez "revelado" el fichero raw con el programa adecuado podremos guardar el resultado obtenido en el formato que deseemos.

Habitualmente se dice que RAW presenta un problema de "mayor peso" pero esto es si lo comparamos con jpeg, porque si lo comparamos con Tiff la cuestión se invierte. Por ejemplo nosotros en Zaragoza debido a los problemas de espacio de almacenamiento (ya vamos por los 8 Tb y seguimos creciendo) un tiff a color de una CANON EOS 1Ds genera casi 31 Mb, mientras que su equivalente RAW pesa 10 MB. Estamos hablando de casi el triple de ahorro y en altos volúmenes de producción es un factor interesante a tener en cuenta.

Tomar la decisión de usar este formato como formato "master" conlleva un riesgo que se puede asumir siempre que uno sea consciente del mismo. El problema es realmente cuando se toma una decisión desconociendo los riesgos que ésta implica.

Uno de los mayores problemas que presenta RAW es precisamente su normalización. Esta falta de normalización lo podemos incluso comprobar al ver como Photoshop tiene que ir publicando continuos "pluggins" para actualizar su programa y que pueda abrir los ficheros RAW de las cámaras fotográficas que van saliendo. En esta línea de actuación surge una iniciativa internacional para intentar conseguir que los fabricantes unifiquen (normalicen) y presenten un único formato RAW, nos estamos refiriendo a OPENRAW

Precisamente un primer intento de mitigar esta falta de normalización surgió de Adobe con el formato DNG que no deja de ser en realidad un símil del formato RAW, llegando incluso a tener la posibilidad de generar un fichero DNG que contenga el fichero RAW del que ha surgido.

Desgraciadamente la iniciativa de Adobe se encontró con el rechazo de los grandes fabricantes de cámaras fotográficas (CANON y Nikón, fundamentalmente) contrarias a dejar más "trozo del pastel" en manos de Adobe.

Es interesante leer el artículo de Michael Reichmann and Juergen Specht "The RAW Flaw" Para tener una aproximación más exacta a las particularidades de este formato que presenta unos niveles de excelencia tan altos como para que prestigiosos fotógrafos de todo el mundo

lo consideren el verdadero “negativo digital”.

Una de las cuestiones más interesantes a la hora de trabajar con el formato RAW es el comprender la diferencia entre las imágenes capturadas con las cámaras fotográ-

ficas de sales de plata y las digitales. Para poder entender mejor esta cuestión voy a trasladar el intercambio de pareceres (traducido) que destacados técnicos anglosajones tuvieron en un foro sobre el tema en cuestión (y que me perdonen mis amigos los filólogos ingleses...)

Fichero raw de 12 Bit	
Dentro del primer F/Stop, que contiene los tonos más brillantes	2048 niveles disponibles
Dentro del segundo F/Stop, que contiene tonos brillantes	1024 niveles disponibles
Dentro del tercer F/Stop, que contiene los Medios tonos	512 niveles disponibles
Dentro del cuarto F/Stop, que contiene los tonos oscuros	256 niveles disponibles
Dentro del quinto F/Stop, que contiene los tonos más oscuros	128 niveles disponibles
Fichero JPG de 8 Bit	
Dentro del primer F/Stop, que contiene los tonos más brillantes	69 niveles disponibles
Dentro del segundo F/Stop, que contiene tonos brillantes	50 niveles disponibles
Dentro del tercer F/Stop, que contiene los Medios tonos	37 niveles disponibles
Dentro del cuarto F/Stop, que contiene los tonos oscuros	27 niveles disponibles
Dentro del quinto F/Stop, que contiene los tonos más oscuros	20 niveles disponibles

Estuvimos bastante tiempo charlando sobre temas digitales a lo largo de nuestro viaje de 3000km por Islandia, y Thomas se extendió en uno en el que no había caído antes; no obstante en cuanto empecé a pensar en ello, observé que tenía mucho sentido.

Empecemos: rango dinámico Supongamos (para los propósitos de esta disertación) que una cámara digital SLR (N. del T: DSLR a partir de ahora) tiene un rango dinámico o latitud de 5 puntos (normalmente se acerca más a los 6, pero no divaguemos). Cuando trabajamos en modo RAW, en el que debería, la mayoría de las cámaras registran una imagen de 12 bits. (Sí, decimos que en modo 16 bits, pero la realidad es que sólo está registrando 12 en un espacio de 16. Mejor que 8, pero no tanto como lo serían 16 bits de verdad.)

Una imagen de 12 bits es capaz de almacenar 4.096 (212) valores tonales diferenciados. Uno podría pensar que por tanto, cada punto o f-stop de los 5 que conforman dicho rango será capaz de guardar unos 850 (4096/5) de dichos valores correlativos. Pero desgraciadamente, no es éste el caso. Lo que realmente sucede es que los datos pertenecientes al primer punto (el más brillante) consumirán 2048 del total de estos pasos; una mitad completa de los disponibles.

¿Por qué? Debido a que los chips CMOS y CCD son dispositivos lineales. Y por supuesto, cada f-stop registra la mitad de luminosidad que el anterior, y por tanto la mitad del espacio que resta.

Esta tabla lo ilustra:

Primer f-stop (tonos más brillantes) 2048 niveles.

Segundo f-stop (tonos brillantes) 1024 niveles disponibles.

Tercer f-stop (tonos medios) 512 niveles disponibles

Cuarto f-stop (sombras) 256 niveles disponibles.

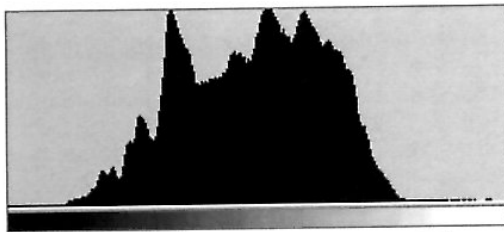
Quinto f-stop (tonos más oscuros) 128 niveles disponibles.

Este descubrimiento trae consigo una serie de lecciones importantes, siendo la más notable que si no utiliza el último quinto de la parte derecha del histograma para registrar parte de su imagen, estará desperdiciando de hecho una mitad de los niveles registrables por su cámara.

No obstante, como todos sabemos (o Vd. debería saber a estas alturas) lo peor que hay en imagen digital es quemar las luces; tanto como si estuviese tirando con diapositiva. Una vez que los ha quemado (ha sobrepasado la parte derecha del histograma) esos datos se han ido para nunca volver.

La lección

Lo que ha de extraer de esta pequeña lección es que debe tender a calcular sus exposiciones de modo que apriete el histograma hacia la derecha, pero no hasta el punto de quemar los altos. Esto suele mostrarse como una alerta parpadeante en las pantallas de revisión de la mayoría de las cámaras. En ese caso, vaya un poco hacia atrás hasta que desaparezca.



Exposición Normal
Histograma Centrado

zación y el ruido que potencialmente pudiera aparecer en las regiones más oscuras de la imagen.

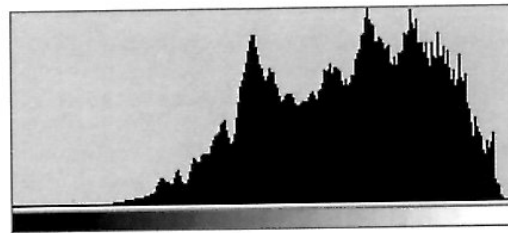
Tenga en cuenta que para obtener los resultados apropiados deberá llevar a cabo dichas correcciones mientras trabaja en el espacio 16 bits (12 bits) de un conversor RAW. A diferencia de lo que algunos piensan, en modo RAW la cámara no realiza ningún proceso no-lineal. Todo el proceso no-lineal lo lleva a cabo el conversor RAW. Ésta es la razón por la que deberá tirar en RAW y reajustar después la imagen en el conversor RAW antes de exportarla a Photoshop. Haciendo esto estará maximizando el ancho de banda disponible del sistema completo. Otra de las razones para tirar en RAW siempre que sea posible.

Tenga también en cuenta que al hacer esto estará reduciendo la sensibilidad ISO efectiva empleada para capturar la imagen, precisando de tiempos de obturación más lentos, o de mayores aperturas. Si está sosteniendo la cámara a mano, o fotografiando sujetos móviles, este inconveniente puede no valer la pena frente al nivel menor de ruido.

Un experimento

Esta técnica no es algo que tenga que emplear a diario necesariamente. Requiere un esfuerzo extra al fotografiar, y tanto esfuerzo como tiempo adicionales al post-procesar en el conversor RAW. Pero si le interesa asegurar el mayor ratio de señal/ruido en una foto en particular, puede valer la pena. Si a pesar de todo piensa que esta técnica es ilógica, o algo vaga, he aquí una idea innovadora: compruébelo usted mismo.

Lógicamente, cuando abra el fichero RAW en su programa de proceso RAW favorito, como Adobe Camera Raw, la imagen será demasiado clara. No hay problema. Utilice los controles deslizantes para modificar el nivel de brillo y contraste hasta que los datos del histograma se repartan equitativamente, y la imagen se vea bien. Con esto conseguirá una serie de cosas. La primera, maximizará la relación señal/ruido. La segunda, minimizará la posteri-



Histograma a la
Derecha

Prepare una fotografía con trípode; digamos que una escena estival típica, que incluya un cielo azul, nubes algodonosas y hierba verde. Mida como hace habitualmente y haga una foto. Revise el histograma. Ahora haga otro disparo sobreexponiendo a +1 o así, pero de modo que no sobreexponga los altos (las nubes por ejemplo). Asegúrese de que no haya nada en el fotograma que parpadee por sobreexposición, y que al mismo tiempo el histograma está todo lo desplazado a la derecha como sea posible.

Cargue la fotografía "expuesta correctamente" en un conversor RAW, haga sus correcciones habituales y mándelo a Photoshop. Cargue después su exposición-desplazada-hacia-la-derecha y haga lo mismo. Asegúrese no obstante de que utiliza primero los controles de gamma, brillo y contraste del conversor RAW para normalizarlos. Cargue esta foto en Photoshop.

Compare ambos. Fíjese en las áreas de sombra más oscura, buscando ruido o posterización. Compruebe si nota alguna diferencia. Si no la nota, no ha perdido nada con experimentar. Pero si sí lo percibe, acaba de aprender una nueva técnica que puede resultarle útil el día de mañana.

Reflexiones posteriores

Este artículo ha generado bastante discusión en la red, así como en el foro de discusión de Luminous Landscape. Uno de los peros que han surgido es que con la posible excepción de la Fuji S2 Pro, todas las demás DSLR actuales solo muestran histogramas globales basados en luminosidad, y no el brillo individual por canal. Esto es lo que Thomas Knoll tiene que decir al respecto...

Thomas Knoll:

Sólo uno o dos de los canales quemados (clipped) resulta problemático a veces. Con captura RAW no es tan pernicioso como con JPEG, dado que con RAW lo que debe importarle es quemar (clip) en el espacio de color nativo de la cámara, y no en el espacio de trabajo RGB (puesto que la conversión a espacio de trabajo tiene lugar después de los ajustes tonales en el conversor RAW). Hay muchos colores que saturan (clip) en RGB o incluso en Adobe RGB que no lo hacen en el espacio de color nativo de la cámara.

La solución ideal sería que los fabricantes modificasen el indicador de sobreexposición de modo que parpadeara en un color basado en alguno de los que constituyen el espacio de color nativo de la cámara, mejor que en la luminosidad. Esta es una de las razones por las que el histograma de Adobe Camera Raw está constituido por la superposición de los canales de color y no está basado únicamente en luminosidad.

En la práctica, este problema no es tan serio, dado que los objetos más brillantes de una escena suelen ser neutros (muchas veces nubes blancas). Cuando no lo tenga claro, siempre podrá recurrir al bracketing y comprobar los histogramas de color resultantes para elegir con cuál trabajar.

El gurú digital Ian Lyons tiene también algunos comentarios para esta técnica, profundizando en la comparación de la fotografía digital frente a ciertos aspectos de la exposición que damos por sentado y que tienen su origen en el trabajo con película.

Ian Lyons:

Básicamente, la exposición ideal es la que describe Michael: procure que el histograma esté todo a la derecha que sea posible sin que el indicador de sobreexposición parpadee en las zonas quemadas. La exposición ideal garantiza que contará con el máximo de niveles disponibles para describir su imagen sin perder detalles relevantes en las luces. Cuanto más se acerque a este ideal, más de estos niveles estarán empleándose en describir las sombras. Si subexpone una imagen hasta el punto de que las sombras se bloquean, que es lo que muchos hacen para proteger sus luces, necesitará abrirlas nuevamente para asegurar que la imagen final resulta como usted quiere. El problema con esta técnica es que sólo contamos con 128 niveles para las sombras. Necesitará emplearse a fondo en las curvas para abrir las sombras, encontrándose entonces con posterización, etc.

Necesitamos desprendernos de los conceptos sobre exposición que tan bien nos has servido con los carretes. Un CCD/CMOS no es película, y no reacciona como tal

en las regiones de luz/sombra. La exposición con película tiende a sobrepasar los límites (roll off) de las sombras o luces con suavidad. En el mundo digital, la captura es lineal, y no hay tal suavizado. Desafortunadamente (como Thomas ha señalado) el comportamiento de estos sensores no es perfecto, y podemos (lo hacemos con frecuencia) saturar uno o dos canales (quemándolos). Esto era un problema serio con la Canon D30; y la versión actual de Adobe Camera Raw simplemente no podía con estas tomas. Thomas sabe cuáles son mis pensamientos, y yo conozco los suyos.

Merece la pena señalar que no he visto todavía una imagen RAW lineal de una Canon EOS D30/D60 o 10D que ocupase toda la escala, las he visto que casi, pero nunca toda. Esto significa que a pesar de todo, los programas existentes de conversión RAW pueden no cumplir sus necesidades. En cuanto a estas aplicaciones, tengo la seguridad de que las cosas mejorarán; en el ínterin urjo a la gente a examinar esta técnica. Recuerde que es muy probable que conserve el fichero CRW durante bastante tiempo. En realidad, es su análogo del negativo; no lo tire a la basura. Incluso si la oferta actual de programas de conversión no es capaz de tratar las luces quemadas, las aplicaciones futuras sí lo harán. No obstante, NADA le permitirá recuperar el detalle perdido en las sombras.

En este ejemplo, podrá ver cuánta información existe todavía en las luces altas de una imagen que los conversores actuales RAW dicen que está quemada. Esta técnica es prácticamente redundante, pero aún así debería facilitar su comprensión de qué cantidad de información retienen las zonas de luces.

Hay quien piensa en la imagen digital como película transistorizada. No es este el caso. La explicación que sigue, de Bruce Lindbloom, le puede ayudar a comprender qué es lo que sucede realmente cuando emplee la técnica anterior...

Bruce Lindbloom:

En la fotografía basada en película, la parte de la escala correspondiente a las luces, está comprimida por la zona del hombro de la curva D/log E. A medida que se fotografían objetos más y más brillantes, el detalle de las luces se comprime más y más hasta que finalmente la película satura. Pero hasta ese punto, la compresión de las luces tiene lugar en una progresión gradual.

Los sensores de estado sólido de las cámaras digitales se comportan de un modo distinto. Conforme la luz incide sobre un sensor, este responde acumulando o disipando una carga (dependiendo de la tecnología del sensor). Su comportamiento es aceptable justo hasta el punto de saturación máximo, en el que la respuesta cesa

abruptamente. Los sensores no perdonan con un cese de respuesta gradual y progresivo como hace la película.

Debido a esta diferencia, calcular la exposición empleando una tarjeta de 18% de gris (como se hace habitualmente con película) no funciona tan bien con una cámara digital. Obtendrá mejores resultados si expone de modo que el blanco más blanco de la escena se acerca, pero no llega, a la escala digital completa (255 para captura a 8bit, 65535 para captura a 16bit). Base la exposición en las luces con cámaras digitales, y en un tono medio (p. ej. tarjeta 18% de gris) con las de película.

TIFF responde a la denominación en inglés "Tagged Image File Format" (formato de archivo de imágenes con etiquetas) se debe a que los ficheros TIFF contienen, además de los datos de la imagen propiamente dicha, "etiquetas" (metadatos) en las que se archiva información sobre las características de la imagen, que sirve para su tratamiento posterior.

El formato TIFF fue desarrollado por Aldus y Microsoft, y es actualmente propiedad de Adobe Systems. La última revisión del formato es la número 6, del año 1992. Hay algunas extensiones, como las anotaciones que utiliza el Imaging de Microsoft, pero ninguna puede considerarse estándar.

Podríamos decir que tenemos actualmente los siguientes grupos:

TIFF para la revisión 6.0

TIFF G4, que es un tiff bitmap con la compresión establecida, limitada a imágenes bitonales. Por ejemplo indicar que el Government Printing Office (GPO) de EE.UU lo tiene como formato master para los ficheros bitonales.

TIFF LZW que es un tiff bitmap con la compresión LZW hasta hace muy poco compresión sometida a patente (al igual que el .gif)

TIFF UNC que es un tiff bitmap sin compresión. Por ejemplo indicar que es considerado como master de conservación por el Florida Digital Archive

TIFF/IT para imágenes de pre-impresión. Viene definido por la norma ISO 12639:2004 (es decir: "Tagged Image File Format / Image Technology") es un estándar para el intercambio y envío de anuncios digitales y de páginas completas. Los ficheros TIFF/IT sólo contienen datos en mapa de bits (bitmap data), sin datos vectoriales. Los ficheros no han sido interpretados (rasterized) —aunque podrían haberlo sido— pero contienen 256 niveles de gris por canal.

Ahora bien lo habitual que podemos leer es el considerar el formato TIFF como el formato definido como "master digital" por excelencia. En una comparación con el formato tiff podríamos resumirla mediante el siguiente cuadro:

RAW	TIFF
MÁXIMA CALIDAD DE IMAGEN	ALTA CALIDAD DE IMAGEN
ALTA CAPACIDAD DE MANIPULACIÓN SIN PÉRDIDAS	POCA CAPACIDAD DE MANIPULACIÓN SIN PÉRDIDAS
ALTA PROFUNDIDAD DE COLOR	PROFUNDIDAD DE COLOR ESTÁNDAR
ARCHIVOS DE TAMAÑO MEDIO	ARCHIVOS DE GRAN TAMAÑO
VELOCIDAD DE ALMACENAMIENTO MEDIA	VELOCIDAD DE ALMACENAMIENTO LENTA

Todas las cámaras disparan en RAW pero no todas las cámaras nos ofrecen la posibilidad de acceder a estos ficheros (recordemos que son los datos en bruto capturados por el sensor).

De esta comparación, y de los datos aportados anteriormente podemos llegar a comprender porque los fotógrafos profesionales cuando tienen que realizar trabajos de calidad con la tecnología digital optan sin ningún tipo de dudas por capturar en formato RAW.

El debate de guardar como master digital el RAW o el TIFF se puede decidir (y eso lo debe hacer cada uno) preguntándose como responden estos formatos a los siete puntos enunciados al principio y luego en las posibilidades que su trabajo puede rendir según su organismo o

entidad.

Dentro de este ámbito la industria se encontró pronto con la necesidad de que los ficheros en color fueran "ligeros" y con una capacidad de representación visual de los colores (o grises) significativa y sobre todo escalable a voluntad por el usuario. En ese instante surgió el formato JPEG (Joint Photographic Experts Group).

El uso de diferentes algoritmos para realizar la compresión permitió un rápido desarrollo de este formato. JPEG es abreviado muchas veces por JPG debido a que algunos sistemas operativos que no permiten extensiones de fichero con más de tres letras.

Realiza una compresión con pérdida, de manera que al descomprimir la imagen no obtenemos exactamente la

misma imagen que teníamos antes de la compresión.

¿Cómo trabaja jpeg? Una explicación sencilla sería el observar un paisaje de un bosque en el que podemos tener (por poner una cifra) 100 niveles de verdes diferentes. Si quisiéramos un jpeg muy comprimido tendríamos una imagen en lugar de con 100 niveles de verdes con 10. Cuanto más comprimamos más se reducirían los niveles de verde. El motivo no es otro que sustituir "elementos" que considera repetidos y que sean sustituidos por otros.

Lógicamente no es una explicación científica pero sirve para que se pueda entender el funcionamiento interno del proceso de jpeg.

Dicho de otra manera el sistema de compresión que usa JPEG se basa en reducir información promediándola en las zonas de degradado.

Es decir que se calcula el valor de color de algunos píxeles en función del color de los píxeles que les rodean. Por esas características este formato es muy eficiente a la hora de almacenar imágenes que posean muchos degradados y matices de color.

Aunque también es preciso tener en cuenta que es casi inútil cuando queremos guardar con este formato dibujos con grandes extensiones de colores planos y uniformes o con bordes muy definidos, definitivamente no se recomienda.

Ese es uno de los motivos por el que una imagen bitonal (mapa de bits) no puede convertirse nunca en jpeg.

El algoritmo de compresión JPEG se basa en dos defectos visuales del ojo humano, uno es el hecho de que es mucho más sensible al cambio en la luminancia que en la crominancia, es decir, notamos más claramente los cambios de brillo que de color. El otro es que notamos con más facilidad pequeños cambios de brillo en zonas homogéneas que en zonas donde la variación es grande, por ejemplo en los bordes de los cuerpos de los objetos.

El algoritmo JPEG, transforma la imagen en cuadrados de 8x8 y luego almacena cada uno de estos como una combinación lineal o suma de los 64 recuadros que forman esta imagen, esto permite eliminar detalles de forma selectiva, por ejemplo, si una casilla tiene un valor muy próximo a 0, puede ser eliminada sin que afecte mucho a la calidad.

Para comparar las diferencias entre el bloque original y el comprimido, se halla la diferencia entre ambas matrices, la media de sus valores absolutos, da una ligera idea de la calidad perdida.

No dejan de ser la aplicación de unas fórmulas matemáticas u otras, por lo que siguiendo los principios establecidos en la normalización del jpeg podemos encontrar programas (fórmulas) más eficientes que otros.

La gran ventaja de JPEG frente a otros formatos similares es el alto grado de homogeneidad conseguido y sobre todo en su escalabilidad.

Un inconveniente que no debemos olvidar es que cada vez que guardamos el documento en jpeg se realiza la compresión. Esto quiere decir que si vamos trabajando con el documento y lo vamos guardando en jpeg entre paso y paso iremos acumulando pérdidas en función de la ratio de compresión. Por eso lo normal es trabajar con tiff o png y sólo en el paso final guardarlo en jpeg.

Pero todo avanza y JPEG no podía ser menos. El avance que está teniendo el llamado JPEG XR (JPEG 2000) está siendo tan importante que ya está planteando que esta evolución del jpeg podría convertirse en un interesante formato como master de conservación.

Se que está adoptando de forma experimental. Es muy interesante la experiencia de la Harvard University Library (HUL) que realizó una migración automatizada de sus TIFF a JPEG2000 con un éxito casi total, de un 99% (los errores se debieron al codec y se pudieron solucionar).

Son muy interesantes una serie de artículos de Margaret Lepley (recomiendo vivamente su lectura), sobre todo por ir a lo concreto y realizar pruebas y comparaciones técnicas de alto nivel. Pero ¿Cuál es la diferencia fundamental del JPEG2000? JPEG 2000 puede trabajar con niveles de compresión mayores a los de JPEG sin incurrir en los principales defectos del formato anterior con altas tasas de compresión: Generación de bloques uniformes y aspecto borroso. También se adapta mejor a la carga progresiva de las imágenes. Sus principales desventajas están en que tiende a emborronar más la imagen que JPEG incluso para un mismo tamaño de archivo (pero sin formar bloques), y que elimina algunos detalles pequeños y texturas, que el formato JPEG normal sí llega a representar. No obstante esto último el desarrollo mediante diferentes patentes nos ofrece diferentes "soluciones" algunas con más calidad que otras. Y esto si que por ahora puede resultar un problema, debido a que no todos los programas o navegadores son capaces de trabajar/visionar estos ficheros.

Trabaja de forma secuencial (baseline), comprime por wavelets (Discreta Wavelet DWT), con frecuencias mediante ondas lo que permite una pérdida menor de detalle. Por otro lado el estándar básico está libre de patentes aunque es un tema muy complejo y no está claro. Las licencias se asocian a las diferentes codificaciones.

El formato JPEG2000 ha sido publicado como norma ISO/IEC 15444-1:2000

Es un formato de futuro y habrá que estar atentos a

las experiencias internacionales.

Un ejemplo del desarrollo sería el formato Musid. MrSID son las siglas de Multi-resolution Seamless Image Database. MrSID se basa en poderosos algoritmos wavelet de codificación y potentes capacidades de descompresión. Está diseñado especialmente para que los profesionales de los GIS consigan plena portabilidad de imágenes masivas, gigantes. MrSID permite la visualización y manipulación instantánea de imágenes, tanto en local como en red, sin sacrificar la calidad de la imagen. Incluye características como multiresolución, descodificación selectiva, mosaicado y navegación en Internet.

Precisamente este es el formato ideal para representar los fondos de las cartotecas, como por ejemplo la de la Library of Congress o el Instituto Cartográfico de Cataluña.



MrSID se basa en una "arquitectura plana", que maneja "bitplanes". Un "bitplane" es la representación de un subconjunto de la imagen, codificada de forma óptima para un grado de calidad y resolución determinados.

- Una imagen MrSID es un conjunto de "bitplanes".
- Todos los "bitplanes" son opcionales y combinables. Es posible conseguir una imagen de alta calidad incluso sin disponer de todos sus "bitplanes".
- Una transacción de una imagen MRSID consiste en extraer y entregar sólo y únicamente los "bitplanes" necesarios para construir la vista requerida por el usuario (escena, escala y calidad visual), independientemente de las restricciones del ancho de banda.

Dejamos para el final dos formatos que pueden funcionar inicialmente con los mismos destinos: PDF y DjVu

El formato de documento portátil (PDF) de Adobe fue inventado por Adobe Systems y perfeccionado durante 15 años, gracias a lo que le permite obtener y visualizar información desde cualquier aplicación y en cualquier sistema informático. Actualmente se ha convertido en un estándar de facto, de manera que enviar un fichero en formato pdf "garantiza" que el receptor del mismo, esté donde esté podrá abrirlo.

¿Que dice Adobe del PDF? Veámoslo:

Es un Formato abierto: es un estándar para el intercambio de información electrónico fiable, reconocido internacionalmente por empresas y organismos gubernamentales. Compatible con estándares de la industria, como PDF/A, PDF/X y PDF/E.

Multiplataforma: se puede ver e imprimir en cualquier plataforma Macintosh, Microsoft® Windows®, UNIX® y múltiples plataformas móviles.

Extensible: más de 1.800 vendedores en todo el mundo ofrecen soluciones basadas en PDF, que incluyen plug-in, consultorías, formación y herramientas de soporte.

Fiable y robusto: el hecho de que haya más de 200 millones de documentos PDF circulando en la red hoy en día es la prueba de la cantidad de empresas que confían en este formato para obtener información.

Mantenimiento de la integridad de la información: los archivos PDF de Adobe tienen el mismo aspecto y muestran la misma información que los archivos originales, como, por ejemplo, texto, dibujos, 3D, gráficos en color, fotos e incluso lógica empresarial, independientemente de la aplicación utilizada para crearlos.

Seguridad de la información: firme electrónicamente o proteja con contraseña los documentos PDF de Adobe creados mediante el software Adobe Acrobat® 7.0 o Adobe LiveCycle™.

Capacidad de búsqueda: Con funciones de búsqueda de texto para encontrar palabras, marcadores y campos de datos en los documentos.

Accesible: los documentos PDF de Adobe colaboran con las tecnologías de asistencia para facilitar el acceso a la información a personas con discapacidades.

Hasta aquí todo son bondades pero PDF presenta también algunos inconvenientes:

El tratamiento que realiza de las imágenes (en jpeg) hace que para conseguir que el mismo pese poco la ratio de compresión tenga que ser muy alta, y esto en algunos entornos es conflictivo, por ejemplo en el campo del patrimonio documental y bibliográfico.

Adobe presentó a principios del 2007 la especificación completa de su formato PDF en versión 1.7 a la Association for Information and Image Management, que se está poniendo manos a la obra para liberarla como un estándar ISO

La medida de Adobe supone un verdadero impulso para uno de los formatos más utilizados actualmente a la hora de transmitir todo tipo de documentos: su Portable Document Format ya está en manos del organismo que se encargará de su publicación como un formato ISO abierto.

La medida, según señalan en DesktopLinux, es la evolución lógica de una migración gradual de diversas versiones específicas del formato a estándares ISO, ya que PDF/Archive y PDF/Exchange ya lo son, mientras que PDF For Engineering y PDF for Universal Access están en vías de serlo muy pronto.

Según Sarah Rosenbaum, directora de gestión de productos en Adobe, esto permitirá la adecuación de PDF en todo tipo de entornos de producción y archivado. En este importante anuncio Rosenbaum afirmó que el creciente aumento del interés por el formato abierto de documentos (ODF) ha sido también clave: "hay una tendencia cada vez mayor en el mercado hacia la normalización en formatos abiertos"

No obstante, según ella esta medida no tiene nada que ver con el formato rival de Microsoft, el llamado Metro, más conocido por sus siglas: XPS (XML Paper Specification), y que está muy presente tanto en Windows Vista como en Office 2007, y no sólo para usuarios empresariales.

Pero PDF por si mismo no es adecuado como formato de archivo:

- Puede incluir características incompatibles con los requisitos actuales de archivo.
 - ◊ Cifrado
 - ◊ Ficheros anexos
- Los documentos PDF no necesariamente son auto-contenidos:
 - ◊ Pueden depender de las fuentes del sistema y otro contenido que resida fuera del archivo.
- Hay múltiples herramientas de creación de PDF's en el mercado
 - ◊ Inconsistencia en el formato de archivo (no todos los pdf's se han creado de la misma manera)

Por eso surge una versión "estándar" (o controlada) de PDF. Se hacía necesario asegurar la preservación de documentos PDF durante largos periodos de tiempo y garantizar que los documentos PDF serán visualizados de manera predecible y consistente en el futuro.

El PDF/A

- es un formato para archivo final, no para archivos vivos.
- ISO especifica el uso de PDF/A para la preservación de documentos electrónicos a largo plazo.
- PDF/A no cubre métodos físicos específicos de almacenamiento de estos documentos, como los medios o condiciones de almacenamiento, o el hardware necesario o los sistemas operativos

Respecto a los PDF/A Nos encontramos con

Recomendaciones:

- Formato persistente e independiente del dispositivo.
- Desarrollo basado en un comité.
- Existen soluciones de otros fabricantes.
- Empleo de metadatos.
- Compresión sin pérdidas

Obligaciones:

- Estructura etiquetada (PDF/A-1a)
- Preservación del aspecto visual (PDF/A-1b)
- Fuentes incrustadas
- Anotaciones claramente identificables.

Prohibiciones:

- Cifrado y protección mediante contraseñas
- Objetos dinámicos incrustados
- Fuentes propietarias o no incrustadas.
- Enlaces externos

Más información sobre PDF/A: www.aiim.org/standards

La evolución del mercado y de las necesidades propició la aparición del otro formato comentado: DjVu

La tecnología de DjVu fue originalmente desarrollada por Yann Le Cun, Léon Bottou, Patrick Haffner y Paul G. Howard en los laboratorios de AT&T en 1996. DjVu es un formato de fichero abierto. Las especificaciones del formato y el código fuente de la biblioteca de referencia están publicadas y se encuentran disponibles. La propiedad de los derechos para el desarrollo comercial del software de codificación ha sido transferido a distintas compañías a través de los años, incluyendo AT&T y LizardTech. Los autores originales mantienen una implementación GPL llamada DjVuLibre.

El año 2002 el formato DjVu ha sido elegido por Internet Archive como formato en el cuál su proyecto "Million Book Project" proporciona libros escaneados de dominio público de forma online (conjuntamente con TIFF y PDF).

DjVu trabaja dentro de un entorno wavelet por el que genera diferentes capas del documento digitalizado. A cada una de estas capas le aplica un formato de compresión determinado.

Lógicamente el quid de la cuestión es lo que el formato interpreta que debe ir en cada capa. Así tenemos (como más importantes) la capa de visión global, la capa de texto y la capa de fondo.

Una traslación completa de la complejidad del formato DjVu la podemos obtener en el documento elaborado por Leon Bottou.

Es necesario tener instalado el plugu de DjVu para

poder acceder al documento, bien el de Lizardtech o bien el desarrollado por la comunidad "open source" llamado DjView 1.0

No obstante finalizo el debate entre PDF y DjVu con un ejemplo.

Dov Isaacs, uno de los mayores expertos de Adobe creó (año 2001) un pdf con las siguientes características:

- ◊ 84 slides from PowerPoint
- ◊ Color graphics on every page
- ◊ 30 font faces subset embedded
- ◊ 17 line art drawings
- ◊ 30 screen shots
- ◊ 28 bitmaps (in addition to the screen shots)
- ◊ Four languages
- ◊ Looks great on-screen
- ◊ Prints like a champ
- ... y un tamaño de 1,14 Mb

En febrero del 2002 Leon Bottou usando el mismo documento lo dejó en Djvu en 0,88 MB, conservando las mismas características.

Tengo claro que PDF mantiene la delantera sobre Djvu un muchos aspectos, por ejemplo su uso e implantación masivo y PDF/A, pero no podemos desdeñar la grandes posibilidades que nos presenta DjVu gracias a la compresión wavelet.

DjVU vs JPEG?

El uso intenso de jpeg nos hace olvidar muchas veces que DjVu no deja de ser "un tipo de jpeg" pero mejorado.

Uno de las cuestiones más prácticas es la conversión de documentos a Djvu asignándole el perfil "photo" de manera que perderemos las capas (si es una foto no tiene sentido crear una capa de texto, de fondo, etc).

De esta manera obtenemos unas imágenes en Djvu (es decir en jpeg mejorado) con:

- Calidad mucho más alta que JPEG's realizados de modo similar, o aproximadamente con la mitad de tamaño para una calidad similar.
- Calidad superior en altas relaciones de compresión. Con degradados más realistas, sin los típicos artefactos digitales de "bloque", y ninguna "solarización" del color.
- Aparece de manera muy rápida la imagen con alta progresividad.
- Muy rápido para mostrar el zoom o la decodificación.
- Uso de poca cantidad de memoria para la descom-

presión lo que facilita el visionado de las imágenes en ordenadores poco potentes.

Como las imágenes valen más que mil palabras os animo a comprobarlo directamente en <http://djvu.org/resources/jpgvsdjvu/comparison.html>

¿Cual puede ser el futuro? Las últimas tendencias nos muestran PDF's que en realidad son DjVu. Es decir, creas un DjVu con su estructura de segmentación (o capas) y manteniéndolas obtenemos un fichero con extensión PDF

Algunos piensan que significará la desaparición de DjVu pero otros muchos pensamos que obtendremos toda la potencia en el tratamiento de los textos y de las imágenes de DjVu con toda la UNIVERSALIZACIÓN de PDF.

Estaremos expectantes al resultado final. Por ahora los intentos comerciales existentes no logran dar el 100% de la potencia de DjVu en su traslación a PDF y presentan además un gran inconveniente (a mi entender). Pretenden vendernos licencias de software limitadas por número de ficheros realizados, algo por lo que nunca deberemos pasar.

Lo bueno es que el futuro ya está aquí.

PRESERVACIÓN DIGITAL

La realidad supera ampliamente a nuestra imaginación cuando tenemos que enfrentarnos al reto del almacenamiento y preservación de las imágenes digitales a largo plazo.

Y hablo de reto usando un término "suave" dada la problemática que se nos avecina. Son muchos los sectores los afectados por el problema del almacenamiento de los documentos electrónicos, pero en todos los casos es necesario que se garantizar:

- Que sean AUTÉNTICOS.
- Que sean FIABLES.
- Que estén INTEGROS.
- Que estén DISPONIBLES.

La problemática y variedad existente hará que nos centremos en este artículo en lo que entendemos todos por imágenes digitales. Dejando a un lado todo lo relacionado con bases de datos, hojas de cálculo y/o aplicaciones corporativas. Todo ello a pesar de presentar elementos coincidentes en cuanto a las políticas de seguridad y almacenamiento.

Como información podríamos citar las normas que actualmente tratan este tema:

- Trabajos previos del TC 171
- ISO/TR 15801

- ISO/TR 18492
- ISO 19005/A
- Proyecto 26102

Las exigencias de digitalización del Patrimonio Documental difieren en gran medida de las de entornos de oficinas y similares. Además de poder garantizar la mayor fidelidad de la imagen reproducida debemos garantizar que en el futuro éstas sean accesibles. Existe una necesidad imperiosa de trabajar de manera multidisciplinar sabiendo que la tecnología digital nos abre muchas puertas, pero que también nos convierte en esclavos de su propia evolución tecnológica. Evolución que por el ritmo que lleva bien podríamos denominar revolución.

Anticiparse y planificar son las palabras claves, y naturalmente vale la pena, ya que de lo contrario se perdería todo lo realizado hasta ese momento.

Estoy de acuerdo con vosotros en que es un problema, pero sabiendo la naturaleza del problema es más fácil solucionarlo. Toda institución que crea ficheros digitales tiene que tener una forma de migrar continuamente sus archivos de soportes obsoletos (o sin soporte seguros) a soportes actuales.

En lugar de pensar en los ficheros de imágenes como un almacén de datos estacionario, podemos pensar en ellos como en una tribu nómada que puede sobrevivir al paso de los años sólo si sabe moverse en busca de mejores territorios de caza.

Todos estamos de acuerdo en la necesidad de archivar los documentos electrónicos pero en el desarrollo de este tema mi objetivo es poder:

- Comentar un programa de autenticación y conservación para garantizar la permanencia del servicio de información.
- Describir los métodos propuestos para garantizar la permanencia y el acceso de los documentos digitales.
- Informar sobre experiencias y proyectos en curso.

Podríamos incluso como inicio plantearnos una pregunta ¿Por qué “archivar” los documentos digitales?. Aquí las respuestas podrían ser variadas pero fundamentalmente para garantizar la permanencia por su valor informativo, administrativo, legal, cultural (patrimonio): Dentro de esta casuística incluiríamos la digitalización de documentos no electrónicos para facilitar su acceso y difusión o como copia de seguridad.

Por ejemplo en la Unidad de Sistemas de Reproducción de Documentos del Ayuntamiento de Zaragoza se realizan dos tareas fundamentales:

- Por un lado la realización de las reproducciones puntuales solicitadas por los investigadores (excluyendo

las fotocopias).

- La labor de microfilmación y digitalización programada de parte de nuestros fondos.

La Ley 30/1992, de 26 noviembre, en su artículo 45.5 nos dice: *“Los documentos emitidos, cualquiera que sea su soporte, por medios electrónicos, informáticos o telemáticos por las Administraciones Públicas, o los que éstas emitan como copias de originales almacenados por estos mismos medios, gozarán de la validez y eficacia del documento original siempre que quede garantizada su autenticidad, integridad y conservación y, en su caso, la recepción por el interesado, así como el cumplimiento de las garantías y requisitos exigidos por esta u otras Leyes”.*

Este artículo fue posteriormente desarrollado por el Real Decreto 263/1996, de 16 de febrero, por el que se regula la utilización de técnicas electrónicas informáticas y telemáticas por la Administración General del Estado. Y las iniciativas, tanto de la Administración General del Estado como de otras administraciones (CC.AA., Entidades Locales, etc.) no hacen más que reafirmar la clara tendencia hacia la e-administración.

Ahora bien ¿cómo definiríamos el archivo de documentos electrónicos? Una buena definición podría ser: *“Conjunto de documentos producidos, recibidos o reunidos por una persona física o jurídica de modo involuntario, natural y espontáneo en el transcurso, y como apoyo, de su actividad, de la que es testimonio, haciendo uso de la electrónica; que se conservan y transmiten también mediante medios electrónicos en depósitos de conservación permanente, tras efectuar una selección a partir de la identificación y valoración de las series, con medidas de autenticación y de preservación adecuadas y con una organización respetuosa con su modo de producción, con el fin de garantizar su valor informativo, legal y cultural, así como de permitir su acceso y uso también mediante las tecnologías de la información”.*

Ya estamos inmersos en el proyecto y deberíamos marcarnos unos principios de gestión mínimos que garantizaran el resultado final. Estos principios deberían pivotar sobre:

- Reunión de los documentos basada en identificación, valoración y autenticación de la integridad documental.
- Plantear una política de conservación preventiva de los soportes y la legibilidad para garantizar la permanencia.
- Representación y organización basadas en la procedencia y el modo de producción documental.
 - ◊ Reconstrucción del orden original.
 - ◊ Descripción multinivel según jerarquía documen-

tal.

- Acceso y difusión por medios electrónicos.

Una vez que nos hemos marcado los mencionados principios de gestión nos encontraríamos con unas Áreas de Gestión que surgen de estos principios:

1. Reunión. Autenticación.
2. Preservación.
3. Organización.
4. Acceso.

La problemática aumenta de una manera exponencial si nos damos cuenta de los posibles modelos de documentos:

- Unitario: Coincide fichero y documento.
 - ◊ Ejemplos: ficheros texto, imagen, sonido y video.
 - ◊ La integridad se puede preservar mediante contraseñas, filigranas (watermarks),...
- Complejo: un documento está compuesto de diversos ficheros y a su vez puede dar origen a múltiples documentos unitarios.
 - ◊ Caracteres:
 - Δ Compuesto de diferentes documentos o partes de ellos.
 - Δ Distribuido entre redes y máquinas.
 - Δ Dinámico o de fácil actualización sin huella.
 - ◊ Ejemplos: bases de datos, workflow, correo electrónico, páginas web, intranets, sistemas de info geográfica...

La intención del tema es centrarnos en el modelo unitario, aunque en algunos casos la complejidad nos hará adentrarnos en el modelo complejo.

Una vez que tenemos (reunión) los documentos y cumplidos todos los pasos necesarios deberemos entrar en la fase de **autenticación**. Fundamental para los documentos electrónicos.

La finalidad no es otra que poder identificar los elementos de los documentos necesarios para mantener su autenticidad a lo largo del tiempo. Y además que nos permita verificar la integridad física y lógica de los documentos.

La autenticación es una actividad complementaria de la conservación ya que deberá ayudarnos a comprobar la calidad y la integridad durante la transferencia, copia y almacén de la dimensión lógica (obsolescencia tecnológica, virus informáticos...).

Los documentos electrónicos presentan una complejidad propia, que los diferencia de los documentos tradicionales, por ejemplo en las cuestiones relativas a lo que se denomina integridad documental. Para verlo con mayor claridad hablamos de un documento íntegro cuando

cumple estos requisitos:

- Completo: conserva todas sus partes (contenido y estructura)
- Fijo: no se puede variar ni su contenido ni su estructura.
- Referenciado: informa sobre su contexto administrativo y documental de producción.
- Fiable: su contenido es cierto. Es una noción relativa, que depende de las condiciones de creación del documento.
- Auténtico: originalidad, sin alteraciones. Depende de las condiciones de uso, transmisión y custodia.

Mientras que la creación y conservación electrónicas plantean:

- Dificultades de identificación (no coinciden entidad física y lógica).
- Dificultades para establecer los límites y partes componentes de los documentos complejos.
- Mayores posibilidades de manipulación. Criptografía.
- Sólo se pueden autenticar documentos conclusos.

La complejidad de las cuestiones relativas a la autenticación, aquí resumidas, nos lo demuestra la cantidad de iniciativas internacionales abiertas para la resolución de este problema...

- EROS (Electronic Records in Office Systems Project) Public Record Office, United Kingdom www.pro.gov.uk/drecordsmanagement/eros/default.htm
- Victorian Electronic Records Strategy Project (Public Records Office, Victoria, Australia) home.vicnet.net.au/~provic/vers/
- CERAR (Center for Electronic Recordkeeping & Archival Research), University of Pittsburgh.
- CEDARS (Curl exemplars in digital archives) www.leeds.ac.uk/cedars
- InterPARES (International Research on Permanent Authentic Records in Electronic Systems) www.interpares.org

De todos los citados podríamos detenernos con un poco más de detalle en el último citado: INTERPARES. Es una apuesta generada en 1999 mediante la creación de un consorcio de siete grupos de investigación repartidos por Canadá, EE.UU., Norte de Europa, Italia, Australia y Este de Asia.

La coordinación recae en la University of British Columbia y la base de todo el proyecto se recoge en una sola línea: The preservation of the integrity of electronic records

Lanza una propuesta de un modelo de actuación con cuatro objetivos:

- Reconocer e identificar un documento en el entorno digital: desarrollo de una tipología a partir de la determinación de sus caracteres internos y externos.
- Determinar qué tipos de sistemas informáticos generan documentos.
- Formular los criterios para segregar los documentos del resto de conjuntos de datos generados por los sistemas informáticos.
- Definir los requerimientos conceptuales para garantizar la fiabilidad y la autenticidad de los documentos electrónicos. Elaboración de guías de conservación.

EL PROGRAMA DE CONSERVACIÓN

Cuando iniciamos un programa de conservación los retos con los que nos enfrentamos son muchos pero podríamos resumirlas en estos tres puntos:

- Adoptar los soportes físico-lógicos más adecuados para conservar original y copia.
- Crear contenedores y depósitos seguros y con medidas ambientales adecuadas.
- Garantizar el acceso y la legibilidad de la información contenida en los soportes.

Este programa de conservación debería consolidarse mediante la creación de un plan de preservación junto con medidas de conservación preventiva y medidas de restauración. Y lo que es peor esto debe cuantificarse en recursos, tanto económicos como humanos, teniendo la absoluta certeza que no podrás prescindir de ellos en el futuro por la propia naturaleza del objeto electrónico y su obsolescencia.

Con un documento tradicional en función de muchos y cambiantes factores podríamos vernos obligados a aplicar técnicas de conservación mientras que con los digitales tenemos la seguridad de tener que aplicar estas técnicas en el futuro independientemente de cualquier otro factor. La diferencia radicarán en el resultado final que obtengamos.

El **plan de preservación** tendrá los siguientes elementos:

- Depósitos acordes con las directrices y normas internacionales de conservación.
- Debe existir un reglamento para el uso y la manipulación de los documentos.
- Plan de emergencia para actuar ante:
 - ◊ Desastres naturales: inundación, terremoto, incendio...
 - ◊ Conflictos bélicos.

- Plan y medidas de seguridad contra el robo:
 - ◊ De los documentos.
 - ◊ De la información. Implantar sistemas contra:
 - Δ Robo mediante medios telemáticos.
 - Δ Modificación o intervenciones no autorizadas.

Las **medidas de conservación** preventiva nos presentarían los siguientes elementos:

- Protección contra agentes de deterioro:
 - ◊ Naturales.
 - ◊ Artificiales.
- Reproducción de documentos como:
 - ◊ copias periódicas de seguridad:
 - Δ Duplicación de documentos.
 - Δ Duplicación de almacenes.
 - ◊ Modo de superar la obsolescencia de:
 - Δ Soportes magnético y ópticos.
 - Δ Hardware y software necesarios para la lectura.

Ahora bien podríamos preguntarnos en que tipo de soportes almacenamos la información ¿soportes magnéticos u ópticos?

Los soportes magnéticos, de aplicaciones digitales, son los discos flexibles, las cintas y los discos duros.

Su mayores inconvenientes como medio de conservación vienen de :

- la inestabilidad del material.
- lo efímero de su soporte debido a la elevada sensibilidad de sus componentes a cambios ambientales: corrosión, ruptura del aglomerante, degradación del adhesivo, etc.
- Riesgos de electromagnetismo.
- Desgaste por el uso: rozamientos y roturas por averías.
- Necesidad de preservar equipos para su reproducción.
- Perduración:
 - ◊ 20 años de expectativa de vida media.
 - ◊ Un siglo es una posibilidad si se almacenan a 5°C

Actualmente podemos encontrar discos duros internos de hasta 1 Tb con unas características impensables hace unos años.

Discos duros extraíbles: Sirve prácticamente todo lo que anteriormente hemos comentado, pero ahora hay que tener en cuenta que estos dispositivos son extraíbles. Con los últimos sistemas operativos (Windows XP) el usar uno de estos dispositivos es realmente sencillo. El sistema

en el momento que conectamos un disco duro externo lo reconoce. Sus capacidades y tamaño han evolucionado de la misma manera que los internos, Por eso ya no es raro encontrarlos de 500 Gb, 750 Gb ó incluso de 6 Tb. Además muchos fabricantes los construyen con la particularidad de poder unirse unos a otros, y así poder aumentar la capacidad de almacenamiento.

Son una opción muy interesante por velocidad y coste, frente a los dispositivos ópticos.

Hoy en día no deberemos en ningún caso adquirir un disco duro externo que no soporte una (o varias de ellas) de las siguientes conexiones: USB 2.0, Firewire I y Firewire II

Hay que decir que la información aparece en los discos duros dentro de las pistas, y estas se colocan de manera concéntrica, mientras que en los dispositivos ópticos (CD's, DVD) es en forma de espiral, lo que implica menor velocidad de acceso para éstos últimos.

Con los discos duros se pueden montar sistemas RAID en los que prima la seguridad sobre la eficacia, de manera que en caso de fallo mecánico de alguno de los discos al estar la información duplicada en otro disco no se producen pérdidas.

¿Que es RAID? Es el acrónimo del inglés "Redundant Array of Independent Disks". Significa matriz redundante de discos independientes. RAID es un método de combinación de varios discos duros para formar una única unidad lógica en la que se almacenan los datos de forma redundante. Ofrece mayor tolerancia a fallos y más altos niveles de rendimiento que un sólo disco duro o un grupo de discos duros independientes. (Ver anexo sobre SISTEMAS RAID)

Esto tiene sentido para información que tiene que estar disponible "al día" para información con una tasa de consulta menor podemos decantarnos por sistemas ópticos en cinta. Con gran volumen de almacenamiento pero una tasa de lectura (y por lo tanto de recuperación) lenta.

Los **soportes ópticos** son los tradicionales CD-ROM, DVD y multiples variedades que siguen surgiendo (Bluy Ray, etc.)

Los estándares de CD han ido evolucionando desde su aparición en el mercado. Están englobados en el libro denominado "El libro del Arco Iris", dividido en varios tomos según el color del láser que se usa en las unidades de CD-ROM (y que cambian de color a cada revisión). Citaré cuatro estándares principales:

- **Libro Rojo:** Definido en 1982 por los inventores del CD de audio: SONY y PHILIPS. Engloba a los CD de au-

dio. Las unidades de CD-ROM pueden leer perfectamente este tipo de CD.

- **Libro Amarillo:** Este estándar definió la forma en que se escriben los datos en el CD-ROM. El libro amarillo añade dos nuevos tipos de pistas, basándose en el libro rojo, conocida como Modo 1 y 2.

- ◊ El Modo 1 del libro amarillo se utiliza para los datos de ordenador.

- ◊ El Modo 2 del libro amarillo se usa para datos de audio y vídeo, normalmente comprimidos.

- **Libro Verde:** Añadido por Philips, incluye los CD-i. Es un formato que permite intercalar datos para ser almacenados en el disco.

- **Libro Naranja:** Trata los CD gravables: CD-R, CD-RW...

Respecto a los soportes de los CD-R, intentar clasificarlos por el color de su superficie es inútil, ya que no afectan al comportamiento del proceso de grabación/reproducción.

Cuando se trata de compatibilidad y calidad, las diferencias entre los soportes tienen mucho que ver con los procesos de fabricación, el master, las materias primas utilizadas y el control de calidad. El verdadero indicador es comprobar que funcionan bien en su equipo, ya sea en los regrabadoras o en los reproductores.

Los discos DVD (acrónimo de Digital versátil Disk) son el sustituto más reciente para los CD-ROM. Físicamente son iguales que ellos pero, lo que fundamentalmente mejoran sobre la tecnología anterior, es la capacidad: teóricamente hasta 17 GB. Esto es así porque se ha mejorado la densidad de escritura y los datos se almacenan en dos niveles gracias a una capa semitransparente. Los DVD, además, se pueden utilizar sobre las dos caras.

Los DVD se diseñaron con el propósito de almacenar datos, audio y vídeo. Existen diferentes modos de elaboración, para lo cual se han establecido 5 "Libros":

- **Libro A:** Para los discos de sólo lectura (DVD-ROM).
- **Libro B:** Para aplicaciones de vídeo (DVD vídeo).

Tipos de DVD

Tipo de disco	Capacidad	Caras	Capas por cara	
DVD-5	4,7 GB	1	1	Solo lectura
DVD-9	8,5 GB	1	2	Solo lectura
DVD-10	9,4GB	2	1	Solo lectura
DVD-18	17 GB	2	2	Solo lectura
DVD-5	2,6 GB	1	1	Múltiples escrituras
DVD-5	5,2 GB	2	1	Múltiples escrituras
DVD-5	3,9 GB	1	1	Una escritura
DVD-R	7,8 GB	2	1	Una escritura
DVD-RW	3 GB	1	1	Múltiples escrituras

© Luis Torres Freixinet

- Libro C: Para reproducción de audio exclusivo (DVD audio).
- Libro D: Para sistemas de escritura única (DVD-R)
- Libro E: para las unidades regrabables (DVD-RAM)

Aparte de estos formatos, existen cuatro tipos de DVD-ROM: simple cara con simple capa o doble capa, y doble cara con simple capa o doble capa.

Los DVD se basan en la misma tecnología de grabación y lectura que los CD, lo que cambia es la longitud de onda del láser, que reduce el tamaño de los agujeros y aprieta los surcos para que quepa más información en el mismo espacio. Los materiales que lo componen son algo diferentes; mientras que el CD-ROM se basa en una capa de resina polímera, con una superficie de material reflexivo (aluminio), el DVD utiliza dos capas de distinto material (una de oro y otra de plata) unidas por una tercera que es adhesiva, con lo que se garantiza que el funcionamiento sea idóneo. Este tipo de dispositivos resulta un tanto más delicado, ya que durante la unión de estas

Comparación entre el CD y el DVD

Aspectos	CD	DVD
Creados con el fin de almacenar:	Datos, audio	Datos, audio, video
Láser utilizado	780 nanómetro, luz infrarroja	635 a 650 nanómetros, haz rojo
Capacidad de almacenamiento	680 MB	Desde 3 GB hasta 17 GB
Tecnología	Una capa, una cara	Una o varias capas, con una o varias caras respectivamente
Tipo de punto utilizado	Mayores que en el DVD	Menores que en el CD
Distancia entre pistas	1.6 mm	0.74 mm
Composición	La capa es de resina polímera, con una superficie de materia reflexivo	Dos capas, una de oro y otra de plata unidas por una adhesiva
Manipulación	Con cuidado	Con mucho cuidado

© Luis Torres Freixinet

dos superficies puede introducirse cualquier partícula del aire, que puede afectar en la reproducción de su contenido. De la misma forma que el CD revolucionó los sistemas de audio hoy ya estamos convencidos de que está revolucionando los sistemas de vídeo doméstico, e incluso los profesionales.

Al principio las unidades lectoras de DVD no podían leer los CD anteriores, pero hoy en día todos los lectores incorporan un segundo láser para poder leer ambos formatos.

Conservación de los dispositivos de almacenamiento.

La conservación de la información almacenada en los dispositivos tratados es un aspecto de gran importancia. Para esto habrán de tomarse ciertas medidas. La realización de copias de seguridad es una de las formas

más sencillas de disminuir los riesgos de pérdidas de datos necesarios.

Los dispositivos magnéticos y en especial los disquetes no son dispositivos seguros para almacenar información, pues los campos magnéticos, las altas y bajas temperaturas, la humedad, los golpes, el polvo, etc. los dañan. Además existe una tendencia de las partículas magnéticas a cambiar de posición y repartirse uniformemente, según transcurre el tiempo, lo que hace que se pierda la distribución inicial de las partículas citadas y la información grabada en ellos. Los discos duros son más seguros, pero a veces fallan, pueden dañarse por oscilaciones de la tensión eléctrica, un golpe, un virus informático, o una manipulación incorrecta.

Los dispositivos ópticos son más seguros; sin embargo ellos deben cuidarse del polvo y su superficie debe protegerse para que no sufran daños, por eso generalmente tienen fundas protectoras. En este sentido, los DVD son más sensibles, sus capas protectoras más finas, por lo tanto están más expuestas a ralladuras. Como se leen con luz, su desgaste físico no es un gran problema. La permanencia de la información almacenada en ello depende de las propiedades del material que la soporta y de las condiciones de su almacenamiento.

Varias empresas aplican distintos métodos para estimar las expectativas de vida de sus propias marcas. Debido a que aún no existen estándares internacionales para estimar la durabilidad de estos materiales, sus resultados no son muy fiables; sin embargo la Prof. Cecilia Salgado en su artículo "Permanencia en CD-R"(1) asegura que los tintes de phthalocianina y cianina estabilizada con metal son bastante duraderos. Si se emplea un quemador compatible con estos tintes y se graba a una velocidad de 2x ó 4x es posible crear discos que duren más de 100 años. Esta autora además hace referencia a un estudio realizado por León-Bavi Vilmont, en el cual obtuvo como resultado que los de oro son más resistentes que los CD-R con tinte de azo y capa reflectante de plata.

Al contrario de lo que muchos piensan, la humedad relativa y la temperatura son parámetros a considerar en el almacenamiento de los soportes ópticos. Los cambios bruscos pueden causar deterioros importantes, porque los componentes de las diferentes capas que los componen tienen diferentes coeficientes térmicos de expansión. Actualmente existen normas internacionales para el almacenamiento de CD-R. Ellos indican que para asegurar su permanencia, a largo plazo, se deberán mantener a una temperatura máxima de 23°C y un 50% HR. Recientemente se ha identificado un nuevo tipo de hongo que, en condiciones climatológicas tropicales, digase 30° C y 90% HR, destruye los CD. Se trata del *geotrichum*; él se reproduce

sobre el soporte y destruye la información almacenada; primero degrada el borde externo del soporte. Esto ocurre porque el hongo se alimenta del carbono y el nitrógeno de la capa plástica de policarbonato, y así destruye las pistas de información. Este hongo crece y se reproduce con facilidad dentro de la estructura de un CD en las condiciones expuestas. Se caracteriza por formar largas cadenas de esporas viscosas e incoloras.

Otro aspecto que se debe considerar para la conservación de cualquier soporte de almacenamiento tecnoló-

gico es el respaldo de medios que aseguren su acceso. Si se analiza, la vida útil de cualquiera de ellos es mayor que la del hardware y el software que se necesita para leerlos; es fácil darse cuenta que para la conservación de estos soportes en el tiempo, existen dos alternativas:

Conservar el hardware y el software adecuado.

Establecer programas de migración de la información a hardware y software más modernos.

La tabla siguiente nos puede servir de referencia.

	Temperatura °C	Variación	Humedad Relativa	Variación
KODAK	10-25°	15°/h	20%-50%	10% en 1 h
UNESCO	20°	1°/h	40%	
IFLA	20°		40%	
ISO/DIS 18925.2	2°	10°/h	20-50%	10% e 1 h
Library of Congress	7-10°	15°/24h	45-50%	5% en 24h

(1) Salgado C. Permanencia en CD-R (discos compactos grabables). Disponible en: <http://www.lmi.com.mx/revista/conservacion/16.html>

Como condiciones de almacenamiento podríamos indicar las siguientes:

- Envases:
 - ◊ Materiales inertes con estabilidad química: poliestireno.
 - ◊ Evitar plásticos de origen celulósico: PVC y espuma.
- Depósitos:
 - ◊ Temperatura y humedad adecuada y constante.
 - ◊ Oscuridad: preservar de rayos ultra violetas.
- Limpieza:
 - ◊ Mezcla de amoníaco y agua, paño sin hilos y en sentido radial para evitar arañazos circulares.
 - ◊ Aire comprimido a presión menor de 275 kPa.
- Perduración: según fabricantes de 50 a 100 años.

La vulnerabilidad de los soportes digitales, además de lo ya comentado, podríamos indicar los siguientes:

- La escasa experiencia internacional sobre su conservación.
- Facilidad de deterioro físico por:
 - ◊ Agentes naturales:
 - Δ Físicos: humedad, temperatura, óxido, corrosión, luz...
 - Δ Bióticos: hongos, usuarios (golpes, arañazos, manipulación)
 - ◊ Agentes artificiales:
 - Δ Electromagnetismo.
 - Δ Virus informáticos.

- Obsolescencia técnica (vida media útil de 5-10 años):
 - ◊ Soportes: obsoletos incluso antes de su deterioro.
 - ◊ Dispositivos de hardware necesarios para su lectura: no hay repuestos más de 10 años.
 - ◊ Incompatibilidad de programas informáticos.

Las opciones que se nos presentan para la reproducción digital son:

1. Cambio de soporte.
2. Migración de formatos: copia del documento en un nuevo formato.
3. Eliminación de la dependencia tecnológica.
4. Emulación de la tecnología de creación.

1.- Cambio de soporte

- Recopiado periódico a soportes más estables:
 - ◊ De magnético a óptico.
 - ◊ De disco óptico a soporte óptico más moderno.
- Salida a microfilm:
 - ◊ Conserva la información pero se pierden las funcionalidades del documento digital.
 - ◊ Garantiza cien años de conservación en condiciones ambientales adecuadas
 - ◊ Util únicamente para documentos de conservación permanente y de gran valor, como copia de seguridad.

Como ejemplo podríamos citar a la mayor asociación de construcción y alquiler de Alemania es "Schwäbisch Hall". Esta empresa con más de 6,3 millones de clientes

(más de un cuarto de todos los contratos de alquiler de vivienda de Alemania) introduce en su sistema unas 60.000 páginas de texto diarias que se deben guardar durante 31 años.

Hasta hace poco, los registros se archivaban en microfilme mediante un dispositivo COM (Computer Output Microfilm – Microfilme de salida por ordenador).

Debido a que había que recuperar alrededor de 300 cartas diarias con promedios de hasta 11 horas la empresa se planteó mayores exigencias de rapidez, fiabilidad pero sin perder garantías de conservación.

Hoy en día el correo de salida de clientes se introduce en un directorio especial de un servidor FTP y los datos se convierten en imágenes TIF, que se importan al archivo Digital Data Finder basado en MSSQL 2000. Para el proceso de archivado a largo plazo, todos los datos relevantes se graban en microfilme mediante el grabador de microfilme de KODAK i9620.

La mayor parte de las recuperaciones de datos tienen lugar en los primeros 18 meses de vida del documento y se realizan en línea con un nuevo sistema denominado DMS. Después de este periodo, la recuperación emplea un escáner de microfilme que convierte las imágenes analógicas del archivo en digitales antes de dejarlas disponibles en el sistema.

La nueva instalación consiguió un ahorro drástico: aproximadamente un 60% de ahorro en costes de mantenimiento de un dispositivo COM y el sistema ya está amortizando su valor. El tiempo medio de recuperación se ha reducido de 11 horas a 85 segundos, y los costes internos se han reducido en torno a un 10%. Por tanto, el conjunto de la inversión se amortizará en menos de catorce meses.

2.- Migración de formatos

Com principio decir que los documentos deben ser accesibles para los sistemas informáticos existentes en cada momento, por lo que se deben transferir a formatos inteligibles por los sistemas actuales.

- Tipos de migración:

- ◊ Compatibilidad retroactiva: lectura y copia a nuevo formato por versiones posteriores del programa de creación.

- ◊ Interoperabilidad: lectura y copia a nuevo formato por programa diferente al de creación, que comparte los mismos formatos.

- ◊ Conversión a formatos estándar.

Es un requisito importante salvaguardar la especificación de las versiones de los formatos junto con los documentos.

Como inconvenientes de la migración encontramos

los siguientes:

- El formato final puede no ser un formato de trabajo.
- La migración pone en duda la autenticidad de los documentos cuando provoca:
 - ◊ pérdidas de datos.
 - ◊ alteraciones en la estructura interna.
- Sujeta a cambios en la estrategia comercial del productor del software.
- Técnica cara en recursos humanos (lenta) y materiales.
- Posibilidad de errores en la grabación.

La imprevisibilidad de la evolución tecnológica y su obsolescencia provoca que cada migración presente problemas específicos. De manera que nos podemos enfrentar a un problema de manera cíclica pero sin la posibilidad de aplicar las mismas soluciones en cada aplicación. Esto implica que es difícil normalizar algo que en su propia esencia nos avisa que cada vez será diferente. Podremos normalizar procedimientos, pero no técnicas, tecnologías y recursos.

En el mercado existen multitud de programas (y empresas) especializadas en la realización de migraciones (más en el ámbito anglosajón), por ejemplo:

- Conversion Plus 4.5 de Data Viz Corporation
 - ◊ 74 formatos fuente y 110 de salida.
 - ◊ Convierte cuatro categorías de ficheros:
 - Δ Texto.
 - Δ Hoja de cálculo.
 - Δ Base de datos.
 - Δ Archivos de imágenes.
- DataJunction 7.0 de Data Junction
 - ◊- 150 formatos fuente 155 de salida.
 - ◊- No convierte ficheros de imagen.

3.- Eliminación de la dependencia tecnológica

La definición o principio podría ser: Eliminación de la dependencia tecnológica no esenciales de los documentos con los programas de creación para facilitar su recuperación mediante tecnología estándar.

Tipos de dependencia evitables:

- Funcionalidades de edición: la conservación permanente exige la inmodificabilidad, por lo que sólo es necesario conservar la función de lectura y los datos contextuales.
- Vestigios de la tecnología original: por ejemplo, índices y listados de una base de datos.

Requisitos:

- Tecnología estándar.
- Normas abiertas.

4.- Emulación de la tecnología de creación

La definición o principio podría ser: Conservar la funcionalidad total o parcial de los documentos respecto al software de creación.

Como tipos más importantes tendríamos:

- Conversión en imagen electrónica mediante paso a formato PDF/A. No nos sirve cualquier PDF por varias razones:
 - ◊ Un PDF puede incluir características incompatibles con los requisitos actuales de archivo, como el cifrado o ficheros anexos.
 - ◊ Los documentos PDF no son necesariamente autointerpretables, es decir, pueden depender de las fuentes del sistema y otro contenido que resida fuera del fichero.
 - ◊ La existencia de múltiples herramientas de creación de PDF's en el mercado le da inconsistencia al formato de archivo de cara a su conservación a largo plazo.
- Desarrollo de aplicaciones emuladoras del funcionamiento del sistema operativo y la aplicación de creación.

Visores (viewers): conserva del software sólo la presentación visual o lectura.

- Ventaja: garantiza fidelidad y autenticidad documental

Inconvenientes: el emulador también está sujeto a la obsolescencia y la enorme dificultad de creación de los emuladores conforme va pasando el tiempo.

Organismo internacionales implicados en políticas de preservación:

- European Commission on Preservation and Access
www.knaw.nl/ecpa
- Council on Library and Information Resources:
www.clir.org
- National Preservation Office of United Kingdom:
www.bl.uk/services/preservation/
- Research Libraries Group: The Commission on Preservation and Access:
www.rlg.org

Mención aparte se merece el modelo OAIS por la importancia que supone en su intento de fijar los metadatos que se deben asociar a un objeto digital para garantizar su preservación y organización permanente.

El modelo OAIS o Sistema Abierto de Información

en Archivos no debe confundirse.

Su denominación puede llevar a equivoco en principio, por lo de sistema abierto de archivo, ya que parece relacionarse con la iniciativa de archivos abiertos, pero en realidad lo único en lo que coinciden ambas iniciativas es en que tratan de sistema de archivos porque el sentido de la palabra abierto tiene diferentes significados en uno y en otros. En la iniciativa de OAI se refiere al acceso libre de la información guardada, pero en el modelo OAIS se refiere al carácter no definido del propio sistema de gestión, es decir su naturaleza flexible y no cerrada para poder adoptarla a las características propias de la organización o de los propios documentos que se van a gestionar. La ISO (International Organization for Standardization), la organización que marca la pauta de como se tienen que realizar cualquier proceso en cualquier campo o área para garantizar cierta calidad, encargo al Consultative Committee for Data Space System, otro organismo internacional formado por más de 50 agencias espaciales, (CCDSS) la realización de este modelo, y así nació. Si queréis saber más información respecto a ambos (OAIS y OAI) aquí os dejo este documento para que le echéis un vistazo.

Fundamentos tecnológicos del acceso abierto: Open Archives Initiative y Open Archival Information System, de Teresa Silió.

Aclara y contextualiza conceptos como el DOI (Digital Object Identifiers), el CrossRef, usado por los editores para enlazar la citación de un artículo o revista con su DOI, y el OAI-PMH, un protocolo, componente del OAIS que sirve "para la extracción de normalizada de los metadatos que están accesibles en los repositorios que cumplen el estándar Dublin Core", que a su vez es un modelo de metadatos muy utilizado en bibliotecas. Como siempre todo se relaciona con todo.

Para los tipos de metadatos dentro de OAIS podemos ver las siguientes "categorías":

- Información sobre el contenido (CI):
 - ◊ Los datos digitalizados.
 - ◊ La descripción del documento digitalizado.
- Información descriptiva de la preservación (PDI):
 - ◊ Información de referencia: identificadores inequívocos.
 - ◊ Información sobre la procedencia: origen, custodia...
 - ◊ Información sobre el contexto documental de producción.
 - ◊ Información sobre la fijación: medidas de autenticación.
- Información sobre el empaquetamiento (PI). *Lo están desarrollando.*

- Información descriptiva (DI): se obtiene de CI y PDI. Sirve para realizar sobre ella las búsquedas. *Lo están desarrollando.*

No quisiera terminar sin mencionar los proyectos internacionales en curso: NARA-ERA, Koninklijke Bibliotheek e-Depot, KOPAL, LOCKSS, Portico, PRESTOSpace, Elsevier, Internet Archive, Library of Congress, DAITSS,.

CONCLUSIÓN

El problema tecnológico, aún sin ser lo más problemático, suele ser el más inmediato y visible. Por ejemplo, lo que más distingue la preservación de materiales digitales de los tradicionales son aspectos como:

- su dependencia del entorno informático;
- velocidad de los cambios de este entorno;
- fragilidad de los soportes;
- facilidad de modificar los datos (y por lo tanto, la dificultad de garantizar la autenticidad y la integridad de los documentos), y
- la intervención activa necesaria a lo largo de la vida del recurso.

Las principales estrategias aplicadas actualmente para frenar la pérdida de información digital son, de forma resumida:

- preservación de la tecnología,
- migración de los datos,
- emulación de las aplicaciones informáticas originales.

Pero, tal y como se afirmó en el informe, Digicult report: technological landscapes for tomorrow's cultural economy; unlocking the value of cultural heritage (Luxembourg: Office for Official Publications of the European Communities, 2002), se consideran estas estrategias como short-term solutions to a long-term problem (p. 15). Es decir, nuestros conocimientos actuales más punteros no nos garantizan la capacidad de preservar lo que estamos creando para un futuro medio y lejano.

La razón por la que nuestros esfuerzos para preservar el bitstream no son suficientes es que otros retos más allá de los tecnológicos representan obstáculos aún más complejos para superar. Se trata de problemas:

- legales: ¿cómo preservar un recurso reproduciéndolo o reformateándolo si no tenemos el permiso del titular?;

- económicos: ¿cómo mantener y actualizar la infraestructura y los recursos necesarios para garantizar la perdurabilidad de los documentos a lo largo de los años?;
- institucionales: ¿cómo asegurar el compromiso institucional permanente? ¿cómo convencer a todos los implicados –los stakeholders– de la necesidad de colaborar? ¿cómo verificar si se cumple todo lo que se promete?

Son pocas las instituciones españolas las que han afrontado plenamente al problema de la preservación a largo plazo, lo que provoca una sensación de “peligro inminente” difícil de sobrellevar.

Lo que está claro es el fundamento común para estos depósitos: el protocolo Open Archival Information System (OAIS), el cual está completamente asumido e implantado por las instituciones que más han avanzado en el terreno de la preservación digital como, por ejemplo, las bibliotecas nacionales de Australia, Gran Bretaña, Holanda y Suecia. El modelo OAIS fue elegido por los consorcios RLG y OCLC como su propuesta sobre depósitos digitales ‘fiables’ (Trusted digital repositories: attributes and responsibilities; an RLG-OCLC report. Mountain View, CA: RLG, 2002). Pero por muy valioso que sea este documento, las pautas que ofrece no son hoy en día suficientemente específicas para la puesta en marcha de depósitos. Varios participantes reclamaron una hoja de ruta para facilitar el inicio del proceso.

Deberemos establecer una necesidad de establecer un sistema de certificación para los depósitos digitales. Básicamente se tendría que validar la capacidad de la institución –y de su personal– para llevar a cabo lo que pretende hacer. Ejemplos de aspectos que requerirán una certificación son:

- responsabilidad administrativa
- viabilidad organizacional
- sostenibilidad financiera
- adecuación tecnológica
- seguridad del sistema
- adecuación de los procedimientos

En definitiva la preservación de los recursos digitales es un reto al que tarde o temprano tendrá que enfrentarse toda institución –biblioteca, archivo, museo– que pretenda preservar para futuros usuarios lo que se genera hoy en formato digital. Comenzar una política de digitalización sin plantearse de manera inicial las cuestiones de la preservación digital lo único que puede producir son problemas a largo plazo, que no por evitarlos durante un tiempo significarán que dejan de existir.

ENLACES ORDENADOS SEGÚN APARECEN EN EL TEXTO

FORMATOS

djvu	http://djvu.org/resources/whatisdjvu.php
Jpg2/.jp2	http://www.dpconline.org/docs/reports/dpctw08-01.pdf
OGC	http://www.opengeospatial.org/
W3C	http://www.w3c.es/
GDFR	http://hul.harvard.edu/gdfr/
OCLC	http://www.oclc.org/global/default.htm
JHOVE	http://hul.harvard.edu/jhove/index.html
OPENRAW	http://www.openraw.org/
Margaret Lepley	http://spiedl.aip.org/vsearch/servlet/VerityServlet?KEY=SPIEDL&possible1=Lepley%2C+Margaret&possible1zone=author&maxdisp=25&smode=strresults&aqs=true
Michael Reichmann and Juergen Specht "The RAW Flaw"	http://www.luminous-landscape.com/TheRawFlaw.pdf
foro	http://www.luminous-landscape.com/tutorials/expose-right.shtml
Florida Digital Archive	http://www.fcla.edu/digitalArchive/pdfs/recFormats.pdf
master de conservación	http://www.dpconline.org/docs/reports/dpctw08-01.pdf
Library of Congress	http://lcweb2.loc.gov/ammem/gmdhtml/gmdhome.html
Instituto Cartográfico de Cataluña	http://vacani.icc.cat/index.php
AIIM	www.aiim.org/standards
DJUV vs JPEG	http://djvu.org/resources/jpgvsdjvu/comparison.html

PRESERVACIÓN

Salgado C. Permanencia en CD-R	http://www.lmi.com.mx/revista/conservacion/16.html
European Commission on Preservation and Access	http://www.knaw.nl/ecpa/
Council on Library and Information Resources	http://www.clir.org/
National Preservation Office of United Kingdom	http://www.bl.uk/npo
Research Libraries Group: The Commission on Preservation and Access:	http://www.oclc.org/community/rlg/transitions/
Fundamentos tecnológicos del acceso abierto: Open Archives Initiative y Open Archival Information System, de Teresa Silió	http://eprints.rclis.org/archive/00005887/01/EPITSilio.pdf
Alice Keefer	http://www.thinkepi.net/15
IweTel	http://www.rediris.es/list/info/iwetel.es.html
<i>Digicult report: technological landscapes for tomorrow's cultural economy; unlocking the value of cultural heritage</i>	http://digicult.salzburgresearch.at/
Dspace	http://www.dspace.org/
CEDARS project	http://www.leeds.ac.uk/cedars/
DELOS project	http://www.delos.info/
Digital Curation Centre	http://www.dcc.ac.uk/index
Digital Preservation Coalition	http://www.dpconline.org/graphics/index.html
Directrices para la preservación del patrimonio digital. Paris: UNESCO, 2003.	http://unesdoc.unesco.org/images/0013/001300/130071s.pdf
ERANET project	http://www.eranet.org/
Implementing preservation repositories for digital materials: current practice and emerging trends in the cultural heritage community. Dublin Ohio : OCLC, 2004.	http://www.oclc.org/research/projects/pmwg/surveyreport.pdf
National Digital Information Infrastructure and Preservation Program	http://www.digitalpreservation.gov/
National Library of Australia. Preserving access to digital information (PADI)	http://www.nla.gov.au/padi/index.html
Preservation Metadata: Implementation Strategies Working Group (PREMIS)	http://www.oclc.org/research/projects/pmwg/
Preserving Cornell's Digital Image Collections: Implementing an Archival Strategy	http://www.library.cornell.edu/preservation/IMLS/