

LA DIGITALIZACIÓN EN ARCHIVOS

Vicent Giménez Chornet

vigicho@har.upv.es

Universidad Politécnica de Valencia

Luisa Tolosa Robledo

ltolosa@upvnet.upv.es

Universidad Politécnica de Valencia

1.- Por qué digitalizar en archivos

La necesidad de realizar copias de la documentación conservada en los archivos, dado que ésta es única, ha estado presente a lo largo de la historia. Los motivos que han generado esta necesidad han sido diversos según las distintas épocas, así como las soluciones aportadas y los soportes utilizados para obtener las copias..

Antiguamente las copias se realizaban generalmente para que una persona, física o jurídica, dispusiese de una información que le permitiría resolver un asunto o una necesidad. El propio organismo realizaba la copia, bien registrándola en libros (ejemplo de la cancillería vaticana o la real cancillería de la Corona de Aragón), bien emitiendo certificados, o bien realizando copias (traslados) firmadas por notarios en el ejercicio de sus competencias como garantes de la fe pública. Estas copias tenían una finalidad eminentemente administrativa o del ejercicio del derecho privado de las personas. El procedimiento utilizado era la escritura manuscrita, o sea, volver a escribir el documento y el soporte predominante era el pergamino o el papel.

Con la revolución tecnológica del siglo XIX y XX aparece un formato de soporte que va a incidir decisivamente en diversas actividades relacionadas con el registro de la información. Nos referimos a la película. Sin entrar en detalle sobre los soportes anteriores de placas de vidrio, la película permitió registrar una captura de imagen para disponer en un futuro de dicha imagen o información. Esta novedad tecnológica, que tanto ha influido en la fotografía y en el cine, se demostró útil también en el campo de la gestión documental llevada a cabo por las organizaciones. La película, en su modalidad de microfilm, va a ser el recurso elegido en el siglo XX para hacer copias de la documentación original, porque, a diferencia de épocas anteriores, la finalidad de la copia ha cambiado, al igual que su soporte y, precisamente, facilitada por éste.

El microfilm permite realizar una copia del documento pero no permite atribuirle los atributos de autenticidad. Aún así algunos organismos de la administración pública, en la época de expansión del uso del microfilm (años 60 y 70), decidieron sacar copias en microfilm de ciertos documentos que habían generado y, dado el volumen que ocupaban estos, destruirlos. En estos casos la copia de microfilm era el único recurso de donde extraer la información que una vez contuvo el documento original. Sin entrar a valorar esta práctica, sin duda alguna fruto de la euforia por ganar espacio físico en el organismo con la utilización del nuevo soporte, la práctica más generalizada en el uso del microfilm se ha aplicado a crear copias de la documentación original para satisfacer dos necesidades: disponer de una copia de seguridad para proteger el original y entregar copias del documento a los usuarios interesados. Como hemos mencionado, en ninguno de los dos casos se plantea la autenticidad de la copia.

Hasta hace pocos años el único formato adoptado en los archivos para realizar la copia de seguridad ha sido el microfilm. Se han valorado hasta la saciedad las virtudes de este formato, pero, en defensa de lo conocido, se han planteado poco sus deficiencias. El nuevo formato competidor, la imagen digital, invita a replantearse la política de creación de copias de seguridad. Como es habitual, las nuevas tecnologías inciden en los métodos de trabajo.

Actualmente al microfilm se le reconocen ciertas deficiencias:

- a) Un coste elevado: que contempla tanto la maquinaria necesaria para la realización de las copias (cámaras, objetivos) y la lectura y reproducción de las mismas (lectores de microfilm, lentes) y su mantenimiento, la adquisición de la película del microfilm y su proceso de revelado, como los gastos de mantenimiento del propio microfilm (condiciones medioambientales específicas que no se dan en el entorno habitual).
- b) Una dificultad en la recuperación de la información: para localizar en un rollo de microfilm el documento deseado se debe buscar o “pasar” dicho rollo por el lector del microfilm. Tarea ardua muy rechazada por los investigadores de los archivos que ven más eficaz y cómoda la consulta del original.
- c) El microfilm se muere: mirando hacia el futuro la industria de la película o celuloide tiende a desaparecer. La industria del cine, primera y gran consumidora del celuloide cambia de soporte, optando por el formato digital. Los ciudadanos abandonan las cámaras analógicas por las digitales. ¿Qué industria va a hacerse cargo de fabricar celuloide para los archivos?
- d) La visualización del microfilm peligrará: la industria encargada de fabricar los lectores de microfilm está en retroceso. Los lectores de microfilm se

encarecen por la escasa demanda y, evidentemente, llegará el momento en que dejen de fabricarlos.

Mirando al futuro, la apuesta por la política de microfilmación es un derroche de dinero: se microfilmará documentación para después digitalizar lo microfilmado. Y ¿qué hacemos con el archivo de microfilm?. El microfilm se deteriora, incluso en las mejores condiciones medioambientales. Pensemos en la industria del cine, principal interesada en conservar las copias master, como se han visto obligados a migrar hacia la copia digital como formato master y de difusión.

La apuesta por la digitalización redundará en varias ventajas. La primera, la industria se ha volcado en el formato digital, hecho que proporciona garantías de supervivencia, mantenimiento tecnológico y bajada de precios. La segunda, el formato digital permite una mayor rapidez en la recuperación de la copia digital deseada, todo depende de los metadatos que enlacemos a dicha imagen. Tercero, la copia digital puede ser firmada digitalmente, autenticando dicha copia. Cuarto, el mantenimiento de los soportes ópticos requiere menos requisitos medioambientales diferentes al de nuestro entorno. Y quinto, la migración a otros soportes o cambios de formato es mucho más rápida que en el microfilm.

El miedo al futuro y a la incorporación de las Tecnologías de la Información y la Comunicación puede provocar una pérdida del trabajo realizado y una pérdida de los recursos económicos invertidos.

2.- El proceso de digitalización

Las características físicas de la documentación original archivística tienen que quedar salvaguardadas en el proceso de la captura digital; su manipulación no debe perjudicar su estado de conservación. El equipo de digitalización tiene

que ser mucho más riguroso y garantizar que no habrá deterioro, o que el deterioro será ínfimo en la pieza documental ya que, a diferencia de las bibliotecas, estamos ante documentos únicos e irremplazables.

Sin olvidar que la finalidad es hacer una copia del documento original para preservar el original y para difundir la información de éste sin su utilización, en el proceso de digitalización debemos prestar especial atención a las siguientes cuestiones¹:

- a) Todo el proceso debe estar supervisado por un técnico en imagen digital, que pueda evaluar tanto la calidad de la imagen resultante como el software utilizado para gestionar el color o incluso el calibrado de los monitores. Unos monitores mal calibrados no permiten determinar la calidad de la captura de la imagen al ofrecer una deficiente visualización.
- b) La preservación de la pieza original, especialmente en los casos de libros con un lomo ancho, requiere que la mesa portalibros disponga de un balancín que permita abrir el libro entre 180° y 120°, sin necesidad de utilizar una prensa o cristal que presione sobre la pieza para extenderla y evitar la curvatura, como generalmente ocurre en la digitalización de libros impresos. En este sentido para garantizar la calidad de la imagen capturada es recomendable que el equipo de digitalización disponga de un software que pueda corregir la curvatura provocada por el libro abierto.
- c) El equipo de digitalización debe reunir unos requisitos técnicos que eviten daños al documento: la iluminación no debe provocar una subida de la

¹ Algunas de estas cuestiones las hemos tratado en Campuzano Juan, Francisco, Giménez Chornet, Vicent, "Recomendaciones técnicas para la captura y digitalización de fondos de archivo". En línea: <<http://www.cult.gva.es/DAIT/Archivos/images-reino/Recdigitalizacion.pdf>> [Consulta: 27-06-2007]

temperatura ambiente superior a 0,5° y no se deben utilizar balancines hidráulicos cuya rotura podría mojar las piezas.

- d) El personal que manipula las piezas documentales ha de ser extremadamente respetuoso con las mismas. Las prisas en la digitalización, como consecuencia de los contratos realizados con empresas externas que buscan una alta rentabilidad económica, puede ocasionar que el personal digitalizador realice una inadecuada manipulación de aquello que queremos preservar, de manera que la pieza pasaría de la digitalizadora al laboratorio de restauración.

3.- Los formatos de archivo digital.

Cuando realizamos la captura de la imagen, bien sea con un escáner o con una cámara, el software que lleve incorporado lo almacenará en un formato de archivo para imágenes digitales. Estos equipos suelen utilizar una reducida gama de formatos (principalmente TIFF, JPEG, y RAW) que posteriormente, dependiendo de nuestro interés o del programa archivístico que hayamos establecido en nuestro centro, transformaremos mediante aplicaciones informáticas al formato o a los formatos que tengamos determinados tanto para la copia master como para la difusión.

A grandes rasgos se pueden agrupar en dos clases los formatos de archivos de imagen digital: los que almacenan toda la información sin pérdida de calidad y los que descartan información en el momento de almacenarlas. Los formatos que almacenan toda la información son los idóneos para la copia master, a partir de la cual podremos transformar en nuevos formatos de archivo con pérdida de información para cubrir otras necesidades, como difusión en una intranet, difusión por Internet o imagen digital disponible para usuarios o clientes.

Entre los formatos que permiten almacenar la captura sin pérdida de información destacamos los RAW, BMP y TIFF.

El formato RAW contiene toda la información que recibe del sensor de captura, es decir, todos los píxeles en bruto que ha adquirido el escáner o la cámara, y se requieren programas específicos para poder procesarla. Está considerado como el “negativo” de la fotografía digital. Las ventajas de este formato para la copia digital de la documentación de archivos estriba en que permite almacenar toda la información de la captura, muy recomendable para la copia master. Entre los inconvenientes destacan, por una parte, el gran volumen de memoria que ocupan (este inconveniente es perfectamente asumible si lo utilizamos para las copias master) y, por otra parte, dos obstáculos más problemáticos, uno que los formatos RAW no están estandarizados, de forma que cada fabricante tiene su propio formato, cuyo código mantiene oculto, y nos vemos cautivos de dicho fabricante, y otro que estos formatos RAW no permiten añadir datos descriptivos en el archivo de la imagen.

El formato BMP (Bit Mapped Picture) es el formato utilizado por Windows y DOS para almacenar las capturas. Se le conoce como formato de mapa de bits porque existe una matriz de píxeles donde se asigna el código de color específico. Soporta hasta 32 bits de color. Entre las ventajas de este formato destaca la gran cantidad de información que almacena y, por lo tanto, también es recomendable para la copia master. Entre los inconvenientes que presenta, además del gran volumen de memoria que requiere, asumible igualmente cuando nos referimos a la copia master, es que no admite imágenes de gran calidad dado que sólo llega a 32 bits de color y tampoco admite datos descriptivos de la imagen.

El formato TIFF (Tagged Image File Format) es uno de los primeros formatos digitales existente y fue desarrollado en los años 80 por Aldus, una compañía que ha sido comprada por Adobe Systems, actual propietaria de las especificaciones. Por su difundido uso en diferentes plataformas (Windows, Mac, Linux, UNIX, etc.) y por ser un formato reconocido por muchísimos programas de edición de imágenes se ha convertido, de hecho, en un estándar para aquellos que quieran almacenar imágenes con gran cantidad de información y, además es, de facto, un estándar reconocido oficialmente por la norma de certificación ISO 12.639. A la ventaja de ser un formato de archivo de alta calidad se le une la característica de que soporta hasta 48 bits de color y la posibilidad de guardar capas y canales alfa. En contrapartida, el volumen de memoria que ocupa el archivo digital es elevado, solamente asumible para copias master². Por otra parte, dispone de una característica que lo hace idóneo para la digitalización en los archivos: admite metadatos de descripción de la imagen insertados en el archivo, tanto EXIF como IPTC³.

Entre los formatos que descartan información a la hora de almacenar la captura o formatos obtenidos a partir de un archivo sin comprimir, destacamos los GIF, PNG, JPEG, JPG2.

El formato GIF (Graphics Interchange Format) se desarrolló en 1987 por CompuServe para la versión 87a, con unas modificaciones en la versión 89a (entre otras permite la imagen animada). Es un formato de archivo que ocupa muy poco volumen, por lo que es útil para la utilización en Internet. Entre los

² Algunos especialistas de la imagen digital también opinan que el formato TIFF es idóneo para la copia master, por ejemplo Fernando de la Ossa: "Tecnología para la digitalización de documentos de archivos", TRIA, nº 8 y 9º (Córdoba, 2001-2002) , p. 11-79.

³ Para conocer EXIF e IPTC, consultar Giménez Chornet "Datos Descriptivos Añadidos a la Imagen Digital. Características EXIF e IPTC" en línea: <http://www.cult.gva.es/DAIT/Archivos/images-reino/metadatos.pdf> [Consulta 24-07-2007].

inconvenientes que presenta, destacamos que tiene muy poca profundidad de color, la imagen puede tener un total de 8 bits por píxel, por lo que los colores quedan reducidos a 256 colores. En este sentido la calidad de la imagen es muy baja. Igualmente es un formato que no permite la incorporación de metadatos descriptivos de la imagen, ni de tipo EXIF ni IPTC.

El formato PNG (Portable Network Graphics), supera a GIF en calidad ya que la profundidad de color alcanza los 24 bits por píxel, considerado como “color verdadero” dado que es el número de colores que alcanza a percibir el ojo humano. La versión 1.0 apareció en 1996, y en 1999 aparece la versión 1.2, aceptada como estándar internacional (ISO 15948:2003) y recomendado por la W3C. Este formato tiene entre sus ventajas que permite una compresión sin pérdida de calidad (es decir, sin reducir el número de colores). Por todo ello es uno de los formatos idóneo para la copia digital de los documentos de archivo que vayamos a destinar a la difusión, bien para entregar a los usuarios o para enviar a la red. El único obstáculo es que no permite la inserción de metadatos descriptivos de la imagen, ni EXIF ni IPTC.

El formato JPEG ha sido desarrollado por la Join Photographic Experts Group, es un formato de fichero que utiliza algoritmos para comprimir la imagen con pérdida de información. Podemos ajustar el grado de compresión, pero al descomprimir la imagen no obtendremos exactamente la misma imagen de antes. El grado de compresión respecto al original puede ser muy elevado manteniendo una buena calidad de la imagen desde la percepción del ojo humano. Es un formato idóneo para la imagen digital que se destinará a la difusión. Por la pérdida de información en la compresión no es recomendable su utilización para la copia master. Utiliza una profundidad de color de 24 bits por píxel, al igual que el PNG.

El sistema de compresión que utiliza el JPEG fue aprobado como estándar en 1994 con la ISO 10918-1. Otra de las ventajas del formato PNG es que admite metadatos descriptivos de la imagen, tanto EXIF como IPTC. Por todo ello es uno de los formatos más recomendables para la digitalización en los archivos, siempre que sea la copia destinada a la difusión.

El formato JPEG2000 ha sido creado igualmente por la Joint Photographic Experts Group, mejorando el algoritmo de compresión (compresión de imágenes basada en transformación de ondas)⁴. Entre sus ventajas destacar, pues, el alto grado de compresión y entre las desventajas encontramos que pocos navegadores web lo reconocen y que no permite la inserción de metadatos descriptivos en el archivo de la imagen.

El formato DjVu⁵ es uno de los más recientes. Es de código abierto aunque la propiedad de los derechos para el desarrollo comercial del software de codificación lo poseen diversas compañías, especialmente LizardTech⁶, que facilita gratuitamente un visualizador de este formato. Este formato realiza una separación de capas de las imágenes, las cuales comprime sin pérdida en el caso de las bitonales (dos colores), lo que permite que el volumen de memoria que ocupa es muy bajo. Ésta es una de las ventajas a tener en cuenta para el caso que nos planteemos utilizar este formato si queremos realizar la copia de documentos para la difusión. La desventaja la encontramos especialmente en el caso de digitalizar documentos de archivo. Este formato crea capas bitonales, por tanto si en una capa de la captura de la letra manuscrita ésta no mantiene una uniformidad en el grado del color de la tinta la información de dicha capa será

⁴ Web oficial de JPEG 2000: <<http://www.jpeg.org/jpeg2000>> [Consulta: 29-06-2007]

⁵ Diversos recursos de este formato se encuentran en: <<http://djvu.org>> y <<http://djvu.sourceforge.net>> [Consulta: 30-06-2007].

⁶ Esta compañía menciona la implantación de DjVu en algunas organizaciones, tanto en Chile como en España. En línea: <<http://www.lizardtech.es>> [Consulta: 30-06-2007]

incompleta, ya que no incorpora los trazos de la letra de una intensidad diferente. Este formato, por el contrario, es eficaz en las letras impresas que mantienen un mismo grado de intensidad de color. Otro de los inconvenientes es que no permite la inserción de metadatos descriptivos de la imagen en el archivo del formato.

4.- Los formatos utilizados en los centros archivísticos.

Desconocemos los formatos master de imagen digital que las instituciones archivísticas han incorporado a su política o plan de digitalización. A través de la difusión de la copia digital de baja resolución que han puesto a disposición de los usuarios en Internet, vemos que predominan dos tipos de formatos, el JPEG y el DjVu. Por la copia difundida en JPEG se intuye que, en algunos casos, no existe copia master de alta resolución (TIFF o RAW) y que por tanto la política de digitalización de la mayoría de los centros archivísticos se han regido por los siguientes criterios:

- a) La captura digital del documento de archivo se almacena directamente en un formato de baja resolución o formato de compresión, generalmente el JPEG, rechazando almacenar toda la información de la captura en un formato que no discrimine información.
- b) No realizar la captura digital del documento original del archivo y realizar la captura de una copia del documento original, es decir, del microfilm, para crear una imagen en JPEG para la difusión.

Partiendo de la premisa de que no hay una política única, ideal y perfecta, que nos permita establecer los criterios para implantar en los archivos una opción por lo que respecta a la copia digital master, valoremos los pros y los contras de los criterios mencionados.

Las ventajas de no almacenar en un formato sin compresión como el TIFF o el RAW radican en dos puntos. Uno, que es bastante evidente, el ahorro económico que puede costar el almacenamiento de estos archivos digitales, dado que el volumen de megas que ocupa dicho archivo es considerable, tanto si se almacenan en DVDs o en un disco duro magnético. Y otro, menos evidente, el tiempo que se invierte en guardar o almacenar formatos sin compresión, por lo que el proceso de digitalización se ralentiza, tanto en dicho proceso de guardar la captura digital como en el uso o manejo de estos archivos.

Por otra parte, la ventaja de digitalizar microfilmes que el propio organismo ya tiene realizados, porque ya tenía una política de preservación de los documentos originales, radica en la rentabilidad económica, ya que en poco tiempo y a partir de un proceso automatizado, se pueden volcar imágenes de microfilm a imágenes digitales, las cuales pueden ser en cualquier formato de los mencionados, pero teniendo presente que es una copia de la copia. Por esta vía, en poco tiempo podemos disponer de un gran volumen de imágenes en formato JPEG, u otro que contemple la compresión, para poner los documentos a disposición de los usuarios.

En contra, la ausencia de una copia master en un formato de archivo que conserve toda la información de la captura, TIFF, RAW o BMP, implica no contemplar los siguientes aspectos en nuestro proyecto de digitalización:

- a) Si disponemos de una copia digital sin comprimir, dada la calidad de la misma, dicha copia se puede depositar en un centro diferente a fin de llevar una política de seguridad en un sistema de prevención de desastres. Dicha copia garantizaría que en otro centro dispusiéramos de una copia de nuestra documentación archivística con todas las garantías

de calidad. Si no disponemos de dicha copia master no podemos implantar un buen sistema de prevención de desastres.

- b) La disposición de una copia sin comprimir nos permitirá en un futuro realizar copias de compresión a otros formatos nuevos que, tal vez, con menos peso puedan contener mayor volumen de información y, por tanto, de calidad. Es, por tanto, una apuesta de futuro, para no tener que volver a digitalizar la documentación original dado que desde el inicio del proyecto ya optamos por un formato sin comprimir. La opción de no aplicar este criterio implicará volver a realizar el trabajo de la captura digital del documento original.
- c) En algunas ocasiones los centros archivísticos necesitan una copia digital de alta resolución (tanto en dpi o en ppp como en formato sin compresión) para resolver diferentes necesidades: para usuarios que necesitan una copia que van a entregar a un impresor en la realización de un libro, para la realización de impresiones de gran tamaño (posters en exposiciones, carteles, etc.), para entregar a algún realizador de un audiovisual que necesite tratar la imagen, etc. Por tanto, la ausencia de la copia master implicaría volver a realizar el proceso de digitalización de la pieza archivística.

Los formatos generalmente utilizados por los centros archivísticos para difundir el patrimonio documental por Internet son el JPEG y el DjVu. Como hemos mencionado el JPEG es un formato que permite un elevado grado de compresión junto a una buena profundidad de color. Las experiencias más antiguas en digitalización de documentación archivística, como el caso del Archivo General de Indias, optaron por el formato JPEG en escala de grises. Si tenemos en cuenta

que la apuesta por este formato tiene su origen en un proyecto de finales de los 80, con un gran apoyo económico del Ministerio de Cultura y de la IBM, es evidente que fue una apuesta de futuro. Para aquella época el volumen que ocupaban dichas imágenes suponían un gran costo económico en almacenamiento y en proceso técnico, sin saber que en un futuro próximo la aparición de Internet iba a revolucionar el sistema de comunicación. Dichas imágenes, en el momento en que el Ministerio de Cultura las sacó por Internet, a principios del siglo XXI (en el portal Archivos Españoles en Red, AER) conservan, en escala de grises, una buena calidad para satisfacer las necesidades del usuario: leer el contenido de los documentos. Otra cuestión diferente es la digitalización del microfilm puesta en Internet, que actualmente podemos contemplar en este portal español, dada la baja calidad de la copia de la copia⁷.

Actualmente, en relación con el coste económico y de tiempo que suponía la captura y almacenamiento en JPEG, el precio que cuesta el proceso de digitalización ha caído en picado. La memoria necesaria para el almacenamiento, tanto de disco magnético como óptico, es muy inferior en relación con las dos décadas anteriores. Y el tiempo que cuesta procesar la digitalización, desde la captura hasta el almacenamiento, también ha disminuido. Por estos motivos, actualmente no tiene mucha justificación la digitalización en escala de grises. Esta opción era también la más elegida por las características de la documentación manuscrita, en cuanto que existe una exigua gama de colores, el de la tinta y el del papel, pero actualmente se pueden conseguir formatos JPEG a color que pueden competir en peso con los formatos en escala de grises, y las ventajas del formato a color son mayores. Radican en que los usuarios ven la copia digital

⁷ Recientemente AER a migrado a PARES (Portal de Archivos Españoles) <<http://pares.mcu.es/>> [Consulta: 2-07-2007]

como si estuviesen viendo el original, se aprecian mejor los detalles del cambio de tonalidad en la tinta y agota menos el proceso de lectura de muchos documentos. Por este motivo algunos Archivos ya apuestan por la imagen a color para la difusión en Internet.

El formato DjVu ha sido adoptado por algunos Archivos en los proyectos de digitalización y difusión por Internet. En España tenemos como ejemplo representativo el Archivo del Ayuntamiento de Zaragoza. Como hemos mencionado es un formato que pesa muy poco, idóneo para transferirlo a gran velocidad por Internet, pero que, para los documentos manuscritos, donde en el trazo de la letra puede variar la intensidad de color de la tinta, las capas que realiza el formato DjVu pueden eliminar algunas partes del trazado de la letra, dificultando la lectura de los documentos⁸.

5.- Los metadatos en las imágenes digitales.

En las bases de datos que se elaboran en los centros archivísticos, con la finalidad de automatizar las descripciones de los fondos documentales, se puede optar por utilizar campos normalizados, bien a partir de la norma ISAD (G) o de la norma EAD, o se puede optar por utilizar campos designados libremente atendiendo, generalmente, a criterios de la tradición archivística anterior a la normalización. No cabe duda que la normalización en los campos reporta muchos beneficios, sobre todo a la hora de implantar un sistema archivístico.

Los registros de las descripciones archivísticas en las bases de datos sirven, fundamentalmente, para la recuperación de la información, es decir, para

⁸ Se puede consultar la serie Diplomática del Archivo del Ayuntamiento de Zaragoza <http://cmisapp.ayto-zaragoza.es/ciudad/usic/archivo/consultaimagenes_Fondo?tema=SERIE%20DIPLOMATICA> [Consulta: 2-07-2007]

poder interrogar en dicha base de datos qué fondos documentales, series o documentos hay sobre la cuestión que el interesado busca.

En algunos casos, el mismo registro de la descripción archivística puede proporcionar un enlace a la imagen o imágenes que son la copia digital del documento original. En estos casos podemos decir que las imágenes digitales tienen unas descripciones enlazadas por el registro de la descripción de la unidad documental de la base de datos. En este caso los datos o “metadatos” asociados a la imagen digital están integrados en la arquitectura de la base de datos. Como ocurre en todos los portales web de los centros archivísticos o de ciertos sistemas de archivos, cuando las descripciones y las imágenes digitales están enlazadas el sistema de recuperación o de búsqueda de los documentos digitalizados pasa por interrogar en los campos del registro de la descripción archivística, bien por los campos normalizados o por campos de descriptores que, generalmente, utilizan un lenguaje documental. Una vez localizado el registro que describe la unidad documental y que permite el acceso a la imagen digital enlazada, el usuario puede, desde su lugar de consulta (domicilio particular, centro de trabajo, sala de consulta del archivo, etc.), descargarse dicha imagen digital, aunque en algunos casos solamente puede consultarlas en el portal⁹. En cualquiera de estos dos casos los datos o “metadatos” enlazados con la imagen digital son completamente eficaces para resolver el objetivo de localizar dicha imagen.

En ninguno de los portales que conocemos, donde se ofrecen a los usuarios imágenes digitales, dichas imágenes contienen insertadas en el formato

⁹ En los casos mencionados de PARES o del Archivo del Ayuntamiento de Zaragoza se pueden descargar libremente. En otros casos se ofrece a la consulta pero no se pueden descargar, como en el Archivo de la Universidad de Santiago de Compostela:
<<http://www.usc.es/archivo/fondos/acceso.htm>> [Consulta: 02-07-2007]

de archivo digital los metadatos relacionados con la descripción de la imagen o sobre el origen del documento que se ha digitalizado.

En el Archivo del Reino de Valencia hemos insertado metadatos descriptivos en el propio formato de archivo digital¹⁰. Tras un análisis de las posibilidades existentes para la inserción de metadatos en el mismo archivo digital se optó por utilizar los campos del estándar EXIF aplicable a los formatos TIFF y JPEG. Tal decisión se tomó antes de la nueva versión de Photoshop CS que, entre sus nuevas funcionalidades, ha incorporado la posibilidad de editar los campos EXIF. Las pruebas realizadas con los metadatos insertados en ciertos campos del estándar EXIF revelaron un excelente grado de eficacia en:

- a) Permanencia de dichos metadatos en las conversiones de formato de TIFF a JPEG.
- b) Permanencia de los metadatos en las migraciones de los archivos digitales de un soporte a otro.
- c) Permanencia de los metadatos en las descargas de la web o en la utilización del correo electrónico.
- d) Recuperación de las imágenes digitales a partir de ciertas aplicaciones comerciales o de ciertas aplicaciones de software gratuito.
- e) Recuperación de las imágenes desde el mismo sistema operativo Windows.

Si bien la mayoría las etiquetas EXIF se destinan a almacenar las características técnicas de la captura digital, unas cuantas etiquetas permiten al autor de la imagen incorporar metadatos relacionados con la descripción de la

¹⁰ Esta experiencia la hemos expuesto en: Giménez Chornet, Vicent; Sellés Carot, Alicia; Roque Izquierdo, Graciela: "Recuperando imágenes digitales a partir de los metadatos Exif: caso del Archivo del Reino de Valencia", *2ª Jornada sobre Gestión de Documentos Digitales*. Madrid: aedocdigital, 2006. Giménez Chornet, Vicent; Sellés Carot, Alicia; Roque Izquierdo, Graciela; Puchades Asensi, Yolanda; Monleón Escribano, Daniel: "Recuperación de información descriptiva en imágenes digitales desde metadatos Exif. Su utilidad para el Archivo del Reino de Valencia.", *10as Jornadas Españolas de Documentación, FESABID 2007*.

imagen y con los derechos de autor de dicha imagen. Es el caso de las etiquetas *Descripción de la imagen, Artista, Copyright y Comentario del usuario*. Tras el análisis de dichos campos, los que mejores resultados han dado de permanencia y posibilidad de recuperación desde cualquier aplicación han sido los campos *Copyright y Descripción de la imagen*. Por ello se determinó incorporar en estos dos campos la información de los campos más relevantes de la ISAD (G).

Correlación de campos ISAD (G) en etiquetas EXIF	
EXIF	ISAD (G)
Copyright	Código de referencia
Descripción de la imagen	Fechas; Alcance y contenido Descriptores: geográfico, onomástico, materia, y instituciones y entidades

Con ello se consigue que el archivo de la imagen digital contenga la información relacionada con la institución archivística que lo custodia, en este caso el Archivo del Reino de Valencia, la signatura del documento original, las fechas, la descripción de dicho documento y el vocabulario índice o descriptores.

La pregunta que se suele plantear a continuación es ¿para qué este proceso de incorporación de la información en los metadatos de la imagen si ya están en la base de datos y con ellos se puede localizar la imagen?. Efectivamente la finalidad no parte de este planteamiento, no se incorporan metadatos a la imagen porque ya está resuelto en la base de datos. La inserción de metadatos en la imagen digital tiene como finalidad otras funciones que, antes de la tecnología digital, eran más difíciles o imposibles de resolver. Destacamos tres finalidades:

- a) **Descripción de las copias de seguridad.** Tradicionalmente en las instituciones archivísticas las copias de seguridad de la documentación se han realizado a partir de microfilmes o microfichas. En todos estos casos la referencia del rollo de película o de la microficha se anotaba en el contenedor de los mismos, cajas o sobres, consistente en una breve referencia a la signatura de la documentación original. Ahora, si optamos por copias de seguridad en formato TIFF, como formato master, podemos disponer de la información descriptiva de la documentación original. Con ello, todas aquellas copias que saquemos con este formato para depositar en otro centro, como consecuencia de la aplicación de medidas de seguridad, guardaran la información relacionada con la signatura y con la descripción del contenido. En estos casos en que la imagen no está incorporada en una base de datos, dicha información es potencialmente muy útil para su localización.
- b) **Descripción de las copias entregadas al usuario.** Como en el caso anterior, es común que cuando los usuarios de los centros archivísticos solicitan copias de los documentos originales se les suministre dicha copia en soporte microfilm o fotocopia en papel. La referencia que se le da al usuario, en el mejor de los casos, es la de la signatura del documento original. Dicha referencia el usuario la puede conservar o la puede perder, pero, en cualquier caso, no dispone de la referencia de la descripción de la imagen. Con la imagen digital podemos entregar a los usuarios la misma descripción de la imagen y la referencia de la signatura, todo ello mejorado con la incorporación de descriptores. Esta finalidad es tan útil para los usuarios presenciales que se acercan al centro como para los usuarios

remotos que consultan la documentación por Internet. En todos estos casos el usuario se descarga la copia digital en el ordenador y puede disponer de la identificación y descripción de la imagen mientras disponga de dicha imagen, y puede realizar búsquedas (recuperación de la información) con resultados eficaces, a pesar de que las imágenes estén fuera de la base de datos.

- c) **Externalización de las copias.** Cuando los centros archivísticos no disponen de suficiente personal, la realización de copias, generalmente en formato microfilm para disponer de copias de seguridad, se encarga a empresas externas y éstas entregan la relación de microfilmes con un dossier donde consta la referencia de la caja o rollo y la referencia de la signatura del documento original. Ahora, con la digitalización, la empresa externa puede entregar los archivos digitales con una referencia insertada de la signatura de la documentación original. Ello nos permitirá no perder nunca dicha signatura y, además, buscar/recuperar rápidamente las imágenes almacenadas en un ordenador, sea cliente o servidor, utilizando el sistema operativo Windows o cualquier aplicación que abra las etiquetas EXIF, algo que en las entregas de microfilm era costoso de realizar. Igualmente, en estos casos las imágenes están fuera de una base de datos pero los metadatos nos facilitan rapidez en la localización y en la identificación e información del documento original.