



RECOMENDACIONES PARA DIGITALIZAR DOCUMENTOS. USO DEL ESCÁNER



**Unidad de Documentación,
Archivo y Registro**

Este documento pretende ofrecer información y recomendaciones en el campo de la digitalización y satisfacer la necesidad de mejorar los conocimientos y prácticas de uso del escáner.

Los objetivos de estas recomendaciones son:

- a) normalizar la terminología
- b) proponer unos parámetros técnicos
- c) presentar unas normas accesibles a usuarios expertos y no expertos

¿POR QUÉ DIGITALIZAR?

La decisión de digitalizar, es decir, la conversión digital de documentos originales no digitales puede tomarse con objeto de:

- Mejorar el acceso a los documentos
- Reducir la manipulación y el uso de documentos originales
- Incrementar los servicios a los usuarios

REQUISITOS TÉCNICOS PARA ESCANEAR

Imagen digital

Es una “fotografía electrónica” convertida en un conjunto de elementos pictóricos (píxeles). El número de píxeles define la resolución de la imagen. La imagen digital de un documento la obtenemos por medio de un escáner.

Escáner

Dispositivo de digitalización que transmite los datos de las imágenes escaneadas a un ordenador. Durante la captura de la imagen los documentos se escanean a una resolución y a una profundidad predefinidas.

Resolución

La resolución del escáner se mide en puntos por pulgada (ppp) o píxeles por pulgada (ppi). A mayor número de píxeles tendremos mayor resolución y mayor capacidad para definir los detalles más precisos, pero también mayor tamaño de los ficheros donde se almacenan las imágenes.

Profundidad del bit

Se define como la capacidad de reproducción de colores. A mayor profundidad de bit pueden representarse mayor número de tonos en grises y color.

Compresión

Se utiliza para reducir el tamaño del fichero de la imagen para su proceso, almacenamiento y transmisión. La compresión puede ser con pérdida o sin pérdida de información.

Formatos

Las imágenes digitalizadas se pueden guardar en diferentes formatos.

TIFF es el líder indiscutido cuando se requiere la máxima calidad. Se trata de un formato estándar y de uso libre, y es utilizado como archivo de conservación. Su principal inconveniente es el alto volumen de almacenamiento debido al mayor tamaño del archivo que genera.

JPEG y PDF son formatos que permiten distintos grados de compresión. Son muy adecuados para distribuir información y mantienen una calidad muy buena. El PDF es perfecto para imprimir y crear documentos secuenciales con páginas múltiples. JPEG y GIF son los preferidos para sitios web y correo electrónico.

ELECCIÓN DEL ESCANER

Existen distintos tipos de escáner con diferentes formatos que se adaptan a cada necesidad de trabajo. La resolución de escaneo dada por el dispositivo de digitalización varía de acuerdo al precio y calidad del aparato. El rendimiento para la transmisión de datos al ordenador depende del tipo de conexión entre el escáner y el PC, desde las que transmiten los datos escaneados más rápidamente como la conexión SCSI o FireWire hasta las más lentas como el USB o de puerto paralelo.

Aspectos a tener en cuenta a la hora de seleccionar el equipo idóneo para escanear documentos.

1. **Tipo y volumen de documentos:** Si hay muchos documentos de una sola página o si en los documentos de varias páginas éstas están sueltas, se recomienda un escáner con alimentador automático de papel. El alimentador automático de hojas sueltas debe tener un sistema de alimentación que desplace la página simplemente evitando giros internos para evitar atascos o incluso roturas de los originales. En función del volumen, se pueden encontrar escáneres de velocidades que están entre 10 y 50 ppm en la gama profesional baja. Es decir, en una hora se podrían escanear entre 600 y 3.600 páginas.
2. **Estado de los documentos:** Si los documentos en papel que se van a digitalizar están en un estado regular de conservación, son antiguos, están rotos, el gramaje del papel es más fino o más grueso del que aceptan los alimentadores de los escáneres, recomendamos su digitalización utilizando un escáner de cristal plano.
3. **Digitalización simplex o dúplex:** ésta es una funcionalidad de mucho valor cuando los documentos en papel están impresos por las dos caras. Los equipos con esta capacidad digitalizan simultáneamente las dos páginas. Normalmente la velocidad baja a menos de la mitad por ello si se van a digitalizar muchos documentos de este tipo es conveniente valorar la velocidad de escaneado en modo dúplex.
4. **Código de barras:** Este sistema es muy eficiente cuando hay que digitalizar muchos documentos con el mismo formato y cuyo dato de referencia es el código de barras. Es ideal en la captura de albaranes, formularios, etc. En este caso, el escáner deberá tener alimentador automático de hojas sueltas y es conveniente la funcionalidad de filtro de colores porque habitualmente suelen ser documentos preimpresos con algún color que puede impedir la lectura correcta de información impresa sobre el mismo.

5. **Calidad de las imágenes para su posterior reimpresión:** normalmente todos los escáneres suelen tener distintos valores de resolución, contraste, brillo, formatos de digitalización, etc., todos ellos configurables para conseguir distintas calidades de las imágenes digitalizadas.

Escáner plano (Fig.1)

En este tipo de escáner el documento se coloca sobre el cristal y luego se cubre con la tapa, el dispositivo de digitalización (CCD) se mueve por debajo del cristal iluminando el original para digitalizar. Es importante que la tapa esté bien cerrada para que no se filtre luz por los costados, porque alteraría la calidad de la digitalización.

Escáner plano con alimentador automático de originales (Fig.2)

Este tipo de dispositivo es como un escáner plano con el agregado de un alimentador automático de papel como el de una impresora o una fotocopidora, este sistema funciona por un rodillo de goma porosa que arrastra el documento hasta el cristal. Su principal ventaja es que no hace falta abrir y cerrar la tapa para poner cada documento, sino que se los carga todos juntos en la bandeja y el alimentador los transporta uno por uno sobre la superficie del escáner. El inconveniente de esta tecnología es que para hacer llegar el original a la superficie de escaneo, el rodillo hace girar el papel por un rodillo y este proceso puede dañar el original, es por esto que antes de utilizar este tipo de escáner se debe controlar que los originales estén en buenas condiciones; si el papel del documento que deseamos digitalizar presenta roturas, arrugas, humedad o se encuentra en un estado de vejez avanzado, puede suceder que se trabe y arrugue dentro del alimentador dañándose gravemente.

Escáner de rodillo (Fig.3)

Los originales son introducidos por un elemento transportador de velocidad constante (alimentador automático de documentos -ADF-) a la fuente de luz/sensores ópticos del escáner. Las páginas pueden recorrer el mecanismo lector en forma horizontal o en forma de "u", pero en cualquier caso serán recogidas en el mismo orden de introducción. Pueden escanearse directamente documentos con múltiples páginas, con tiempos de respuesta rápidos. Desventajas: los originales deben ser hojas sueltas y no admite grosores de papel superiores a 150g/m².



Fig.1-Escáner plano



Fig.2- Escáner plano con alimentador



Fig.3-Escáner de rodillo

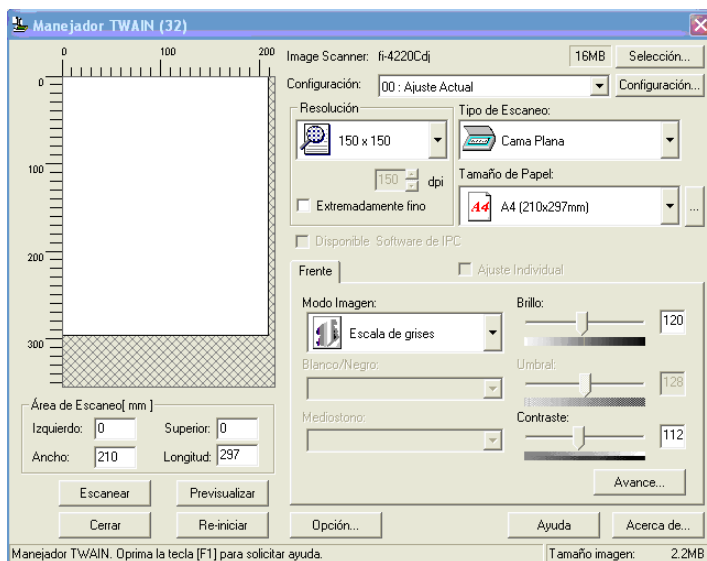
PROCESO DE CONVERSIÓN

1. Preparación del PC e instalación del escáner

Es necesario disponer de un PC con las características adecuadas para un buen funcionamiento del escáner y realizar una instalación apropiada del dispositivo. La instalación de un escáner es una tarea sencilla; basta enchufarlo, conectar el cable al puerto USB del ordenador (la mayoría utiliza esta conexión) y seguir las instrucciones del CD que le acompaña.

El funcionamiento de los escáneres es muy sencillo. Al instalar el dispositivo, se instala también un programa que permite manejarlo por parte de las diferentes aplicaciones de nuestro ordenador. Lo más habitual es que el escáner utilice la tecnología Twain. El driver Twain es un protocolo creado por fabricantes de software y hardware para que cualquier aplicación pueda adquirir imágenes desde dentro del propio programa, de manera que desde cualquier aplicación instalada en nuestro equipo se puedan digitalizar imágenes, modificarlas y salvarla en el formato deseado.

El driver TWAIN es propietario de cada marca y modelo, pero debe cumplir con las normas para que sea reconocido por cualquier aplicación compatible con TWAIN.



Ejemplo Interfaz Twain para un escáner mixto plano y rodillo

Muchos controladores de escáner ofrecen una **previsualización** de la imagen antes de adquirirla. En esta previsualización podemos seleccionar la parte de la imagen que nos interesa captar, con una herramienta de selección estándar (el recuadro de línea discontinua), y el tipo de lectura que hará el aparato (color, gris, blanco y negro, resolución). Una vez decidido, el dispositivo lee la imagen y envía la digitalización de la misma al programa desde el que hemos empezado la tarea.

2. Resolución correcta para cada proyecto

Elegir una resolución suficiente para capturar los más sutiles detalles requeridos de los documentos que se van a escanear. Antes de iniciar un trabajo con el escáner hay que decidir qué

resolución mínima puede escogerse sin perder la calidad necesaria en función del uso que se dará a la imagen. Si va a digitalizar un documento para:

- Utilizarlo en Internet, enviarlo por correo o visualizarlo en su pantalla, no necesita más de 72 ppp.
- Imprimirlo en una impresora láser o de inyección, entre 200 y 300 ppp es más que suficiente.



3. Elegir una profundidad de bit de acuerdo con las características de los documentos originales:

- Escaneo en blanco y negro (los documentos son digitalizados en binario, blanco y negro).
- Escaneo en escala de grises (los documentos son digitalizados con 256 gradaciones de negro a blanco).
- Escaneo en color (los documentos son digitalizados a colores).

4. Utilizar las técnicas de compresión adecuadas para guardar las imágenes y llevar a cabo ensayos de acceso a los ficheros con el fin de encontrar un equilibrio entre una calidad visual aceptable y el tamaño del fichero, de modo que el ordenador pueda acceder al mismo en un tiempo aceptable.

Un formato de imagen digital con compresión degenerativa como el JPEG debe ser utilizado para copias de consulta que requieran un tamaño en bytes reducido para su mejor consulta On-line tanto en una Intranet como en Internet. Los Masters o copias en alta resolución no deben estar procesados con un formato de compresión con pérdida de datos. Un archivo de extensión TIFF por ejemplo, puede ser recuperado por cualquier sistema operativo con un software de visualización estándar, no pertenece a ninguna empresa en especial y no comprime la información.

5. Archivo de ficheros

Los ficheros de imágenes producidos como resultado de la digitalización deben organizarse, nombrarse y describirse de manera adecuada a los objetivos establecidos.

