

Proceso D: Escaneo

Recursos necesarios:

- Personal: Digitalizador
- Computador
- Escáner
- Software para escanear
- Criterios de calidad (Catálogo)

Insumo: Documentos a digitalizar (copia u original).

Objetivos de calidad:

- Definir un tamaño óptimo para los archivos electrónicos, que facilite su manipulación en los distintos medios electrónicos.
- Definir una estructura lógica para las carpetas que contendrán los documentos en el computador, así como una nomenclatura estandarizada, que facilite la ubicación y acceso a estos nuevos documentos electrónicos, y la oportuna automatización del proceso.
- Conservar o mejorar, en la medida de lo posible, la imagen electrónica generada con respecto al documento original.

Resultados esperados:

- Documento escaneado de un modo completo, legible con una secuencia lógica de sus partes y consecutividad de sus páginas.
- Documento escaneado con una calidad de imagen no menor a la del documento original.
- Documento escaneado ubicado en la carpeta y con el formato de imagen elegido.
- Tamaño de los archivos generados apropiadamente definido para su uso en los distintos medios electrónicos.
- Contar con reglas de nomenclatura para los archivos electrónicos generados, que permitan el acceso rápido y eficiente a los documentos electrónicos, y las futuras automatizaciones del proceso.

Cuadro sinoptico de tareas. Proceso D: Escaneo

Tareas		Páginas
D.1	Corroborar el estatus del documento en el respectivo sistema de control	14
D.2	Crear la carpeta correspondiente al documento	14
D.3	Dividir el contenido del documento en varios archivos	15
D.4	Determinar tratamiento de imágenes	15
D.5	Guardar el documento en las carpetas correspondientes	18
D.6	Revisar las imágenes escaneadas	18
D.7	Revisar que los archivos estén completos en la carpeta correspondiente	18

Tarea D.1 Corroborar el estatus del documento en el respectivo sistema de control

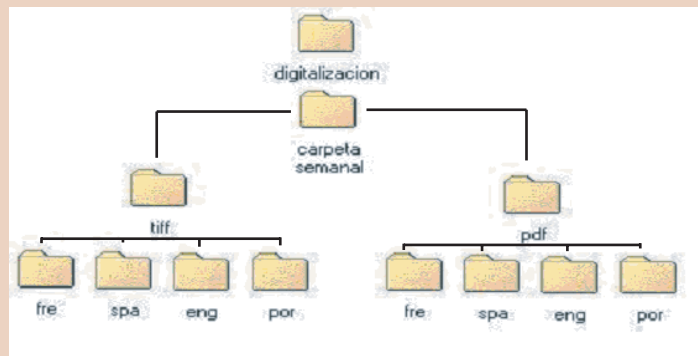
El digitalizador debe comprobar, en el sistema de control, que el documento que trabajará está ingresado con el estatus "en proceso" y no como digitalizado.

Además ingresará su nombre y fecha (mes y año) para contribuir al seguimiento del proceso.

Tarea D.2 Crear la carpeta correspondiente al documento

En este punto se requiere tener definida una estructura básica para las carpetas que contendrán los nuevos documentos, con la finalidad de organizar mejor la información.

Por ejemplo:



Es importante definir ciertas reglas de nomenclatura para poder establecer un orden, que permita el fácil acceso y ubicación de los archivos, así como una futura automatización del proceso, en el caso de producirse altos volúmenes.

Se recomienda utilizar un código único para cada documento y utilizar nombres cortos, sin tildes, mayúsculas, eñes, u otros caracteres especiales, a fin de:

- Asegurar la compatibilidad con cualquier sistema operativo y medio de almacenamiento que se planee utilizar.
- Mantener la capacidad de visualizar, recuperar y utilizar estas colecciones digitales en un marco tecnológico y organizativo que cambia con mucha rapidez.

El tamaño que se defina para cada uno de los archivos generados, así como las conexiones de red (Internet) y el tráfico en ella, son variables que definen la velocidad de acceso a la información.

La siguiente tabla muestra algunas referencias, en cuanto al tiempo de descarga de archivos con un tamaño aproximado a 1 Megabyte (1024 Kb):

Velocidad (Kbps)	Tiempo requerido (Segundos)
56	184
64	160
128	80
256	40
512	20
1024	10
2048	5,0
4096	3,0

Tarea D.3 Dividir el contenido del documento en varios archivos

Corresponde hacer esto si existen restricciones en cuanto al tamaño de los documentos. Esta división del contenido se basará en un estudio previo sobre el tamaño de archivo que se pretende manejar por los distintos medios electrónicos: CDS, Internet, etc.

La división del documento debe regirse por un tamaño definido. Igualmente importante es no dañar su estructura lógica, dejando las ideas inconclusas en una sección al separarlas inapropiadamente.

Por esta razón, es importante que también se definan algunas reglas de nomenclatura para la asignación de archivos a un mismo documento.

Por ejemplo, los archivos que forman parte del documento con código 1024 podrían nombrarse como doc1024-1, doc1024-2, etc.

Como referencia, se puede tomar la siguiente tabla, basada en la experiencia adquirida por el CRID.

TABLA DE REFERENCIA PARA LA ESTRUCTURACION DE DOCUMENTOS		
Descripción del archivo	Número aprox. de paginas	Tamaño aprox. del archivo Escaneado con una resolución de 200dpi
Texto simple, relativamente limpio y legible	15-20	700kb a 850kb
Texto simple, un poco "sucio" y legible	12-15	700kb a 800kb
Una fotografía en blanco y negro, tamaño de 1 pg	1	700kb a 1Mb
Una pagina de color	6 - 10	750Kb a 1.2Mb

Tarea D.4 Determinar tratamiento de imágenes

En este momento hay que analizar muy bien el tratamiento que se le dará al documento impreso, lo cual depende del estado que presenten sus imágenes.

Entre las opciones más importantes para el tratamiento de imágenes, durante el proceso de escaneo, tenemos:

Resolución:

Su aumento permite capturar detalles más precisos. Sin embargo, en algún punto, una mayor resolución no tendrá como resultado una ganancia evidente en la calidad de la imagen, sino un mayor tamaño de archivo. La clave es determinar la resolución necesaria para capturar todos los detalles importantes que están presentes en el documento fuente

Se puede utilizar la siguiente información como referencia:

Tipo de documento	Resolución Recomendada
A color	200 dpi
Texto Blanco y negro, tamaño de la letra normal y bien definida	200 dpi
Texto B/N, tamaño de la letra normal y no muy bien definida (letra clara)	250 dpi
Texto B/N, tamaño de la letra pequeña y bien definida	250 dpi
Texto B/N, tamaño de la letra pequeña y no muy bien definida (letra clara)	300 dpi
Fotografías B/N	200 dpi

Figura 1. Página B/N escaneada a 50 dpi.

3.4.6 Título sexto

De los residuos peligrosos

Las disposiciones especiales de
residuos peligrosos durante su transporte

- El transporte de residuos peligrosos

Figura 2. Página escaneada a 400 dpi.

3.4.6 Título sexto

De los residuos peligro

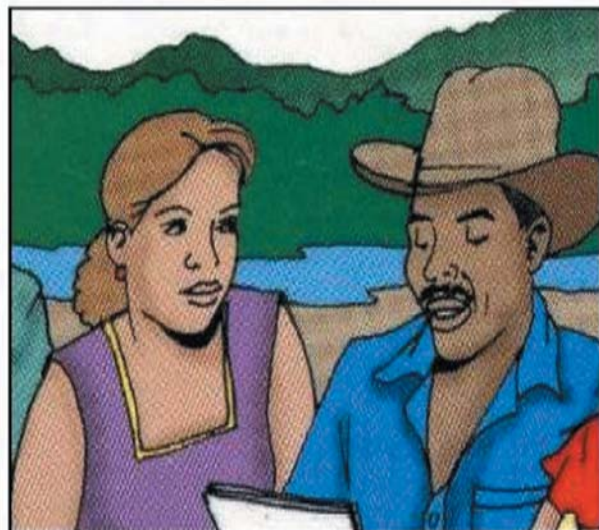
Las disposiciones especial
residuos peligrosos durante su tran

- El transporte de residuos

Figura 3. Página color escaneada a 75 dpi



Figura 4. Página color escaneada a 400 dpi



Profundidad de bits:

El aumento de la profundidad de bits o de la cantidad de bits utilizados para representar cada pixel (punto o elementos de la figura), permite capturar más matices de gris, o tonos de color. Como ya se comentara, una imagen bitonal está representada por pixeles, de 1 bit cada uno, que pueden representar dos tonos (típicamente negro y blanco). Una imagen a escala de grises está compuesta por pixeles representados por múltiples bits de información, que típicamente varían entre 2 y 8 o más. Una imagen a color está típicamente representada por una profundidad de bits entre 8 y 24 o superior.

"Rango dinámico" es el término utilizado para expresar la gama de tonos, desde el más claro de los claros hasta el más oscuro de los negros, que contiene una imagen.

La capacidad de un escáner para capturar el rango dinámico está regulada por la profundidad de bits que utilice y genere, así como también por el rendimiento del sistema.

Color (fidelidad del tono):

La reproducción correcta del color depende de una serie de variables, como el nivel de iluminación al momento de la captura, la profundidad de bits capturada y generada, las capacidades del sistema de escaneado y la representación matemática de la información del color.

Compresión

Se utiliza para reducir el tamaño del archivo de imagen para su almacenamiento, procesamiento y transmisión.

El tamaño del archivo para las imágenes digitales puede ser muy grande, complicando las capacidades informáticas y de redes de muchos sistemas. Todas las técnicas de compresión abrevian la cadena de código binario, en una imagen sin comprimir, a una forma de abreviatura matemática, basada en complejos algoritmos.

Existen técnicas de compresión estándar y otras patentadas (ver la tabla que se adjunta en el tema de formatos de archivo). En general, es mejor utilizar una técnica de compresión estándar (ITU-T.6, JBIG/JBIG2, JPEG, entre otros) y ampliamente compatible, en lugar de una patentada (LZW, ImagePac, entre otros) que puede ofrecer una compresión más eficiente y/o mejor calidad, pero que puede no prestarse a un uso o a estrategias de preservación digital a largo plazo.

Los sistemas de compresión también pueden caracterizarse como "sin pérdida" o "con pérdida".

Los sistemas sin pérdida, como el que utiliza el formato TIF, no desechan la información. Es decir, la imagen capturada es idéntica a la original. Los sistemas con pérdida, como el que utiliza el formato JPG, "desechan" la información menos importante, desde el punto de vista visual, aunque ésta, en algunos casos, puede ser imperceptible para el ojo humano.

Se recomienda utilizar la compresión sin pérdida en el escaneado de material impreso. La compresión con pérdida se puede aprovechar para trabajar con imágenes tonales, y en particular con imágenes de tono continuo.

Formatos de archivo:

Los formatos de archivo varían en términos de resolución, profundidad de bits, capacidades de color, y soporte para compresión

El formato de archivo más recomendable para almacenar imágenes es el TIFF, porque permite manejar varias páginas en un mismo archivo, además de ser un formato estándar y uno de los más utilizados para este tipo de tratamiento. Se utiliza para intercambiar archivos entre distintas aplicaciones y plataformas de la computadora, soporta un método de compresión con menor pérdida y, de hecho, ayuda a no perder detalles en la imagen almacenada.

La siguiente tabla resume algunas de las principales características de los formatos más comunes, en cuanto a la profundidad de bits, compresión, si es o no estándar, entre otros:

de sus partes y consecutividad de sus páginas, así como con una calidad de imagen no menor al del documento que lo originó.

Nombre	TIFF (Tagged Image File Format)	GIF (Graphics Interchange Format)	JPEG (Joint Photographic Expert Group)	PNG (Portable Network Graphics)	PDF (Portable Document Format)
Extensión (Extensiones)	.tif, .tiff	.gif	.jpeg, jpg, .jif, .jfif	.png	.pdf
Profundidad (es) de bits	Bitonal a 1 bit; escala de grises o color de paleta de 4 u 8 bits; hasta color de 64 bits	Bitonal, escala de grises o color entre 1 y 8 bits	Escala de grises a 8 bits; color a 24 bits	1-48 bits; color a 8 bits, escala de grises a 16 bits, color a 48 bits	Escala de grises a 4 bits; color a 8 bits; soporta hasta 64 bits para color
Compresión	Descomprimido sin pérdida	Sin pérdida	Con pérdida	Sin pérdida	Descomprimido Sin pérdida: ITU-T.6, LZW Con pérdida: JPEG
Estándar / patentado	Estándar	Estándar	Estándar	Estándar	Estándar
Soporte de Web	Conexión o aplicación externa	Originario desde Microsoft® Internet Explorer 3, Netscape Navigator® 2	Originario desde Microsoft® Internet Explorer 2, Netscape Navigator® 2	Originario desde Microsoft® Internet Explorer 4, Netscape® Navigator 4.04	Conexión o aplicación externa
Comentarios	Acepta imágenes y archivos múltiples	Se puede reem- plazar por PNG; Soporte de en- trelazado y transparencia a	JPEG progresivo, ampliamente so- portado por los	Puede reemplazar a GIF	Preferido para imprimir y ver documentos de

Tarea D.5 Guardar el documento en las carpetas correspondientes

El documento escaneado debe ser guardado en la carpeta correspondiente. Hay que tomar en cuenta las reglas de nomenclatura, anteriormente definidas, para nombrar cada archivo electrónico generado.

Es importante recordar, en este punto, que es imprescindible estandarizar la forma en que se nombrarán los archivos, para así facilitar la ubicación y acceso a estos nuevos documentos electrónicos y, en caso de requerirse, una oportuna automatización del proceso.

Tarea D.6 Revisar las imágenes escaneadas

Se recomienda revisar que el documento electrónico generado esté completo, legible con una secuencia lógica

Tarea D.7 Revisar que los archivos estén completos en la carpeta correspondiente.

Verificar la existencia de los archivos en la carpeta y su nomenclatura.