

MODULO IV

Digitalización de información impresa

INDICE

| | |
|--|----|
| Introducción | 3 |
| Sinopsis general: Procesos y tareas..... | 10 |
| Proceso A Selección de documentos..... | 15 |
| Proceso B: Preparación del material para digitalizar..... | 17 |
| Proceso C: Fotocopiado..... | 19 |
| Proceso D: Escaneo..... | 20 |
| Proceso E: Edición de imágenes | 30 |
| Proceso F: Control de calidad..... | 34 |
| Proceso G: Respaldo de la información | 35 |
| Proceso H: Diseño de páginas HTML..... | 37 |
| Proceso I: Presentación de documentos electrónicos en texto completo con motor de búsqueda..... | 43 |
| Proceso J: Presentación de documentos electrónicos en texto completo, sin motor de búsqueda, en el sitio Web | 46 |
| Proceso K: Producción de CDS temáticos con funciones interactivas | 50 |
| Bibliografía consultada: | 69 |

Introducción

El presente módulo ha sido elaborado como un recurso didáctico para la capacitación de personal en el proceso de digitalización de documentos en texto completo.

1. El proceso de digitalización. Aspectos generales

Dicho de una forma muy general, la digitalización es un proceso que permite introducir información a computadoras. Es dar *formato digital* a cualquier tipo de información gráfica, de audio o vídeo, es decir, convertirla al lenguaje del computador.

Existen miles de accesorios que permiten digitalizar información: teclados, escáneres, cámaras digitales fotográficas o de video, prótesis auditivas digitales y muchos otros.

La digitalización de documentos en papel consiste en su conversión a imágenes electrónicas, por medio de un escáner, las cuales son capturadas por medios de almacenamiento electrónico. Cuando se digitaliza material de audio y vídeo, se utilizan uno de los muchos procesos de conversión de análogo a digital, en el cual una variable continua (análoga) es cambiada a una señal de varios niveles (digital), sin alterar su contenido esencial.

2. Alcances y ventajas de la digitalización de documentos

A la hora de analizar el papel que la digitalización cumple en la actual “sociedad de la información”, podría ser suficiente con recordar que, sin este proceso, no sería posible producir información digital y, por ende, no tendríamos colecciones digitales, bibliotecas digitales ni tampoco un recurso actualmente utilizado en su trabajo y entretenimiento cotidiano por millones de personas: Internet.

Sin embargo, no estará de más referirse a algunas ventajas específicas de la digitalización, que se hacen notar particularmente en la capacidad de servicio de centros de información y en el trabajo de los usuarios:

- a) Disminuye el tiempo profesional invertido en la búsqueda y traslado de documentos y los gastos de impresión.
- b) Facilita significativamente la diseminación de los materiales reunidos. El acceso *on-line* a documentación permite eliminar costos de impresión múltiple de un mismo documento. Los documentos pueden consultarse simultáneamente. Para efectos de distribución masiva de información, se cuenta también con la posibilidad de almacenar esta información en CD-Roms, DVD's o difundirla por conducto de una red programada.
- c) Evita el almacenamiento de grandes cantidades de papel y, con ello, facilita un mejor aprovechamiento del espacio físico y la eliminación de costos directos y ocultos.

- d) Evita el desgaste o rotura que el manejo manual frecuentemente provoca en los documentos originales.
- e) Se reduce el riesgo por robo o extravío de documentos.
- f) Hace posible la formación de colecciones para la puesta en marcha de futuras bibliotecas virtuales. Más aún, al estar montados sobre una plataforma web, los procesos de digitalización brindan una posibilidad de crecimiento hacia nuevos servicios.
- g) Los usuarios pueden obtener información al instante y efectuar consultas inmediatas, mediante distintas opciones de extracción de información (índices de búsqueda, texto dentro del documento, reportes de un sector específico del documento, etc.). Además, los investigadores y el personal de los servicios especializados encuentran en la información digitalizada un recurso que posibilita intercambios y nexos de cooperación que antes eran lentos y difíciles.

3. Requerimientos sugeridos para la digitalización de documentos

i) Recursos humanos

Serán descritos más adelante, al examinar cada uno de los procesos involucrados en este quehacer.

ii) Computadora

Sus características más adecuadas dependerán del volumen de producción estimado. Por ejemplo, para producir 5.000 páginas mensuales (aproximadamente 2 GB, considerando formatos **pdf** y **tiff**), se puede utilizar una computadora con las siguientes características:

- Pentium III o similar
- Memoria RAM de 128 MB
- Disco duro de 20 GB
- Quemador de CD o ZIP para efectuar respaldos

La capacidad de la computadora tendrá que ser mayor, en términos de velocidad de procesamiento y almacenamiento, cuanto más grande sea el volumen de producción.

iii) Escáner

Un escáner es un artefacto diseñado para captar ópticamente imágenes o textos y llevarlos a un formato electrónico.

Físicamente, lo que hace es iluminar el objeto o documento a ser digitalizado y dirigir la luz reflejada (por lo general a través de una serie de espejos y lentes) sobre un elemento fotosensible. En la mayoría de los escáneres, el medio sensible es un circuito electrónico integrado, sensible a la luz, conocido como “dispositivo acoplado cargado” (CCD). Los fotositos sensibles a la luz, dispuestos a lo largo del CCD, convierten los niveles de brillo en señales electrónicas, que luego se procesan en una imagen digital.

En lo que respecta a la organización de la información en el formato digital, ésta se lleva a cabo mediante unidades discretas de datos o *bits*, que pueden ser tratadas por separado (grupos de 8 bits, usualmente llamados *bytes*).

Un escáner captura una imagen o texto y lo convierte en archivo de imagen, confeccionando un mapa en forma de cuadrícula de puntos o elementos de la figura (*píxeles*). A cada pixel se le asigna un *valor tonal* (negro, blanco, matices de gris o color), el cual se representa por el código binario (ceros y unos). Los dígitos binarios (bits) para cada pixel son almacenados por la computadora en una secuencia, y frecuentemente son reducidos mediante una representación matemática (comprimida). Luego la computadora lee e interpreta los bits para producir una versión analógica apta para ser visualizada o impresa.

Ejemplo: Valores de píxel. Imagen bitonal.

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 |

En este ejemplo, se le asigna a cada píxel un valor tonal: 0 para el negro y 1 para el blanco.

En la selección de un escáner hay que tomar en cuenta varios aspectos:

- Resolución. Se mide en puntos o píxeles por pulgada (dpi) y representa la nitidez o el grado de fidelidad de la copia con respecto al documento original. A mayor cantidad de píxeles, mayor nitidez. Se recomienda que ésta sea de 150 a 300 *dpi*.
- Opción de lector óptico de caracteres (OCR).

- Capacidad para comprimir formatos.
- Capacidad de producción: cantidad de páginas que admite la bandeja o alimentador de páginas.
- Velocidad, medida en páginas escaneadas por minuto. El escaneo a color es aproximadamente un 30% más lento que en blanco y negro.
- Posibilidad de escanear por un sólo lado (escáner “simplex”) o por ambos (escáner “duplex”).
- Tamaños de papel que acepta: carta, legal, etc.
- Solidez y durabilidad del aparato.

Hay dos *tipos* de escáner, que utilizan la misma tecnología básica:

i) **Planos** (o de cama plana): Disponen de un sensor de luz y una fuente de luz, ambos montados sobre un brazo móvil que pasa sobre un documento que se coloca sobre una placa de vidrio. Son versátiles y fáciles de manejar y también los más conocidos y vendidos. ii) **Con alimentador de páginas**: tanto el sensor de luz como la fuente de luz permanecen fijos mientras las páginas a escanear son transportadas por medio de rodillos, cinta, tambor o vacío. Maximizan el rendimiento, aunque por lo general a expensas de la calidad. Se emplean habitualmente cuando se requiere escanear grandes volúmenes de documentos.



PLANO



CON ALIMENTADOR DE PÁGINAS

Ventajas y desventajas comparativas

| Tipo de escáner | |
|---|-----------------------------------|
| Plano | Con alimentador de páginas |
| Menor precio (alrededor de \$ 300) | Más costoso (entre \$ 500 y 3000) |
| Mayor calidad de salida | Menor calidad de salida |
| Permite escanear por un solo lado | Permite escanear por ambos lados |
| Ambos permiten escanear en blanco y negro o a color | |

v) Software de escaneo

Normalmente todo escáner trae su propio software para escanear y editar imágenes. Sin embargo, sólo unos pocos incorporan todas las opciones necesarias para editar imágenes de calidad. Por tal razón, también se utilizan alternativamente otras aplicaciones:

- *Adobe Photoshop®*,
- *Pagis Pro®*,
- *Textbridge®*,
- Las conocidas como *Imaging®*,, *Paint®*, y *Microsoft Photo Editor®*,, que vienen con el Microsoft ®,y el *MS Office®*.

Los elementos más importantes a tomar en cuenta para que una aplicación de software logre producir un documento digital de calidad son:

- Los tipos de resolución (recomendable de 150dpi a 300 dpi)
- Manejo del contraste (imagen clara a oscura)
- Opción de lector óptico de caracteres (OCR)
- La capacidad para comprimir formatos
- La capacidad de eliminar del texto impreso todo aquello que le reste calidad en su presentación (manchas, puntos oscuros, sobrescritos, texto torcido), como consecuencia de su antigüedad, manipulación constante, fotocopiado o cualquier otro tratamiento al que generalmente es sometido un documento impreso.

v) Sistema de control

Es aconsejable utilizar algún sistema de control en la producción de documentos electrónicos, a fin de monitorear el estatus en que se encuentra cada uno de los documentos seleccionados, generar un reporte estadístico/histórico de todos los documentos digitalizados y contar con un sistema de información que permita realizar diversas consultas sobre los documentos ya digitalizados o que están en proceso de serlo.

La manera de llevar este control depende de la cantidad de documentos que se planea manejar en el proceso. Varía desde una hoja de Excel hasta una pequeña base de datos en Access.

Se recomienda definir los siguientes campos o variables:

- ID del documento (Número de identificación del documento)
- Estatus: En proceso – Digitalizado
- Mes
- Año
- Digitalizador

v) Otros recursos necesarios:

- Fotocopiadora
- Espacio físico necesario para acomodar al personal, cuyo número depende directamente de la cantidad de producción proyectada,

Sin dejar de lado otros aspectos ergonómicos, se recomienda un área de trabajo mínima de 1.5m x 2m por puesto físico de trabajo (ubicación de una persona con el equipo necesario para realizar la labor de digitalización).

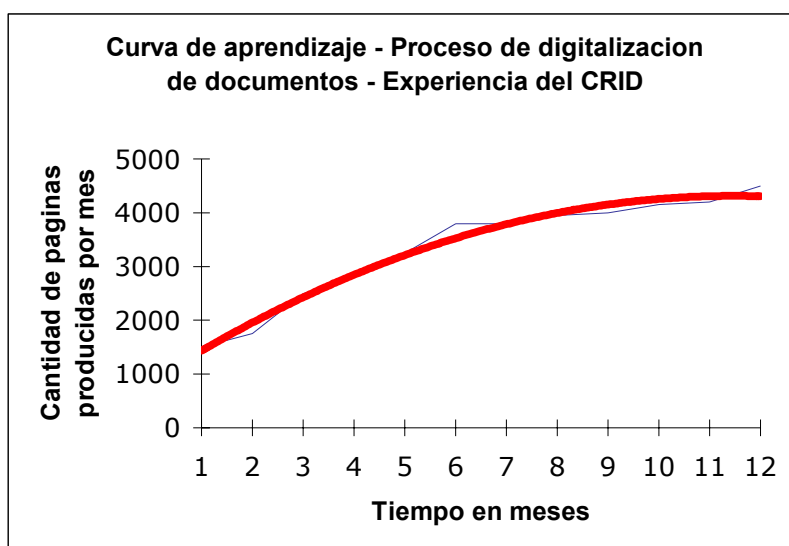
- Materiales: papelería, CDs, ZIPs entre otros.

2. Aspectos relativos a la evaluación de los procesos de digitalización

i) Curva de aprendizaje

Las curvas de aprendizaje se pueden aplicar a individuos, equipos de trabajo u organizaciones. Son gráficas que muestran la mejora obtenida cuando las personas repiten un proceso y adquieren habilidad o eficiencia a partir de su propia experiencia.

Con base en la experiencia del CRID, se puede estimar que, al cabo de 6 meses, una persona a tiempo completo estará en capacidad de producir alrededor de 4000 páginas digitalizadas por mes (considerando los factores especificados bajo el gráfico siguiente), con un nivel de calidad bastante bueno, alcanzando en este punto el nivel máximo de producción con la calidad requerida.



-**Texto impreso:** Dibujos de líneas simples – representación en base a bordes definidos, sin variación de tono, como un libro que contiene texto y gráficos de líneas simples.

-**Tono continuo:** elementos tales como fotografías, y algunos dibujos que exhiben tonos que varían suave o sutilmente.

-**Combinado:** documentos que contienen las dos categorías mencionadas anteriormente. Por ejemplo: libros ilustrados.

- **Tipo de escáner utilizado:** Escáner con alimentador de páginas, con una velocidad de 15 páginas por minuto.

ii) Productividad

En general, la productividad del desempeño individual, de equipo u organizacional expresa la siguiente relación:

$$\text{Producción} = \frac{\text{Rendimiento} + \text{Calidad}}{\text{Consumo material y humano}}$$

En nuestro caso, el punto óptimo de productividad corresponde a la mayor cantidad de producción de páginas con el mínimo porcentaje de rechazo (páginas devueltas por no cumplir con los requisitos de calidad definidos)

Según la experiencia del CRID, una productividad óptima sería 4.000 páginas mensuales por persona (a tiempo completo) con un nivel aceptable de rechazo de aproximadamente el 1% (es decir, un máximo de 40 observaciones que incumplan los criterios de calidad definidos).

En el caso de que la productividad no sea la esperada, se recomienda analizar el método de trabajo utilizado y las condiciones en que se encuentra el material y el equipo. Asimismo, si se plantea la posibilidad de incrementar la producción, convendrá analizar si con ello se pone en riesgo la calidad del material digitalizado. Si a esto se agregara un mayor consumo de recursos, se afectaría doblemente la productividad de las operaciones.

iii) Calidad

En principio, es el conjunto de condiciones que determinan que un servicio de digitalización resulte satisfactorio para las necesidades de los usuarios. No obstante, no se trata de un criterio a asumir pasivamente, porque la organización siempre tiene la posibilidad de mejorar sus servicios e influir educativamente sobre las expectativas de quienes los reciben.

Siempre es importante responder con tiempo y profesionalismo a esas expectativas, aumentar el valor que los usuarios perciben en los productos o servicios que se les suministran y actuar preventivamente para disminuir los costos derivados de fallas en la calidad del servicio.

Tomando en cuenta que la producción de documentos en formato electrónico es un proceso esencialmente dinámico, que evoluciona según las necesidades de información de los usuarios y el progreso tecnológico que se produce día a día, se convierte en parte integral de la gestión de calidad, orientada a su mejoramiento continuo.

En las tareas de digitalización, deben utilizarse todas las opciones que permitan obtener un producto satisfactorio y mejorar, o al menos mantener, la presentación de la información. No sólo están de por medio ciertos aspectos estéticos, sino también la optimización de los procesos que repercuten directamente en la funcionalidad del producto final.

Para unificar criterios en la evaluación de este quehacer, es importante elaborar un documento, de preferencia ilustrado, con las especificaciones de calidad pertinentes.

Procesos involucrados en la digitalización de documentos. Sinopsis

En este módulo, el quehacer general de digitalización se presenta desagregado en once "procesos". De éstos, siete son definidos como "básicos" y cuatro como "complementarios". Teniendo en cuenta que en este quehacer puede participar personal especializado con diferentes funciones, esta desagregación puede contribuir a organizar su desempeño como equipo.

Sinopsis general: Procesos y tareas

| | |
|---------------------------------|--|
| Procesos básicos | <p>A: Selección de documentos</p> <p>B: Preparación del material para digitalizar</p> <p>C: Fotocopiado</p> <p>D: Escaneo</p> <p>E: Edición de imágenes</p> <p>F: Control de calidad</p> <p>G: Respaldo de la información</p> |
| Procesos complementarios | <p>H: Diseño de paginas HTML</p> <p>I: Presentación de documentos electrónicos en texto completo con motor de búsqueda</p> <p>J: Presentación de documentos electrónicos en texto completo sin motor de búsqueda</p> <p>K: Producción de CDs temáticos con funciones interactivas (en general: elaboración de productos innovadores a partir de la digitalización de documentos)</p> |

| Cuadro sinóptico de tareas. Proceso A: Selección de documentos | |
|--|----------------|
| Tareas | Páginas |
| A.1 Definir criterios de selección de documentos | 16 |
| A.2 Definir criterios de selección de documentos | 16 |
| A.3 Verificar si corresponde pedir permisos de autor | 16 |
| A.4 Revisar la integridad física del documento | 16 |
| A.5 Ingresar el documento seleccionado al sistema de control de producción de documentos electrónicos | 17 |
| A.6 Enviar el documento al coordinador o encargado del proceso de digitalización | 17 |

| Cuadro sinoptico de tareas. Proceso B: Preparación del material para digitalizar | |
|---|----------------|
| Tareas | Páginas |
| B.1 Determinar condiciones de los documentos | 18 |
| B.2 Separar los documentos para el proceso de digitalización | 18 |
| B.3 Corroborar el listado de documentos generados | 18 |
| B.4 Ingresar al sistema de control de documentos la cantidad de páginas por documento. | 19 |
| B.5 Entregar los documentos al departamento de fotocopiado o al departamento de digitalización | 19 |

| Cuadro sinóptico de tareas. Proceso C: Fotocopiado | |
|---|----------------|
| Tareas | Páginas |
| C.1 Fotocopiar el documento y entregarlo a los digitalizadores | 19 |

| Cuadro sinoptico de tareas. Proceso D: Escaneo | |
|---|----------------|
| Tareas | Páginas |
| D.1 Corroborar el estatus del documento en el respectivo sistema de control | 21 |
| D.2 Crear la carpeta correspondiente al documento | 21 |
| D.3 Dividir el contenido del documento en varios archivos | 23 |
| D.4 Determinar tratamiento de imágenes | 23 |
| D.5 Guardar el documento en las carpetas correspondientes | 30 |
| D.6 Revisar las imágenes escaneadas | 30 |
| D.7 Revisar que los archivos estén completos en la carpeta correspondiente | 30 |

| Cuadro sinoptico de tareas. Proceso E: Edición de imágenes | |
|---|----------------|
| Tareas | Páginas |
| E.1 Editar imagen | 32 |
| E.2 Revisar el tamaño del archivo | 33 |
| E.3 Subdividir el archivo | 33 |
| E.4 Guardar el documento en formato TIFF | 33 |
| E.5 Convertir el archivo a otro formato (si fuera necesario) | 33 |
| E.6 Cambiar el estatus del documento en el sistema | 34 |
| E.7 Sellar el documento original como digitalizado | 34 |
| E.8 Empacar fotocopias de documentos | 34 |

| Cuadro sinoptico de tareas. Proceso F: Control de calidad | |
|--|----------------|
| Tareas | Páginas |
| F.1 Revisar y aprobar la calidad del documento | 34 |

| Cuadro sinoptico de tareas. Proceso G: Respaldo de la información | |
|--|----------------|
| Tareas | Páginas |
| G.1 Respaldo los documentos procesados | 36 |

| Cuadro sinóptico de tareas. Proceso H: Diseño de paginas HTML | |
|--|----------------|
| Tareas | Páginas |
| H.1 Seleccionar o diseñar una plantilla predefinida desde el programa de software. | 40 |
| H.2 Llenar cada uno de los campos en la tabla de la pagina HTML | 41 |
| H.3 Revisar cada uno de los enlaces del documento (links) y la información general. | 41 |
| H.4 Revisar los aspectos estéticos de la página y el texto que la contiene | 41 |
| H.5 Control de calidad en el código | 41 |
| H.6 Guardar la página Web | 42 |

| Cuadro sinóptico de tareas. Proceso I: Presentación de documentos electrónicos en texto completo con motor de búsqueda | |
|---|----------------|
| Tareas | Páginas |
| I.1 Agregar la dirección electrónica, en el campo correspondiente de la base datos | 43-44 |
| I.2 Actualizar las carpetas que contienen los documentos electrónicos en el Servidor Web | 43-44 |
| I.3 Actualizar las carpetas que contienen los documentos electrónicos en el Servidor Web | 43-44 |

| Cuadro sinóptico de tareas. Proceso J: Presentación de documentos electrónicos en texto completo sin motor de búsqueda | |
|---|----------------|
| Tareas | Páginas |
| J.1 Actualizar las carpetas de documentos electrónicos en el Servidor Web | 47 |
| J.2 Digitar los datos correspondientes a los documentos por ingresar al Sitio | 47 |
| J.3 Hacer los enlaces correspondientes a las carpetas en el Servidor Web y verificar su funcionalidad. | 47 |
| J.4 Actualizar las plantillas html en el Servidor Web | 47 |

| Cuadro sinóptico de tareas. | |
|---|----------------|
| Proceso K: Producción de CDs temáticos con funciones interactivas | |
| Tareas | Páginas |
| K.1 Seleccionar los documentos por incluir en el CD | 49 |
| K.2 Clasificar los documentos seleccionados por temas | 49 |
| K.3 Preparar las plantillas HTML | 50 |

Proceso A Selección de documentos.

Recursos necesarios:

- Personal: Analista de información
- Computador
- Sistema de control de producción de documentos electrónicos
- Criterios de selección que permitan definir temáticas de interés

Insumo: Documentos impresos con clasificación numérica para su identificación.

Objetivos de calidad:

- Seleccionar documentos que satisfagan los niveles de demanda de los usuarios por temas específicos y actuales.
- Seleccionar documentos que presenten condiciones físicas aptas para ser digitalizados, es decir, sean completos, legibles y sus páginas se encuentren ordenadas consecutivamente.
- Mejorar, mediante el proceso de digitalización, la legibilidad, nitidez y apariencia general de aquellos documentos impresos deteriorados por diferentes causas (manipulación, humedad, polvo y otras condiciones).
- Respetar los derechos de autor cuando el documento lo expresa en su contenido.

Resultados esperados:

Este proceso debe dar como resultado documentos que:

- Tengan características físicas aptas para ser digitalizados.
- Correspondan a temáticas de interés de los usuarios.
- Sean legalmente reproducibles
- Se encuentren ingresados en el sistema de control de producción de documentos digitalizados (ID – Estatus)

Cuadro sinóptico.
Proceso A: Selección de documentos

| Tareas | Páginas |
|---|----------------|
| A.1 Definir criterios de selección de documentos | 16 |
| A.2 Verificar la no existencia del documento en formato electrónico. | 16 |
| A.3 Verificar si corresponde pedir permisos de autor | 16 |
| A.4 Revisar la integridad física del documento | 16 |
| A.5 Ingresar el documento seleccionado al sistema de control de producción de documentos electrónicos | 17 |
| A.6 Enviar el documento al coordinador o encargado del proceso de digitalización | 17 |

Tarea A.1 Definir criterios de selección de documentos

Se deben tomar de cuenta los criterios generales de selección mencionados en los objetivos de calidad, así como otros pertinentes.

Responsable: Analista de información.

Tarea A.2 Verificar la no existencia del documento en formato electrónico.

Para este fin es recomendable:

- a) Asegurarse mediante una búsqueda general en Internet.
- b) Llevar un control físico de estatus, sellando cada documento una vez que haya sido digitalizado. Adicionalmente, esto permite agilizar las decisiones con respecto al proceso de digitalización.

Tarea A.3 Verificar si corresponde pedir permisos de autor

Es necesario determinar si el documento tiene derechos de autor y como operan en ese caso específico. Si existe prohibición para publicarlo, se recomienda gestionar por escrito el permiso del autor para reproducirlo.

Tarea A.4 Revisar la integridad física del documento

Es necesario efectuar una revisión general del documento, para determinar si es legible, se encuentra completo y en lo posible tiene sus páginas ordenadas consecutivamente.

Tarea A.5 Ingresar el documento seleccionado al sistema de control de producción de documentos electrónicos

Este sistema de control, necesario para darle seguimiento al documento del caso, puede ser montado de diversas maneras, desde una hoja Excel hasta una pequeña base de datos en Access.

Se recomienda que el perfil de seguridad definido en la base de datos le permita, al analista de información, ingresar o borrar en el sistema el ID del documento y su estatus.

Tarea A.6 Enviar el documento al coordinador o encargado del proceso de digitalización

El analista debe enviar el documento al encargado del proceso de digitalización, una vez ingresada la información al sistema.

Proceso B: Preparación del material para digitalizar

Recursos necesarios:

- Personal: Encargado del proceso de digitalización
- Computador
- Sistema de control de la producción de documentos electrónicos.

Insumos: Documentos aprobados por analista de información

Objetivos de calidad:

- Preservar la integridad física del documento original.
- Proporcionar el material en las condiciones necesarias para una eficiente manipulación en el proceso de digitalización.
- Controlar la salida y entrada de documentos al proceso de digitalización.

| Cuadro sinòptico. | |
|--|----------------|
| Proceso B: Preparación del material para digitalizar | |
| Tareas | Páginas |
| B.1 Determinar condiciones de los documentos | 18 |
| B.2 Separar los documentos para el proceso de digitalización | 18 |
| B.3 Corroborar el listado de documentos generados | 18 |
| B.4 Ingresar al sistema de control de documentos la cantidad de páginas por documento. | 19 |
| B.5 Entregar los documentos al departamento de fotocopiado o al departamento de digitalización | 19 |

Resultado esperado: Lograr que el material quede correctamente registrado en el sistema de control de documentos y en condiciones aptas para ser procesado.

Tarea B.1 Determinar condiciones de los documentos

Hay que revisarlos para asegurarse de que presentan condiciones aptas para ser digitalizados.

Tarea B. 2 Separar los documentos para su digitalización

Corresponde al encargado de digitalización decidir si los documentos originales pasan directamente a los digitalizadores o deberán ser previamente fotocopiados.

Tal decisión debe estar basada en aspectos como: tipo de escáner (plano o con alimentador) con el que se cuenta, ritmo de producción con el cuál se quiera trabajar (volumen alto o bajo de producción), preservación del buen estado físico de los documentos y calidad de imagen.

Se recomienda fotocopiar los documentos originales cuyas hojas sean difíciles de manipular (empastados), para evitar su maltrato en el escáner.

Tarea B.3 Corroborar el listado de documentos generados

El sistema de control de documentos permite controlar la entrada de documentos originales al proceso de digitalización, con el objetivo de evitar pérdidas de material por manipulación durante el proceso.

Tarea B.4 Ingresar, al sistema de control de documentos, la cantidad de páginas por documento.

Se recomienda llevar un control de las páginas digitalizadas, a fin de conocer mejor el ritmo de producción del proceso. Con este dato, se puede proyectar la cantidad y la capacidad de producción del personal involucrado. Adicionalmente, esto puede ser útil si la remuneración correspondiente se calcula por unidad producida.

Tarea B.5 Entregar los documentos al departamento de fotocopiado o al departamento de digitalización, (según lo explicado en Tarea B.2).

Proceso C: Fotocopiado

Recursos necesarios:

- Personal encargado
- Fotocopiadora
- Criterios de calidad (Catálogo)

Insumo: Documentos aprobados para ser fotocopiado

Objetivos de calidad:

- Velar por la integridad física de los documentos originales.
- Mejorar o mantener la calidad de su imagen
- Producir copias que conserven la características que hacen a los originales aptos para ser digitalizados: legibilidad, consecutividad de las páginas y estar completos.

| Cuadro sinóptico. Proceso C: Fotocopiado | | |
|---|--|---------------|
| | Tareas | Página |
| C.1 | Fotocopiar el documento y entregarlo a los digitalizadores | 19 |

Tarea C.1 (única) Fotocopiar el documento y entregarlo a los digitalizadores

a) El documento debe ser inspeccionado mientras se fotocopia, ya que algunas imágenes pueden mejorarse con este proceso. Las opciones de fotocopiado - densidad, resolución, ampliación - pueden establecer una diferencia positiva entre el documento impreso y el que será digitalizado.

b) Una vez que se tiene la certeza de que las fotocopias están completas, compaginadas consecutivamente y en buenas condiciones para iniciar el proceso de digitalización, se les entregan a los digitalizadores.

Nota: En el caso de que las fotocopias obtenidas empeoren la imagen del texto impreso, se recomienda entregar el documento original junto con ellas, para que así el digitalizador pueda escoger el material que juzgue más conveniente para su trabajo.

Proceso D: Escaneo

Recursos necesarios:

- Personal: Digitalizador
- Computador
- Escáner
- Software para escasear
- 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 sinóptico. Proceso D: Escaneo | |
|---|---------------|
| Tareas | Página |
| D.1 Corroborar el estatus del documento en el respectivo sistema de control | 21 |
| D.2 Crear la carpeta correspondiente al documento | 21 |
| D.3 Dividir el contenido del documento en varios archivos | 23 |
| D.4 Determinar tratamiento de imágenes | 23 |
| D.5 Guardar el documento en las carpetas correspondientes | 30 |
| D.6 Revisar las imágenes escaneadas | 30 |
| D.7 Revisar que los archivos estén completos en la carpeta correspondiente | 30 |

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

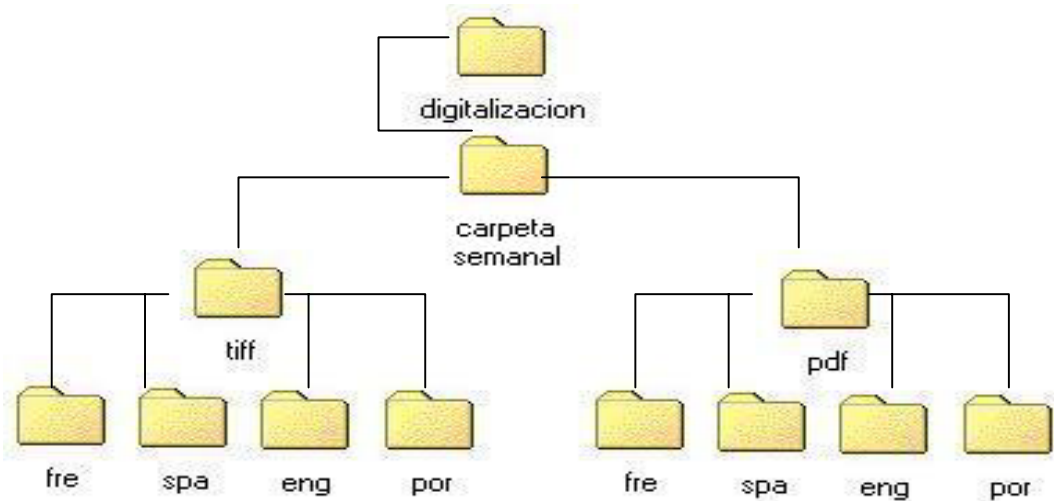
El digitalizador debe comprobar, en el sistema de control, que el documento que trabajará está ingresado con el estatus “en proceso” y no como digitalizado.

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 |

OBSERVE:

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:

| 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, .jiff | .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 reemplazar por PNG; Soporte de en-trelazado y transparencia a través de la mayoría de los navegadores Web | JPEG progresivo, ampliamente soportado por los navegadores Web | Puede reemplazar a GIF | Preferido para imprimir y ver documentos de páginas múltiples |

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

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.

Proceso E: Edición de imágenes

Recursos involucrados:

- Recurso humano: Digitalizador
- Computador
- Software para edición de imágenes
- Software para conversión de formatos
- Criterios de calidad (Catálogo)

Insumo: Documento electrónico

Objetivo de calidad:

- Eliminar del texto todo aquello que le reste calidad en su presentación - manchas, puntos oscuros, sobrescritos, texto torcido – como producto de su antigüedad, manipulación constante, fotocopiado o cualquier otro tratamiento.

Resultados esperados:

- Obtener un documento digitalizado de calidad, esto es:
 - con resolución de calidad no menor a la del documento original
 - sin elementos en mal estado (texto tachado, texto ilegible, manchas, bordes sombreados, ni torcido)
 - legible
 - completo y con sus páginas ordenadas consecutivamente
- Secuencia lógica en la nomenclatura de los archivos
- Archivo en algún formato conveniente para el uso que se le dará al documento electrónico
- Registro en el sistema de control de documentos
- Plan para la distribución del material fotocopiado

| Cuadro sinóptico. Proceso E: Edición de imágenes | | |
|---|---|----------------|
| | Tareas | Páginas |
| E.1 | Editar imagen | 33 |
| E.2 | Revisar el tamaño del archivo | 33 |
| E.3 | Subdividir el archivo | 33 |
| E.4 | Guardar el documento en formato TIFF | 33 |
| E.5 | Convertir el archivo a otro formato (si fuere necesario) | 33 |
| E.6 | Cambiar el estatus del documento en el sistema | 34 |
| E.7 | Sellar el documento original como digitalizado. | 34 |
| E.8 | Empacar fotocopias de documentos | 34 |

Tarea E.1: Editar imagen

Una vez escaneada la imagen, normalmente se hace necesario editar el archivo, aplicando una serie de técnicas para lograr la calidad y nitidez final deseada en el documento:

Limpieza

Consiste en eliminar/borrar todo aquello que no sea parte del contenido del documento, tal como: manchas, rayas, efecto granulado y otras inconsistencias producto de la manipulación.

Recorte

Para suprimir bordes oscuros de la página, producidos por su escaneo y fotocopiado, o para mover alguna imagen del texto

Enderezado

Corregir eventuales inclinaciones de las páginas, producidas por su escaneo y/o fotocopiadas.

Reducir el tamaño de la imagen

Esto se hace con frecuencia ya que, al reducir el tamaño original de una imagen, ésta ocupa menos espacio de almacenamiento.

Si no se utiliza un buen algoritmo de amplitud versus altura, normalmente se degradará un poco la resolución de la imagen.

Rotación

Cambiar la orientación de las páginas, lo cual puede hacerse desde 0 hasta 360 grados.

Maximización visual (zooming)

Agrandamiento temporal de la imagen visual. Muy útil cuando se quiere corregir detalles con mayor exactitud.

Edición:

Permite la modificación de la imagen. Puede ser muy útil cuando se requiere reconstruir documentos



Refiérase al catálogo de calidad (Anexo 1) para observar los resultados del proceso de edición de imágenes

Tarea E.2 Revisar el tamaño del archivo

Si anteriormente se definió una restricción en este aspecto, debe chequearse el tamaño del archivo después de la edición de las imágenes que lo contienen, ya que algunas ocasiones su tamaño podría aumentar.

Tarea E.3 Subdividir el archivo

Si el archivo sobrepasó el límite definido en su restricción por tamaño, es necesario subdividirlo, y adaptar de nuevo las reglas de nomenclatura definidas.

Tarea E.4 Guardar el documento en formato TIFF

Este es el formato de salida estándar más recomendable para almacenar imágenes, porque permite manejar multipáginas.

Además, este formato se utiliza para intercambiar archivos entre distintas aplicaciones y plataformas de la computadora, y soporta un método de compresión con menor pérdida

Tarea E.5 Convertir el archivo a otro formato (si fuere necesario)

Aquí se presentan dos situaciones diferentes, según se emplee:

a) El **reconocimiento óptico de caracteres (OCR)** en el formato de texto. En este caso, se reconocen los caracteres contenidos en la imagen de un documento y éste pasa a ser un texto editable.

El proceso OCR no siempre reconoce en su totalidad el contenido del documento. No interpretará correctamente algunas palabras, o algunos caracteres pueden ser erróneos, dependiendo de la complejidad del programa, así como de la calidad y el tipo de original que se maneje.

b) El **Formato PDF (Portable Document Format)** : Este formato de archivo es uno de los más utilizados para desplegar imágenes de documentos, razón por la cual la mayoría de los manuales y documentación que encontramos en Internet están en formato PDF.

Algunos softwares de escaneo traen la opción de guardar la imagen capturada directamente en este formato o la de convertir un archivo TIFF a PDF.

Por otro lado, es posible conseguir algún software gratuito en Internet para convertir imágenes a PDF. Sin embargo, si se desea tener acceso a una mayor cantidad de opciones para manipular y corregir los archivos PDF, es recomendable comprar la licencia del Adobe Acrobat®.

Hay varias razones que hacen aconsejable guardar los archivos digitalizados en formato pdf:

- Para abrir un archivo PDF sólo se necesita tener el software gratuito Adobe Reader®.
- Los archivos PDF son compactos, permiten la realización de búsquedas y puede accederse a ellos en cualquier momento.

Tarea E.6 Cambiar el estatus del documento en el sistema

Registrarlo como **Digitalizado**, incluir el Nombre del digitalizador y la fecha (mes, año). Debe quedar completa la información de la base de datos o sistema de control de la producción.

Tarea E.7 : Sellar el documento original como digitalizado.

Se recomienda sellar el documento como digitalizado, ya que así la colección tendrá información acerca de su existencia en formato electrónico y eso contribuye a agilizar la selección de documentos.

Tarea E.8 Empacar fotocopias de documentos

Los documentos que hayan sido fotocopiados deben empacarse, debidamente identificados, y entregarse al centro de información para que se encarguen de su distribución.

Proceso F: Control de calidad

Recursos necesarios:

- Personal: Encargado del proceso o inspector de calidad
- Computadora
- Software para visualizar el documento electrónico producido
- Criterios de calidad (Catálogo)

Insumo: Documento electrónico

Objetivo de calidad:

- Proporcionar documentos electrónicos con la mejor calidad de imagen posible.
- Identificar posibles fallas en los métodos de trabajo

Resultados esperados:

- Documento electrónico: legible, completo, con sus páginas consecutivamente ordenadas, texto en una correcta posición, imágenes claras y nítidas.
- Reglas de nomenclatura correctamente aplicadas.
- Tamaño del archivo adecuado a posibles restricciones definidas.

| Cuadro sinóptico. Proceso F: Control de calidad | |
|--|---------------|
| Tareas | Página |
| F.1 Revisar y aprobar la calidad del documento | 34 |

Tarea F.1 (única) Revisar y aprobar la calidad del documento

- a) Revisar totalmente el documento electrónico, basándose en los criterios de calidad definidos anteriormente.
- b) Llevar un registro de los tipos de errores encontrados en la inspección, con la finalidad de definir nuevos métodos que minimicen las fallas que se hayan presentado en procesos anteriores.
- c) Notificar al digitalizador respectivo del porcentaje de error obtenido así como el desglose de los mismos para su modificación.
- d) Verificar que el trabajo notificado haya sido corregido.

Proceso G: Respaldo de la información

Recursos necesarios:

- Personal: Encargado del proceso
- Computador
- Carpeta de documento electrónico
- Servidor
- Cintas magnéticas, Zip, Cds

Insumos:

- Base de datos o sistema de control de la producción

- Material digitalizado

Objetivo de calidad:

Asegurar la integridad y la disponibilidad de toda la información digitalizada.

Resultado esperado:

Respaldos de información digital actualizados mensualmente en el disco duro del computador de digitalización, en el disco duro del servidor y en cintas magnéticas.

| <i>Cuadro sinóptico. Proceso G: Respaldo de la información</i> | |
|--|---------------|
| <i>Tareas</i> | <i>Página</i> |
| G.1 Respaldo los documentos procesados | 36 |

Tarea G.1 (única) Respaldo los documentos procesados

Debe efectuarse a nivel local y externo.

A nivel local, se requiere que la información contenida en el disco duro de la computadora permanezca ahí durante la semana y tener un respaldo, en ZIP o CD regravable, de todos los documentos en proceso.

A nivel externo: debe realizarse un respaldo mensual del disco duro, que contiene la información digitalizada y la base de datos, en el servidor y en las cintas magnéticas.

Se recomienda realizar cortes de producción semanales para llevar un mejor control de los documentos digitalizados y evitar la acumulación de material.

Al finalizar la producción de un determinado mes, deben quedar respaldos generales, que contengan los archivos TIFF así como los PDF y la base de datos o sistema de control de la producción, en los siguientes medios: en el disco duro del computador de digitalización, en el disco duro del servidor y en cintas magnéticas.

PROCESOS COMPLEMENTARIOS

Los procesos que se analizan a continuación no forman parte del proceso de digitalización en sí, pero le dan un valor agregado:

Proceso H: Diseño de páginas HTML

Recursos necesarios:

- Personal: Encargado del proceso
- Computadora
- Software para editar páginas Web HTML

Insumo: Archivo(s) electrónicos del documento

Objetivo de calidad:

- Posibilitar que el documento electrónico se visualice como un todo en la página HTML, facilitándose con ello el acceso al mismo.
- Garantizar, mediante el formato utilizado, la preservación digital del documento.

Resultados esperados:

- Todos los enlaces de la página Web funcionando correctamente.
- Información contenida en la página correcta y completa.
- El código generado en el diseño de la página HTML se apega, en la medida de lo posible, a las normas definidas en XHTML, para así garantizar la preservación digital de documento.

ASPECTOS GENERALES DEL LENGUAJE HTML

Las páginas HTML diseñadas, permiten informar al usuario sobre el contenido y datos generales de los documentos digitalizados. Facilitan su manejo y constituyen una forma atractiva de presentar los documentos tratados.

HTML (HyperText Markup Language) se traduce como: "Lenguaje de hipertexto basado en etiquetas o marcas", es un lenguaje que permite presentar información de forma estructurada y agradable, con enlaces (hyperlinks) que conducen a otros documentos o fuentes de información relacionadas, y con inserciones multimedia (gráficos, sonido...) La descripción se basa en especificar en el texto la estructura lógica del contenido (títulos, párrafos de texto normal, enumeraciones, definiciones, citas, etc) así como los diferentes efectos que se quieren dar (especificar los lugares del documento donde se debe poner cursiva, negrita, o un gráfico determinado).

El **XHTML (Extensible Hypertext Markup Language)**, es el nuevo lenguaje estándar generalizado para la Web. Proporciona una serie de normas o estándares, que permiten estructurar de una mejor manera un documento HTML, garantizando con esto la compatibilidad de este tipo de documentos con las distintas plataformas o browsers, es decir, que puedan ser leídos por cualquiera de estos dispositivos, en el presente y en el futuro.

La estructura básica de un documento HTML es la siguiente: comienza con la etiqueta **<html>**, y termina con la etiqueta **</html>**. Entre ambas, hay dos zonas bien diferenciadas:

- el encabezamiento, delimitado por **<head>** y **</head>**, que sirve para definir diversos valores válidos en todo el documento.
- el cuerpo, delimitado por **<body>** y **</body>**, donde reside la información del documento.

En resumen, la estructura básica de un documento HTML queda de la forma siguiente:

```
<html>
<head>
<title>Título</title>
</head>
<body>
Texto del documento, menciones a gráficos, enlaces, etc.
</body>
</html>
```

El ejemplo siguiente muestra una plantilla utilizada por el CRID:

Archivo Edición Formato Ver Ayuda

```

<html>
<head>
<title>Atención al ambiente en emergencias de origen volcánico.</title><meta htt
</head>
<body alink="#003366" link="#003366" bgcolor="#ffffff">
<br />
<table bordercolor="#ffffff" bgcolor="#deedfe" cellspacing="0" cellpadding="0" w
<tr valign="top">
  <td valign="top" width="145" bgcolor="#deedfe" height="192">
    <div align="center">
<br />
    
</td>
  <td valign="top" width="445" heig
<table width="95%" align="center" bor
<tr valign="top">
  <td width="20%" height="19"><font fac
Documento: </b></font></td>
  <td width="80%" height="19"><font fac
740</font></td>
</tr>
<tr valign="top">
  <td width="20%"><b><font face="Verdar
Título: </font></b></td>
  <td width="80%"><font face="verdana"
Atención al ambiente en emergencias d
<tr valign="top">
  <td width="20%"><b><font face="Verdar
Autor: </font></b></td>
  <td width="80%"><font face="verdana"
Colombia. Ministerio de Salud.<br />
Bogotá, CO; 1986</font></td>
</tr>
</table>
  <br />
  <br />
<table cellpadding="0" width="95%" a
<tr valign="top">
  <td>
<div align="right">

```

| | |
|---|---|
|  | <p>Documento: 740</p> <p>Título: Atención al ambiente en emergencias de origen volcánico.</p> <p>Autor: Colombia. Ministerio de Salud. Bogotá, CO; 1986</p> <p style="text-align: right; font-size: 0.8em;">Formato PDF</p> |
|---|---|

| Contenido | KB |
|---|-----|
| Portada | 44 |
| Índice | 58 |
| Introducción | 102 |
| Objetivos | 95 |
| Capítulo 1: Estructura orgánica | 233 |
| Capítulo 2: Plan de emergencias de origen volcánico | 790 |
| Capítulo 3: Recursos | 431 |

El diseño de páginas HTML requiere software adicional. Es posible utilizar un software especializado, como **Dreamweaver®** o **Frontpage®**. Ambos son editores visuales de páginas Web, de fácil manejo, que permiten crear y modificar páginas de una forma intuitiva, sin necesidad de saber programación. Presentan algunas diferencias que es necesario mencionar:

FrontPage sobrecarga el código fuente con etiquetas. Se ha comprobado que al generar una misma página con DreamWeaver y con FrontPage esto puede dar lugar a que esta varíe en espacio, ofreciendo al usuario una diferencia que luego se nota al cargar la página. Esto a la vez genera un código inútil que podría resultar problemático a la hora de crear una página Web puesto que, como es ampliamente conocido, las páginas deben ser creadas pensando en diferentes navegadores e incluso en diferentes plataformas como Linux, Apple, etc.

El interfaz de ambos programas es radicalmente diferente. En el DreamWeaver, muchos de los menús están de manera flotante, así como determinadas ayudas y referencias. En FrontPage, este interfaz se realiza de la manera habitual de Microsoft, es decir, mediante menús desplegables.

Por último, la adquisición del programa de software Dreamweaver es mucho mas costosa que FrontPage.

| Cuadro sinóptico. Proceso H: Diseño de paginas HTML | | |
|--|---|----------------|
| Tareas | | Páginas |
| H.1 | Seleccionar o diseñar una plantilla predefinida desde el programa de software. | 40 |
| H.2 | Llenar cada uno de los campos en la tabla de la pagina HTML | 41 |
| H.3 | Revisar cada uno de los enlaces del documento (links) y la información general. | 41 |
| H.4 | Revisar los aspectos estéticos de la página y el texto que la contiene | 41 |
| H.5 | Control de calidad en el código | 41 |
| H.6 | Guardar la pagina Web | 42 |

Tarea H.1 Seleccionar o diseñar una plantilla predefinida desde el programa de software.

Cualquiera de los dos programas permite hacer un diseño muy simple. Existe también un catálogo de plantillas HTML predefinidas, que puede utilizarse para este fin.