

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 (análogo) 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

Para llevar a cabo el proceso de digitalización de documentos, se necesita la participación de varias personas con diferentes perfiles:

- Analista de información: Persona encargada de seleccionar el material a ser digitalizado, debe tener criterios para evaluar aspectos como contenido, actualidad, relevancia, estado físico del documento, demanda de solicitud del documento.
- Digitalizadores: Personas encargadas de llevar a cabo el proceso de transformar un documento impreso a formato electrónico, realizando cada una de las etapas que conlleva este proceso.
- Evaluador de calidad o inspector de calidad: Persona encargada de evaluar el resultado del proceso de digitalización llevado a cabo por los digitalizadores. Evalúa aspectos como calidad, rendimiento en la producción, implementa métodos, políticas y estándares de calidad en busca de un mejoramiento continuo del proceso.
- El analista de información puede realizar dos funciones simultáneamente, ser analista de información y evaluador de calidad, o analista de información y digitalizador; lo que no se recomienda es que una misma persona realice las funciones de digitalizador y evaluador de calidad porque corre el riesgo de carecer de objetividad al evaluar los resultados obtenidos de su propio trabajo.

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 pixel un valor tonal: 0 para el negro y 1 para el blanco.

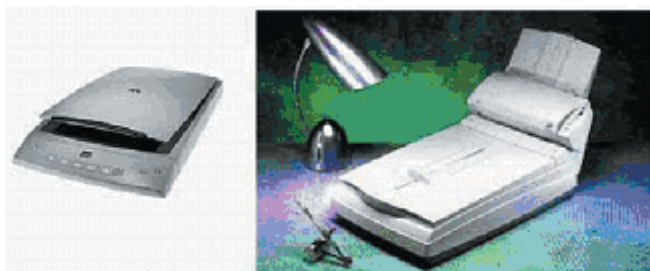
En la selección de un escáner hay que tomar en cuenta varios aspectos:

- 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
Ambos permiten escanear en blanco y negro o 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).

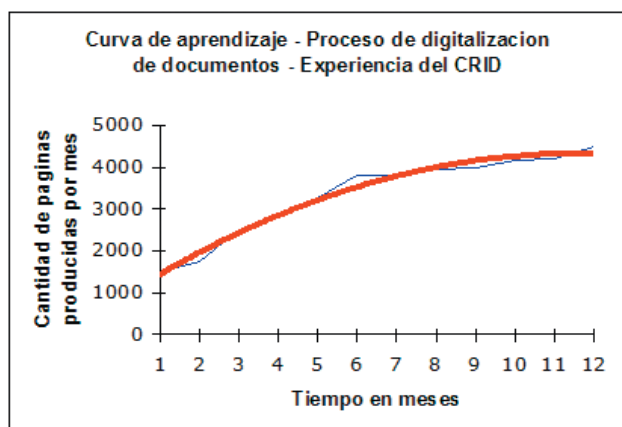
- Materiales: papelería, CDs, ZIPs entre otros.

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.