

# El reto de la Gestión Documental

---

## Introducción

Quizá la pregunta más habitual que nos hacemos al considerar soluciones de Gestión Documental sea “¿cómo puedo digitalizar la enorme cantidad de documentos que tengo en formato papel de una forma rápida y eficiente?”. Cualquiera que contemple la gestión de documentos o el archivo electrónico de papel debe enfrentarse a esta pregunta y debe determinar lo que mejor se adapte a sus circunstancias particulares.

Además de considerar los volúmenes de los archivadores de papel ya existentes, se debe planificar el mejor modo de procesar los documentos que se generan diariamente durante los procesos de negocio.

El propósito de este documento es el de introducir la Gestión Documental y proporcionar una guía sobre cómo enfrentarse a los procesos de digitalización. El proceso completo es conocido indistintamente como captura de documentos electrónicos o captura electrónica de documentos.

## ¿Por dónde empezar?

Existen muchos aspectos a considerar antes de comenzar a diseñar un proceso de digitalización, que podríamos dividir en cinco funciones primarias:

- ✓ Preparación de la documentación
- ✓ Digitalización
- ✓ Control de calidad
- ✓ Indexación o clasificación
- ✓ Migración a la solución de Gestión Documental

## Preparación de los documentos

Este proceso implica un análisis completo de los documentos existentes y de los procesos necesarios para que estos documentos se puedan digitalizar. Es el momento más crítico del proceso y el que más tiempo consume. La extracción de los documentos de las carpetas, la eliminación de grapas, clips, notas pegadas, post-it, y la limpieza general del documento se produce durante esta etapa. Una vez analizados los documentos a capturar, se deberían poder responder las siguientes preguntas:

1. **¿Cuál es el tamaño de los documentos? ¿Son todos DIN A4, o algunos son mayores o menores? ¿Están a simple cara o a doble cara? ¿Cuál es la proporción de cada tipo?**

Determinar estos aspectos ayuda en la elección de los dispositivos necesarios para llevar a cabo el trabajo: alimentador manual o automático, simple o doble cara, etc. El escáner elegido debería tener la capacidad de digitalizar todo su rango de tamaños soportados sin necesidad de cambiar los parámetros del escáner. Deben tenerse en cuenta escáneres especiales para necesidades especiales, como tarjetas de visita o grandes planos de AutoCAD.

2. **¿Cuáles son las características de los documentos? ¿Están en blanco y negro, o en color? ¿Tienen zonas resaltadas, marcadas, escritas a mano o con tinta?**

Estas preguntas ayudarán a decidir si se necesita un proceso de mejora de las imágenes por parte del propio escáner o si se necesita añadir algún software al mismo. Sin este proceso de mejora para los papeles de color, para las anotaciones manuales y otras características similares del documento, podrían generarse resultados de poca calidad e imágenes inútiles (demasiado oscuras o claras, con demasiados puntos, con poco contraste, etc.)

3. **¿Son todos los documentos del mismo tipo, como facturas, albaranes, órdenes de compra, o son de tipo mixto, como historias clínicas, ficheros de recursos humanos, ficheros de clientes, etc.?**

Esto ayudará a determinar el modo de clasificar los documentos antes del proceso de digitalización. Si fueran todos del mismo tipo, se pueden digitalizar por lotes mediante la categorización de documentos. Si son de tipo mixto, se tendrán que preparar para ser categorizados además a nivel de carpeta. Esta categorización es conocida como separación de documentos. A medida que se digitalizan los documentos, el software de captura debe determinar nuevas categorías y direccionarlos a carpetas en función de páginas separadoras. Para este cometido se utilizan páginas en blanco, páginas con códigos de barras o etiquetas, patrones comunes o números predeterminados de páginas.

4. **¿Sería necesario poder buscar por el texto de los documentos una vez almacenados? Si es así, ¿cuál es la calidad de los mismos? ¿Proviene de impresoras laser, impresoras matriciales o de máquinas fax?**

La búsqueda de texto dentro de una imagen digitalizada requiere de un proceso previo de OCR (**O**ptical **C**haracter **R**ecognition) o Reconocimiento Óptico de Caracteres, sobre la imagen. Si el documento requiere OCR, la calidad de la imagen será muy importante para obtener los resultados deseados. El proceso de OCR nunca alcanza una fiabilidad del 100% y su calidad depende en gran medida de las condiciones del documento original, de la resolución del escáner, del motor de OCR utilizado y del contenido del documento original. El OCR debería utilizarse siempre junto a un buen esquema de indexación, y nunca como el único argumento para localizar una imagen en el sistema de Gestión Documental.

## 5. ¿Cuántos documentos hay que digitalizar y con qué rapidez debe llevarse a cabo el trabajo?

El hecho de determinar correctamente el volumen de información a digitalizar ayudará a elegir el hardware adecuado, así como el número de personas necesarias para la preparación y finalización del proyecto en el periodo de tiempo establecido. Los escáneres se clasifican por el número de páginas por minuto que son capaces de digitalizar, de modo que este es un dato muy a tener en cuenta en el cálculo del tiempo. Si a este tiempo le sumamos el utilizado en preparar, clasificar e indexar los documentos, tendremos el tiempo total de finalización del proyecto.

## Indexación de los documentos

La indexación de los documentos es imprescindible para que su localización, una vez convertidos a formato electrónico, sea sencilla y efectiva. Los parámetros de indexación se pueden determinar una vez respondida la siguiente pregunta: ¿qué información utilizaría yo para buscar este documento rápidamente?

Simplificando mucho, se puede pensar en indexar un documento por un “Título”, y añadirle después otros campos clave para poder localizarlo de una forma sencilla dentro del sistema de Gestión Documental. La indexación también puede ayudar a la estandarización de los documentos dentro de la Organización, de modo que cada persona que necesite buscar algo sepa cómo conseguirlo gracias al acuerdo alcanzado para el esquema de nombres.

El software de captura debe incorporar opciones de indexación como introducción manual de los datos, desplegables para incorporar valores fijos, comunicación con bases de datos para rellenar campos de forma automática, OCR para la indexación automática de zonas de la imagen, lectura de códigos de barras y opciones de arrastrar/soltar desde la imagen.

Cada uno de estos métodos de indexación es beneficioso en determinadas circunstancias para conseguir la más rápida introducción de los valores a indexar. Cualquiera que sea el método que se decida utilizar, es muy importante encontrar el modo más eficiente de indexar los documentos, de modo que se puedan procesar tan rápidamente como sea posible.

## Digitalización del papel generado a diario

Todas las consideraciones mencionadas hasta ahora se pueden aplicar a las tareas de digitalización diarias. Sin embargo, es probable que requiera mucha menos preparación de los documentos que el archivo histórico.

Si el volumen diario es grande (>500 páginas), el método de digitalización más eficiente sería un proceso de captura de gran volumen. La digitalización diaria es también un proceso altamente dependiente de la fuente de los documentos. Si estos se generan por sistemas “propios” de la Organización, es posible mejorar el proceso de digitalización con códigos de

barras pre-impresos en los propios documentos. Esto permite al software de captura la indexación y clasificación de los documentos de forma automática, ahorrando una considerable cantidad de tiempo en la preparación y el indexado.

Para aquellos documentos que vienen de fuentes externas, se usará normalmente un indexado y clasificación manuales. Durante el proceso de estos tipos de documento es importante disponer de un esquema estandarizado de nombres. El uso de patrones de nombre y de listas desplegables asegura que los documentos se van a clasificar de forma correcta y consecuente, especialmente cuando hay varias personas implicadas en el proceso de digitalización y archivo.

El proceso de digitalización puede diseñarse para usar varios escáneres departamentales o centralizar la captura en un único punto con un escáner de alto volumen, como un dispositivo multifunción (MFP). Dependerá de los tipos de documentos, su tamaño, su forma de generarse, etc. En cambio, parece recomendable que el proceso de indexación se distribuya a nivel de cada puesto para clasificar y archivar los documentos en el sistema de Gestión Documental.

## **Preparación y ejemplos de indexación**

### **Ejemplo primero**

Nos encontramos con varios archivadores llenos de facturas antiguas. La compañía desea tener las viejas facturas digitalizadas lo más rápidamente posible. Después de revisarlas, se observa que todas las facturas fueron impresas en impresoras laser, y pueden ser de una o varias páginas. Todas son DINA4 y a una sola cara. Algunas de ellas tienen zonas resaltadas e información escrita a mano sobre sellos estampados.

A la compañía le gustaría poder buscar por el contenido de las facturas y prefiere el formato PDF. Se calcula que tienen unas 100.000 facturas en total y van a contratar temporalmente a algunos estudiantes de secundaria para digitalizar durante el verano. Los campos a indexar de cada factura deben ser: número de factura, número de proveedor, número de orden de compra y cantidad total.

Basándose en la calidad de las facturas y en las propiedades definidas anteriormente, la compañía selecciona un escáner de 50 páginas por minuto con procesamiento interno de imagen, para poder capturar el texto resaltado y el texto escrito a mano en los sellos. El software de captura seleccionado produce ficheros PDF con texto obtenido por OCR para que el sistema de Gestión Documental soporte las búsquedas por contenido. La información a indexar se encuentra ya en la base de datos de contabilidad, lo que permite que el usuario introduzca una cantidad mínima de datos, en este caso el número de factura, suficiente para ir a buscar el resto de los valores a la base de datos mencionada.

Antes de digitalizar, y para separar los documentos en facturas individuales, se coloca una etiqueta con un código de barras en la primera página de cada factura. El software encargado de digitalizar crea un nuevo documento cada vez que detecta un código de barras en el mismo momento de la captura. Se estima que la duración de este proyecto sería de unos 40 días de trabajo, a un ritmo de 2.500 facturas diarias.

## Ejemplo segundo

El departamento de Recursos Humanos de una gran empresa necesita digitalizar los expedientes de los empleados en la aplicación de Gestión Documental para facilitar su acceso a los trabajadores del departamento. Cada empleado tiene una carpeta de documentos en el sistema de ficheros del departamento de RRHH.

Después de revisar los documentos, se observa que hay varios tipos diferentes en cada carpeta de empleado. El tamaño de los documentos es generalmente DIN-A4, existiendo otros menores y también algunos en papel de color. Los documentos pueden clasificarse en seis categorías, con diferentes "Títulos de documento" para cada una de ellas. Cada categoría puede tener más de un título asociado.

Se decide utilizar el formato TIF Grupo IV y no dar la posibilidad de buscar por contenido. El volumen estimado es de 500 archivadores de empleados con una media de 50 páginas por carpeta. La compañía elige un escáner de 30 páginas por minuto con proceso integrado de la imagen para poder capturar eficientemente los documentos en color.

El proceso de preparación incluye eliminar grapas y clips para separar los documentos. Se crea un código de barras para cada uno de los seis tipos de documento, y se coloca en la primera página de cada uno de los tipos de documento que requiere un título. Se crea un separador de carpetas con páginas de papel reutilizables, que se usará como separador de cada carpeta de empleado. Si el nombre de carpeta no existe, este separador hará que se cree una carpeta automáticamente en la base de datos del Gestor Documental.

Se configuran también listas desplegables para indexar las seis categorías en las que se han dividido los títulos de los documentos. Esto permitirá a los usuarios elegir el título de la lista sin necesidad de teclearlo. Además de aumentar la velocidad de indexación, esta configuración asegura que todos los documentos siguen un esquema fijo de indexación independiente del operador que haga el trabajo. La compañía tiene poco personal disponible para este proceso.

Se estimó un tiempo aproximado de 2 meses para digitalizar los documentos. Este plazo fue cumplido con los parámetros seleccionados.

## Conclusión

La clave para implementar con éxito una solución de digitalización comienza en la etapa de planificación. Son aspectos vitales del proceso: el conocimiento de los tipos de documentos a digitalizar, los requerimientos de hardware, los métodos de indexación disponibles y un plazo de tiempo realista para llevar a cabo la conversión del papel a ficheros digitales.

Si no se revisa minuciosamente el archivo físico de papel y no se entiende el nivel de esfuerzo necesario para preparar los documentos, los resultados pueden ser decepcionantes. Mediante una planificación cuidadosa y la ejecución de ensayos del proceso completo, pueden cumplirse razonablemente las expectativas marcadas. Y recuerde: si necesita apoyo, siempre habrá profesionales dispuestos a ayudarle con el proceso completo.