

# Cómo digitalizar tu archivo de documentos sobre derechos humanos

<b>¿Qué es digitalizar?</b>	1
<b>¿Por qué digitalizar?</b>	2
<b>Fases de un proyecto de digitalización</b>	3
Definición de metas y alcance	3
Diagnóstico del archivo	4
Acceso y utilización del archivo	5
Estimación de recursos humanos, temporales y equipamiento	7
Digitalización, almacenamiento e indizado	9
Control de calidad y medidas de preservación digital	11

## ¿Qué es digitalizar?

Digitalizar es convertir documentos y archivos en soporte no electrónico (papel, película o fotografía) o soporte electrónico analógico (VHS, casete) en formato digital. El ejemplo de digitalización más frecuente, al que nos referimos en este documento, es el escaneado de documentos en papel.

Las organizaciones con una larga trayectoria en el campo de los derechos humanos suelen acumular documentos a lo largo de los años, mayoritariamente en papel, aunque también archivos fotográficos e incluso películas y diapositivas.

Existen múltiples razones para plantearse la digitalización total o parcial de estos documentos. En muchas ocasiones estos documentos ya han cumplido su función y se guardan físicamente para dar cuenta de la trayectoria de la institución, o bien se desea preservar documentación de relevancia que corre riesgo de destrucción si se mantiene en formato físico; también para posibilitar su sistematización en un sistema de datos, favorecer su accesibilidad de forma remota o como un modo de asegurar y contar con copias y proteger información en riesgo.

## ¿Por qué digitalizar?

Los principales motivos por los que digitalizar una colección son los siguientes:

- a) Mejorar la accesibilidad: los documentos digitales publicados en un repositorio de acceso abierto pueden consultarse en cualquier momento y desde cualquier lugar. La digitalización también facilita la forma de compartir documentos de forma privada y por cauces seguros.

Más allá de la ruptura de las barreras espaciales y temporales en el acceso, la digitalización suele conllevar una descripción exhaustiva del contexto y el contenido del documento digital, lo que aumenta las opciones de búsqueda.

- b) Incrementar la seguridad: si tus documentos físicos contienen información sensible existe el riesgo de que alguien irrumpa en tus instalaciones y los confisque o destruya. Hacer copias digitales evita que la información almacenada se pierda para siempre. A la hora de evitar el acceso a información sensible por parte de terceros, el formato digital permite adoptar medidas de seguridad para blindar el acceso a los documentos.
- c) Ahorrar espacio: los documentos en papel ocupan mucho espacio en comparación con el almacenamiento digital. La escasez de espacio suele ser un problema recurrente para las organizaciones, sobre todo cuando deciden almacenar toda la documentación que pasa por sus manos, sin aplicar criterios de selección.

- d) Reducir amenazas propias de la conservación física: los documentos en papel están expuestos a distintos tipos de amenazas: humedad, temperatura, insectos, hongos, pérdidas al ser manipulados por personas, inundaciones, incendios, etc. Al realizar una copia digital aseguramos que la información que contienen no desaparecerá, aunque la conservación digital también se enfrenta a amenazas, como veremos más adelante.

## Fases de un proyecto de digitalización

### 1. Definición de metas y alcance

Todo proyecto arranca con la definición de metas y lo más lógico es que las de un proyecto de digitalización estén relacionadas con alguna o varias de las razones expuestas en el apartado anterior. Las metas deben quedar plasmadas de forma clara, ya que serán la referencia para medir posteriormente el éxito del proyecto.

Al definir las metas de un proyecto de digitalización conviene ir más allá del propio proceso de escaneo y pensar en cómo se consultarán y aprovecharán los documentos digitalizados, ya que de ello dependerá en buena medida el diseño del proyecto.

En lo que respecta al alcance, es importante preguntarse qué se quiere digitalizar y por qué, ya que los recursos casi siempre son escasos. No es necesario digitalizar el 100% del archivo. Dependiendo del propósito, estos son algunos ejemplos de materiales que te puede interesar digitalizar:

- Si vas a iniciar un procedimiento legal, los documentos que puedan tener valor como evidencia.
- Si lo haces por razones de seguridad, aquellos documentos que contengan información sensible.
- Si tu objetivo es preservar la información, los documentos más deteriorados, amenazados por la humedad o ubicados en una sala con riesgo de inundación.

## 2. Diagnóstico del archivo

Para tomar decisiones adecuadas es necesario contar con un diagnóstico previo que evalúe, al menos, las siguientes dimensiones:

- a) Cantidad: en ocasiones, el volumen de documentos es tan grande que no permite tener una idea clara a simple vista. Existen fórmulas para realizar estimaciones basadas en el número de cajas o estanterías en el almacén. Si no podemos contar todos los materiales, al menos debería ser posible averiguar el número de documentos de varias cajas o varias estanterías para obtener una media y después multiplicar por el número total. Lo ideal para nuestro propósito es realizar la estimación total en páginas (especificando si por una o dos caras), ya que esta es la medida en la que basaremos muchos de los cálculos del proyecto.
- b) Estado de conservación: hay factores ambientales (humedad, temperatura, luz) o biológicos (hongos, insectos, roedores) que pueden afectar a la conservación de los documentos en papel. Es importante que revise una muestra relevante del archivo (documentos ubicados en distintas zonas, a diferentes alturas, en distintas salas, etc.) para tener una idea del estado de conservación de los documentos, ya que esto puede hacer que priorice algunos materiales sobre el resto o que necesite equipamientos especiales. Estas condiciones también afectan a la tarea de preparación de documentos que se lleva a cabo antes de pasarlos por el escáner.
- c) Tipología de los documentos: un archivo relacionado con los derechos humanos puede contener documentos legales, fotografías, documentos escritos a mano o a máquina, transcripciones de entrevistas y testimonios, libros, investigaciones, recortes de prensa, etc. No es lo mismo digitalizar una fotografía en la que necesitamos apreciar un elemento con gran nivel de detalle que digitalizar un documento escrito cuyo único requisito es que sea legible. Además, la tipología suele estar relacionada con el soporte (papel fotográfico, papel de celulosa, película, etc.) y con las dimensiones del documento que se va a digitalizar (por ejemplo, un mapa o un periódico suelen ser más grandes que un documento legal o una fotografía).
- d) Sistema de clasificación: ¿cómo están organizados actualmente los documentos? Podrían estar ordenados cronológicamente, por temática, por zona geográfica a la que hacen referencia, por investigaciones, etc. Entender cómo está organizado el archivo proporciona información sobre qué contiene y anticipa muchas de las

decisiones que luego tomaremos para facilitar el acceso a los documentos digitales. Si los documentos de tu archivo están organizados por país, está claro que ese será uno de los criterios de búsqueda una vez que estén digitalizados. Sin embargo, la gran ventaja del formato digital es que podremos añadir fácilmente otros criterios de búsqueda que ayuden a localizar lo que se desea.

- e) Propiedad de los documentos: ¿Quiénes son los autores y propietarios de los documentos? La organización que custodia los documentos no tiene por qué ser la propietaria de los mismos, o puede no contar con derechos para su publicación (su permiso puede estar limitado al uso interno, por ejemplo). También es necesario considerar si hay cláusulas de confidencialidad o similares.
- f) Necesidades de retención: hay ciertos tipos de documento que, por ley, deben ser guardados durante cierto tiempo. Esta circunstancia suele darse, sobre todo en el campo legal y en el financiero. Para evitar problemas, conviene revisar la legislación nacional para estar al tanto de estas condiciones.

Este diagnóstico será la base del diseño del proyecto de digitalización, por lo que no conviene escatimar tiempo y otros recursos en él. Para tener una idea clara de todos los aspectos mencionados, lo más útil es investigar directamente en el archivo, tomando muestras representativas. Los resultados de la evaluación deben quedar plasmados en un documento claro y conciso, así evitaremos tener que regresar al archivo para valorarlo de nuevo.

### 3. Acceso y utilización del archivo

Aunque todavía estamos diseñando el proyecto y no hemos comenzado a escanear, podemos plantear cómo se accederá a los documentos digitalizados y cuál será su uso. Las principales opciones al respecto consisten en publicar la colección digitalizada en una plataforma abierta al público o mantenerla para uso interno, con envíos selectivos a otras personas u organizaciones cuando sea necesario.

Tanto si vamos a publicar la colección o mantenerla para uso interno, UWAZI, la plataforma de código abierto diseñada por HURIDOCs, permite organizar y analizar colecciones de documentos. Con UWAZI podremos describir documentos mediante metadatos personalizados para hacerlos fácilmente recuperables a través de su interfaz. Además, la plataforma posee las características necesarias para mantener altos estándares de

seguridad y se puede adaptar para distintos propósitos, por lo que constituye una solución ideal para almacenar y publicar las colecciones digitalizadas.

En caso de que los costes de mantenimiento de la colección sean demasiado elevados, se puede considerar delegar su custodia en otra institución que se encargue de incorporarlos a su archivo y ponerlos a disposición del público. Existen varias instituciones internacionales que custodian archivos relacionados con derechos humanos, aunque todas tienen ciertos requisitos a la hora de incorporar colecciones que hay que revisar de antemano. Estas son algunas de las instituciones a las que nos podríamos remitir:

- The Open Society Archives (<https://www.osaarchivum.org/>)
- Duke University Human Rights Archive (<https://library.duke.edu/rubenstein/human-rights>)
- Human Rights Web Archive at Columbia University (<https://www.archive-it.org/collections/1068>)
- University of Texas Libraries' Human Rights Documentation Initiative (<https://legacy.lib.utexas.edu/hrdi>)

Con independencia del destino de la colección, conviene plantearse cómo vamos a estructurar la información y, de forma más específica, como vamos a nombrar los archivos generados, ya que es un error frecuente que al finalizar la tarea de escaneo no sepamos a qué corresponde cada fichero generado. Algunos consejos básicos para la nomenclatura de archivos serían:

- Incluir el tipo de archivo o qué contienen: *declaración testigo, recorte periódico, sentencia, etc.*
- Incluir la fecha, preferentemente en formato año-mes-día para que los archivos se ordenen cronológicamente. Por ejemplo *20191117*.
- Incluir el lugar geográfico al que se refieren: país, departamento, municipio, etc.
- Usar números secuenciales o fechas para documentos del mismo tipo: *sentencia\_001, sentencia\_002* o *sentencia\_20190212, sentencia\_20190421*.

- Combinar varios de estos criterios y otros que se ajusten a la colección, pensando en los métodos más comunes para encontrar los documentos a posteriori.
- Nombrar los archivos de forma sistemática una vez que se haya establecido el criterio.

Las mismas pautas pueden aplicarse a la hora de nombrar carpetas, aunque evitando las redundancias con el nombre del archivo. La jerarquía de carpetas permite estructurar los archivos en varios niveles. Por ejemplo, si vamos a seguir el criterio geográfico, un primer nivel sería el continente, el segundo el país, el tercero el departamento, y así sucesivamente. Si fuese cronológico, estructuraríamos por década, año, mes, etc.

Las consideraciones de esta etapa, relacionadas con las necesidades de los usuarios y con la estructura de la información, son decisivas cuando la digitalización se integra en un proyecto más amplio que abarca el diseño de un sistema de información.

## 4. Estimación de recursos humanos, temporales y equipamiento

El diagnóstico previo te permitirá tener una idea certera de los recursos necesarios para acometer el proyecto.

Tiempo. Para calcularlo, es necesario tener claro el volumen total de documentos a digitalizar y el tiempo medio empleado para cada documento, considerando la preparación del mismo (extraerlo de las cajas o archivadores, quitar grapas, estirar dobleces, etc.), el escaneo, la reubicación en el lugar de origen y la descripción del documento digital para su almacenamiento.

Personal. El tiempo total del proyecto dependerá también del número de personas implicadas, pese a que existen cuellos de botella que hacen que la reducción de tiempo no sea proporcional al número de personas contratadas.

Por ejemplo, si la duración total del proyecto se estima en 300 horas de trabajo para una persona, la contratación de otra persona no implica que la duración se reduzca a la mitad (150 horas), ya que muchas de las tareas no pueden llevarse a cabo de forma simultánea y hay tiempos, como el del escaneo, que no dependen del personal.

A la hora de contratar personal adicional o de calcular cuántas personas de la organización se involucrarán en el proyecto es necesario contar con el diseño completo flujo de trabajo. Esto también nos permitirá determinar si se requieren perfiles especializados.

El personal que vaya a participar en el proyecto debe estar capacitado para las tareas que le serán asignadas, en especial en lo que respecta al manejo de las herramientas tecnológicas con las que se llevará a cabo el proyecto, que suelen quedar fuera del trabajo cotidiano en las organizaciones de derechos humanos.

### Equipamiento:

El escáner es un equipamiento indispensable en cualquier proyecto de digitalización. Dependiendo de la envergadura del proyecto y de la urgencia del mismo se puede considerar adquirir o alquilar más de un equipo. Para escoger el modelo que mejor se adapta a tus necesidades debes considerar, al menos, los siguientes aspectos:

- Velocidad: suele calcularse en páginas por minuto. Para tener una idea de qué es un escáner rápido y qué es un escáner lento lo mejor es comparar modelos, aunque un buen punto de partida suele ser una velocidad de 30 páginas por minuto.
- Tamaños admitidos: la evaluación del archivo debería haber servido para identificar si hay materiales con dimensiones que escapan de lo convencional, como mapas o fotos de gran tamaño. Los dispositivos para grandes formatos suelen ser mucho más caros que los convencionales.
- Bandeja alimentadora y posibilidad de doble cara: para calcular el volumen de trabajo necesitas saber cuántas páginas se pueden colocar a la vez en la bandeja desde la cual los documentos pasan al escáner, y si el dispositivo escanea por las dos caras de forma automática. De lo contrario, tendrá que haber una persona permanentemente para escanear por ambos lados.
- Resolución óptica: la resolución se mide en número de píxeles por pulgada (ppp) y determina el nivel de detalle con el que se puede observar una imagen sin perder calidad al hacer zoom. Si nuestro propósito al digitalizar es, simplemente, que el documento se pueda leer, no necesitaremos mucha resolución. Por contra, si estamos escaneando imágenes a color



necesitaremos mayor resolución y, por tanto, tendremos que adquirir un escáner más caro.

- Formatos de salida: lo más común es generar archivos PDF para documentos escritos y JPEG o PNG para imágenes, aunque en caso de que queramos conservar una copia de las fotografías con la máxima calidad posible (y por tanto mayor tamaño), lo ideal es el formato TIFF. La mayoría de los escáneres del mercado admite estos formatos.

OCR son las siglas en inglés de Optical Character Recognition. Se trata de un software que permite reconocer los caracteres en una imagen y procesarlos como texto, de forma que se puedan hacer búsquedas y otros análisis sobre el documento. Esto es especialmente útil para encontrar la información que queremos en un documento escaneado. Estas herramientas suelen tener un coste elevado, aunque también las hay gratuitas y de código abierto. Este artículo del proyecto *Source de Open News* (en inglés) reseña algunas de estas herramientas: [Our Search for the Best OCR Tool, and What We Found](#).

El almacenamiento digital tiene costes, tanto si lo hacemos offline (en un disco duro en nuestras oficinas, por ejemplo) como en la nube. Para calcular el almacenamiento necesario podemos multiplicar el tamaño medio de un archivo escaneado por el total de archivos.

Una vez que hayamos estimado los recursos, es recomendable reflejar todos estos rubros en un presupuesto completo, añadiendo también un apartado de costes imprevistos.

## 5. Digitalización, almacenamiento e indizado

Ahora ya tenemos todas las especificaciones del proyecto, por lo que podemos empezar con el proceso de escaneado y todas las tareas que lleva asociado. No obstante, para terminar de ajustar las estimaciones realizadas y anticipar imprevistos, conviene realizar una prueba con una muestra del archivo. Una vez que todo esté listo, podemos empezar paso a paso:

1. Mantener un inventario. Los desplazamientos de materiales van a ser constantes durante el proceso y, manejando grandes volúmenes de documentos, corremos el riesgo de que alguno acabe extraviado. Reflejar el flujo de documentos con un inventario sencillo que muestre cuándo los materiales se sacaron y se devolvieron a su lugar de origen puede ser de gran ayuda.
2. Localizar los archivos físicos en su ubicación, sacarlos de las cajas o archivadores (si procede) y llevarlos a la mesa de trabajo.
3. Preparar los materiales: quitar grapas, sacar de fundas acrílicas, deshacer dobleces, limpiar polvo, desencuadernar, etc.
4. Cargar la bandeja alimentadora. Antes de cargar, revisa si los materiales deben ser escaneados por una o por ambas caras y configura el equipo para ello. Verifica también la carga máxima de la bandeja alimentadora y la posición correcta de los materiales.
5. Escaneado. Decide si vas a realizar un escaneo página a página (cada una generará un archivo independiente) o de un documento completo (un solo archivo con varias páginas). Comprueba una vez más que los parámetros del escáner son los correctos para el tipo de documento.
6. Nombre y descripción del archivo digital mediante metadatos. Cuando el documento ha sido escaneado y figura en la computadora es el momento de nombrarlo y describirlo según los criterios establecidos. En caso de que el archivo físico y digital se vayan a mantener, es frecuente que uno de los campos de descripción del archivo digital incluya una referencia a la ubicación del archivo físico original. Este paso es previo a la carga de los documentos digitalizados en una base de datos, que corresponde a una etapa posterior.
7. Recolocación del material físico en su lugar de origen. Si pretendes conservar tus archivos a largo plazo, es mejor no volver a colocar grapas y clips metálicos, ya que con el tiempo se oxidan y dañan el papel. Cuando los materiales hayan sido devueltos, asegúrate de indicarlo en el inventario.

## 6. Control de calidad y medidas de preservación digital

Pese a que no podemos tener un control absoluto de los materiales digitalizados, es importante realizar un seguimiento que nos ayude a detectar desviaciones durante el proceso. Con él verificaremos si los pronósticos (sobre todo temporales y de recursos) se están cumpliendo y si los resultados de digitalización son los esperados.

El control de calidad sirve para conocer el grado de cumplimiento de los requisitos previamente establecidos. Un documento mal escaneado (sin la resolución suficiente, con las páginas inclinadas, demasiado oscuro, etc.) puede provocar, por ejemplo, que el software de OCR sea incapaz de reconocer los caracteres. Si no detectamos el problema a tiempo, habremos perdido gran cantidad de recursos.

Lo mismo sucede a la hora de asignar metadatos o nombrar carpetas y archivos. Si no asignamos metadatos obligatorios en el momento que corresponde (por ejemplo, la referencia a la ubicación del archivo físico), la tarea se puede volver mucho más ardua a posteriori.

En general, todas las fases del proyecto deben cumplir ciertos parámetros de calidad concretados en indicadores y rangos de valores admitidos. Durante el escaneado, el control de calidad se basa en visualizaciones periódicas de los ficheros generados, para asegurar que no presentan problemas o variaciones significativas con respecto al documento físico original.

Por último, antes de cerrar un proyecto hay que valorar las amenazas para la conservación de los documentos digitales. Ni la tecnología ni los soportes son para siempre: la obsolescencia del formato puede hacer que un archivo quede inutilizable y un disco duro dañado puede hacer que perdamos información valiosa.

Aunque la preservación digital es un campo de trabajo muy amplio que queda fuera del alcance de este documento, enumeramos algunos consejos para mitigar estas amenazas:

- Utiliza formatos interoperables y extendidos. Estos son los recomendados por la Biblioteca del Congreso de los Estados Unidos:  
<https://www.loc.gov/preservation/resources/rfs/TOC.html>
- Actualiza los formatos cuando identifiques que se están quedando obsoletos.

- Realiza copias de seguridad siguiendo una estrategia predefinida. Este artículo del blog de Huridocs (en inglés) puede ayudarte:  
<https://www.huridocs.org/2016/12/fridaydiscobackup-and-other-pro-backup-tips/>
- Revisa el estado de tus soportes de almacenamiento (pen drives, discos duros, etc.) y renuévalos periódicamente.