

Guías para la digitalización de los materiales de archivos

De O.S.G., Box 459, Grand Central Station, New York, NY 10163

INTRODUCCIÓN

Vivimos en un mundo de constante evolución tecnológica. Un creciente número de archivos y bibliotecas están haciendo uso de repositorios digitales y experimentando con la tecnología, prácticas que no siempre dan los resultados deseados. Puede ser un reto crear un repositorio digital, en parte porque implica una serie de asuntos y problemas divergentes. Si estás pensando en hacer una biblioteca digital, es probable que tendrás que abordar cuestiones tales como la metodología para crear un repositorio digital, los gastos involucrados y la capacidad para tomar las decisiones sobre qué materiales digitalizar.

Además tendrás que conocer los criterios técnicos para manejar los diversos materiales en tu colección, tales como documentos escritos a mano, fotografías, revistas, libros, materiales de audio, películas y otros. La seguridad del repositorio, acceso y uso de los materiales, servidores y el sistema de respaldo del contenido (o *backup*) son otros asuntos que habría que atender.

Dado que con regularidad se nos consulta en los Archivos de la Oficina de Servicios Generales sobre la digitalización de las colecciones de los Archivos de A.A. locales, decidimos responder aquí a algunas de las interrogantes. Estas directrices tienen un propósito informativo. No abordan todo lo que se sabe sobre cómo mantener un repositorio digital, pero constituyen un buen punto de partida. Ninguna base de datos, sin importar su tamaño, puede digitalizar toda su colección. La digitalización puede ser costosa y requiere muchísimo tiempo, y algunos materiales puede que no se presten a la digitalización o que no valgan la pena digitalizar. No todos los archivos de área o de distrito podrán afrontar el costo de la digitalización, pero eso está bien.

CÓMO MANTENER LA PRIVACIDAD Y EL ANONIMATO

Una cuestión fundamental que hay que tener en cuenta al mantener archivos digitales es la protección de la privacidad y el anonimato de los miembros de A.A. Los procedimientos que se deben seguir parecen ser cuestión de la política local de los comités de archivistas, pero la necesidad de proteger la confidencialidad de la correspondencia y el anonimato de los correspondientes es, sin duda alguna, una consideración importante y una

confianza depositada en todos los archivistas de A.A. y en los comités de archivos. Podrías pensar en la implementación de una política para el uso de los materiales digitalizados e incluir directrices sobre la difusión de los materiales por la red. Para más información sobre el anonimato en línea consulta "Las Guías de A.A. sobre el Internet." (SMG – 18).

¿QUÉ ES LA DIGITALIZACIÓN?

La digitalización es el proceso de convertir información, tales como textos, fotografías o archivos de sonido o video, en formato digital. Permite preservar el contenido del material mediante la creación de una accesible fuente alternativa, y salvaguardar el material original del perjudicial uso ulterior. Luego se mantienen los objetos digitales en un repositorio digital que presta un medio útil de almacenar, mantener, acceder a las fuentes alternativas y preservarlas.

Fundamentalmente, los repositorios digitales tienen tres funciones. La primera es adquirir o captar información digital. La segunda es almacenar y mantener la información digital. La tercera función conlleva acceder a la información digital y crear materiales entregables (es decir, lo que se puede hacer con el contenido digital). Estas funciones requieren equipos físicos y programas de computación (o sea, hardware y software). Su costo puede fluctuar entre ningún costo y muy costoso, según el hardware o software que se escoja.

CONSIDERACIONES BÁSICAS AL PLANIFICAR TU PROYECTO

Antes de crear tu base de datos digital debes tener en cuenta las siguientes cuestiones: ¿Cuáles son tus objetivos? ¿Cuáles de los objetivos se ajustan más a la misión de tus archivos? Por ejemplo, algunas razones para digitalizar pueden incluir mayor acceso y uso, la seguridad, la preservación, el mantenimiento y la autenticidad. En los Archivos Históricos de la Oficina de Servicios Generales nuestras iniciativas para digitalizar han sido beneficiosas en cuanto al acceso a la información, además de mejorar la preservación de las piezas originales al reducir su uso físico. Tómate el tiempo que necesites para decidir cuál equipo de digitalización conviene para tu re-

positorio. Se debe hacer la investigación básica del hardware y software antes de adquirirlos. Estudia los peligros potenciales para el repositorio, tales como los medios de almacenamiento, el hardware, el software y/o cortes de luz. Algunos puntos que hay que tener en cuenta para evitar consecuencias desastrosas son los siguientes:

- ¿Podrían los archivos sostener a largo plazo el costo del hardware y software necesarios?
- ¿Qué medios utilizarías para almacenar los datos?
- ¿Qué sistema de respaldo utilizarías para tus dispositivos de almacenamiento?
- ¿Has pensado en la obsolescencia del hardware y software?
- ¿Cómo controlarías el acceso a los datos?
- ¿Cómo protegerías los datos del malware electrónico, los virus, etc.?

Los requisitos de software:

Primero, escoge la aplicación para escanear que mejor cumpla con tus necesidades. Para colecciones locales pequeñas, se puede usar aplicaciones sencillas basadas en Windows o MAC. Se recomienda el sistema de gestión de contenido (y gestión de archivos) para una colección apreciable, os sea, de por lo menos 5.000 imágenes.

Es importante tener buenas herramientas de software para el escaneo, la edición de archivos y la funcionalidad de archivos PDF. Muchos proveedores de escáneres incluyen el programa controlador para el dispositivo de escaneo con el software. Hay software de código abierto en abundancia que se puede descargar del Internet. A menudo, muchos vendedores de software y hardware ofrecen sus productos con descuento a entidades sin fines de lucro.

Si un programa es de código abierto, el código fuente les es disponible gratuitamente a los usuarios. No hay que pagar derechos de licencia ni hay otras restricciones sobre el software. Por lo tanto, un usuario puede tomar el código fuente, modificarlo y pasárselo a otros. Los programas de código abierto raras veces vienen acompañados de asistencia técnica, y los usuarios dependen de una comunidad en línea para orientarse y recibir ayuda. Se requiere conocimiento técnico para utilizar el software de código abierto. Por otra parte, el software de código cerrado y patentado consiste en programas distribuidos a un costo por una marca fiable. Los derechos de licencia del software patentado incluyen restricciones que

excluyen la modificación y redistribución del producto. El software propietario ofrece apoyo técnico, que por lo general es un recurso positivo para los usuarios que carecen de destrezas técnicas avanzadas. Sin importar el tipo de software que uses, se requerirán ciertas modificaciones y un profesional de TI para iniciar y echar a andar tu sistema. Es importante estudiar bien los términos y condiciones de la licencia del software antes de firmar un acuerdo. Un factor clave que se debe tener en cuenta es el grado de control que tú retienes en cuanto al acceso y la edición de tus datos cuando se venza la licencia del software.

Los requisitos de hardware:

Normalmente los requisitos básicos de hardware incluyen una computadora (preferiblemente con alta velocidad de procesamiento), un monitor de alta calidad, una unidad de CD-ROM, un disco duro con 40 GB (o más) de espacio, una impresora láser a color y en blanco y negro, un escáner que pueda reproducir fotografías y, como opción, un escáner de mano. La decisión sobre cuáles materiales se van a escanear (texto, fotografía, obra de arte) pesa sobre la decisión respecto al equipo y el tamaño del equipo que se debe adquirir. Se recomiendan los escáneres de mesa para la mayoría de los proyectos de digitalización, mientras que el escáner de mano tiene su uso en los casos de piezas muy grandes, materiales frágiles y tomos encuadernados.

Los medios óptimos para el almacenamiento de los ficheros maestros probablemente son los sistemas de disco duro. Es importante indicar que poner los ficheros maestros en CD-R, DVD-R, Blu-Ray y en particular en una memoria flash a fin de tener fácil acceso a ellos no les asegura larga vida. Sin embargo, si se almacenan materiales en los CD-R y los DVD-R, lo aconsejable es utilizar los CD-R y DVD-R de alta calidad o de calidad archivística (tales como Mitsui Gold Archive CD-Rs) y almacenar estos discos en un entorno con temperatura y humedad controladas.

Las tecnologías de software y hardware exigen atención en todo momento, dado los continuos y rápidos avances tecnológicos. Por lo general, cuando emerge una nueva tecnología de información digital o sistema computerizado, ésta desplaza la versión antigua en poco tiempo, y cuando se abandona una tecnología de software o se deja de fabricar un dispositivo de hardware, los archivos digitales hechos con esa tecnología corren

el riesgo de perderse. A esto se le llama obsolescencia digital. Es extraño que hoy día todavía se pueda leer en su idioma original documentos centenarios, y sin embargo, en bastante poco tiempo, la tecnología de los medios digitales ha evolucionado del disquete floppy al disquete, los CD, los DVD, los discos Blu-Ray y los medios de memoria flash. Recuerda que cualquier tipo de almacenamiento puede fallar en cualquier momento. Almacena tus datos en más de un medio y revisalos con regularidad para estar al tanto de cualquier problema.

Debes tener cuidado con “las compañías de la nube” que parecen ofrecer una opción ideal para el almacenamiento de copias de respaldo. El sistema de la nube requiere mandarle por la red una copia de los datos a un servidor externo. El servidor suele ser un proveedor de servicios de terceros, que cobra una tarifa según el ancho de banda, la capacidad o el número de usuarios. La mayor parte de estas compañías toman control por completo de los datos que uno deposita bajo su custodia. Antes de entrar en la nube, analiza cuidadosamente los términos y condiciones del contrato y ten en cuenta los siguientes puntos de importancia:

- Cuestiones de seguridad. ¿Qué seguros están tus datos?
- ¿Qué pasa si decides cancelar tu servicio? ¿Puede quedarse la compañía de la nube con una copia de tus datos?
- ¿Qué pasa si la compañía quiebra y vende tus datos?
- La inflexibilidad. Asegura que no hayas puesto tus datos irremediamente en una aplicación o un formato patentado.

LA SELECCIÓN DE MATERIALES ORIGINALES PARA DIGITALIZAR

Muchos archivistas enfrentan la abrumadora interrogante de qué digitalizar. Algunos se informan de los esfuerzos por digitalizar de otros repositorios y piensan que ellos mismos no están haciendo lo debido. Puedes estar seguro de que tu trabajo por preservar la historia de A.A. es un buen punto de partida. No podemos decirte directamente qué debes digitalizar porque el alcance de tu colección es singular.

Debes ser selectivo al determinar qué piezas vas a digitalizar, concentrándote en la calidad y no la cantidad. También recomendamos que te centres en la digi-

talización de materiales pertinentes a tu área o distrito, tales como las actas, volantes, correspondencia y otros artefactos producidos por los grupos de A.A. locales. Al digitalizarlos, obtienes un duplicado electrónico, fácil de acceder y utilizar, a la vez que prescindes del uso excesivo de los originales.

Algunos objetos “nacen” digitales, o sea, se crean en forma digital, por ejemplo ediciones electrónicas, cámaras digitales, etc. Sin embargo, los debes importar a tu depósito digital y guardar con nombres de ficheros afines con el sistema de identificar ficheros que hayas establecido para objetos digitales. Se pueden digitalizar los textos siempre y cuando el proceso de digitalización no les cause daños.

A los documentos que no caben en un escáner plano convencional se les llama documentos de gran tamaño. Se deben digitalizar tales piezas, al igual que los libros encuadernados, con un escáner de mano, una cámara digital para escanear o una cámara digital convencional. No es aconsejable escanear tomos encuadernados en un escáner plano ya que puede causar daños permanentes en el lomo del libro y en su encuadernación.

CÓMO ARMAR LA ESTRUCTURA PARA LOS DATOS

Después de haber abordado estas cuestiones básicas, debes formular un plan para organizar los datos de forma inteligente. También es el momento para determinar el mecanismo para ponerles nombres de ficheros a las carpetas y ficheros. En aras de la coherencia, se debe dar este paso antes de iniciar la tarea de digitalización. Los nombres de ficheros descriptivos deben asociarse con el objeto digitalizado y pueden incluir el nombre del objeto y el número de registro. Considera la convención de nomenclatura de ficheros físicos de tu colección como modelo para tu repositorio digital.

Hay una página en el *Libro de trabajo de Archivos Históricos* titulada “Calendario de la Colección” de los Archivos de la OSG de A.A. Es un sistema sofisticado de organización y clasificación de los archivos de la OSG de A.A. Se debe emplear el mismo tipo de sistema eficaz para organizar el contenido de los repositorios digitales.

A continuación se da un ejemplo del sistema de ordenar carpetas en la base de datos electrónica de los Archivos de la OSG:

Boletines (Carpeta principal)

➔ *Acerca de A.A. (subcarpeta)*

- *Acerca de A.A., Verano 1972* (nombre de fichero)

➔ *Box 4-5-9 (subcarpeta)*

- *Box 4-5-9, Otoño 2010* (nombre de fichero)

LOS METADATOS

Cualquier individuo inmerso en un proyecto de escaneo se topará con la palabra “metadatos.” Dicho de otro modo, los metadatos constituyen “información sobre la información.” Es información que se emplea para describir, localizar y acceder a los ficheros almacenados en una biblioteca digital. Es la clave para asegurar que haya fácil acceso a las fuentes electrónicas en el futuro. En Microsoft Windows y demás software, los metadatos también se conocen por el nombre de “propiedades.” Ambos términos tienen el mismo significado.

La mayor parte de los metadatos, tales como el tamaño y la fecha de un fichero, se produce automáticamente cuando se crea el fichero. El software de las aplicaciones también produce metadatos al crearse un fichero, por ejemplo el nombre y tipo del fichero.

Los nombres y los valores deben ser sencillos y coherentes. Una forma de lograrlo es con un vocabulario controlado. Esto simplemente significa una lista de palabras y frases para identificar la información digital y recuperarla por medio de una búsqueda. Se deben emplear palabras clave descriptivas y nombres de carpetas y ficheros que sean claros (no ambiguos) y significativos.

Un nombre de fichero es un metadato y debe proporcionar una descripción del contenido.

En el ámbito bibliotecario, existen varias y complejas normas o esquemas para describir los ficheros digitales de diversos tipos de piezas (libros, fotografías, materiales de sonido, etc.), y cada uno tiene su propia estructura. Sin embargo, siempre y cuando se sea coherente, fácilmente puedes crear tu propio sistema sencillo de metadatos, de acuerdo a las fuentes y los requisitos de tu repositorio.

El programa de software que escojas para la base de datos tendrá su propio sistema para incorporar nuevos metadatos. Por ejemplo, el repositorio electrónico de los Archivos Históricos de la OSG contiene miles de documentos de grupo. Para encontrar la información precisa de un grupo, creamos una estructura de metadatos exclusivamente para los registros de Grupo. Sus campos incluyen los siguientes: “Nombre del grupo,” “Número del grupo,” “Sede de reuniones,” “Área,” “Distrito,” “Estado” y “País.” Esto le permite al programa localizar un grupo eficazmente al introducir datos de cualquiera de estos campos. No se queda aquí el asunto de crear metadatos, pero lo más importante es no olvidarse de mantener un sistema sencillo y coherente.

LAS ESPECIFICACIONES DE DIGITALIZACIÓN

Diversos factores determinan la apariencia de las imágenes, sean expuestas o impresas. Por lo tanto, al escanear, hay diferentes requisitos en cuanto a la resolución

| Formato de ficheros | Características técnicas | Uso recomendado |
|--|--|--|
| TIFF (Tagged Image File Format) | <ul style="list-style-type: none">• Apoya la mayoría de las plataformas.• Sin pérdidas (no hay disminución de calidad debido a la compresión).• Sirve para ficheros muy grandes.• Formato flexible.• Formato de imagen preferido para la preservación. | Formato recomendado para el archivo. El más universal y ampliamente aceptado. |
| JPEG (Joint Photographic Expert Group) | <ul style="list-style-type: none">• Aceptable para ficheros de imágenes fotográficas si se utiliza el modo menos comprimido.• No se pierde calidad al comprimir y guardar un fichero JPEG. | Sólo para uso en ficheros de acceso. No se recomienda para los ficheros maestros de producción. |
| PDF (Portable Document Format) | <ul style="list-style-type: none">• Formato de fichero complejo. | Sólo para uso en ficheros de acceso. No se recomienda para los ficheros maestros de producción o para archivar. Se puede procesar para que reconozca texto OCR y se pueda buscar el fichero. |

y tipo de fichero según los diferentes tipos, formatos e índole de los originales. Sin embargo, hay demasiados rasgos particulares en un documento, cada uno de los cuales requiere parámetros específicos.

- Cuando se escanean documentos impresos (por ejemplo, impresos con láser o compuestos tipográficamente), por lo general se recomienda un mínimo de 300 ppi en el modo de escala de grises.
- Si estás escaneando documentos de poca legibilidad, escritos a mano, copias en carbón o fotografías, el parámetro recomendado es de 400 ppi con un espacio de color de 24 bits.
- Se debe usar los formatos de fichero abiertos ya que son de acceso público y muchos programas los reconocen. En el cuadro que figura más abajo, se encuentra información sobre los tipos de formatos de fichero abiertos utilizados en proyectos de digitalización.

Open Office XML también es un formato de fichero abierto que se emplea en Microsoft Office 2007, 2010 y 2013. Algunos ejemplos de estos ficheros incluyen DOCX (documento producido con un procesador de textos), XLSX (hoja de cálculo) y PPTX (presentaciones). Versiones anteriores de Microsoft Office puede que no estén en formato de fichero abierto y, por lo tanto, sólo se puedan abrir con Microsoft Office.

EL CONTROL DE CALIDAD

Una revisión de la calidad de las imágenes escaneadas es crucial y constituye un paso significativo del proceso de digitalización. Es, además, aconsejable que se inspeccione el escáner periódicamente para asegurar que no quede materia física, tales como polvo o partículas de papel, sobre la cama plana del escáner.

Recuerda siempre mantener tu área de trabajo limpia. Esta tarea incluye el mantenimiento diario del equipo de escanear. Muchos de los documentos y libros viejos suelen tener bastante polvo y hasta moho y puede que dejen sustancias potencialmente dañinas en la cama plana del escáner. De vez en cuando, se deben revisar visualmente los ficheros de imágenes digitales para asegurar la calidad en las siguientes áreas:

- Se ha usado el formato (TIFF, JPEG, etc.) y la resolución debidas.

- Está correctamente orientada la imagen original.
- Para los documentos de múltiples páginas, que estén todas las páginas y en orden correcto.
- La imagen digital debe estar entera y no recortada.
- La imagen digital no debe estar torcida.
- No deben aparecer en la imagen artefactos digitales (polvo, partículas de papel, etc.).
- Se mantiene la calidad de la imagen en cuanto al color, al tono, a la nitidez y al contraste.
- Nombres relacionados con los metadatos — los ficheros digitales deben estar correctamente identificados.

LA PRESERVACIÓN DE MATERIALES ORIGINALES DESPUÉS DE ESCANEARLOS

El resultado final de un proyecto de digitalización es una réplica digital que corresponde a los originales. Es importante cuidar bien de los materiales originales. Aunque tengas la copia digital, aún es crucial proteger los documentos históricos del deterioro mediante el uso de materiales de calidad archivística y un entorno con condiciones óptimas para el almacenamiento.

Consulta las “Directrices para la preservación de los Archivos de la OSG” para un examen más completo y profundo de la preservación de materiales de archivo.

TEN PRESENTE . . .

Para la respuesta a cualquier pregunta y listas de otros recursos, no dudes en ponerte en contacto con los Archivos de la OSG por correo electrónico, archives@aa.org, o llamando al (212) 870-3400. Para más información de valor, visita el sitio Web de la OSG de A.A., www.aa.org.

FUENTES DE INFORMACIÓN ADICIONAL

Ten en cuenta que los Archivos Históricos de la OSG ni respaldan ni avalan estos sitios Web y los provee sencillamente como fuentes externas de recursos útiles.

National Archives and Records Administration (NARA)
[Administración Nacional de Archivos y Documentos]
<http://www.archives.gov/preservation/technical/guidelines.html>

Digital Preservation Coalition
[Coalición para la Conservación Digital]
<http://www.dpconline.org/advice/preservationhandbook>

