

Tratamiento digital de imágenes en archivos. Modelos de gestión y preservación

Dra. ANTONIA SALVADOR BENÍTEZ

F. de Biblioteconomía y documentación. U. de Granada

1. INTRODUCCIÓN

La introducción de sistemas de gestión electrónica de documentos ha incorporado nuevas prácticas en el tratamiento documental integrando en un mismo sistema la imagen y la información descriptiva. El proceso de conversión digital de las fotografías comienza con el escaneado o captura mediante una cámara digital del que se obtiene una copia digital del original. La información descriptiva asociada a cada imagen también debe adaptarse a esta plataforma y para ello es necesario aplicar un proceso de codificación que va a convertir la estructura informativa en una nueva compuesta de metadatos.

La imagen -en su nueva naturaleza digital- va a requerir para su tratamiento, gestión y difusión, el empleo de una infraestructura tecnológica lo que implica el tomar una serie de decisiones de carácter técnico, de presupuesto y el desarrollo de una metodología de trabajo que el profesional responsable debe conocer y valorar. En este proceso, la preservación constituye un aspecto fundamental que requiere una continua supervisión ya que el material digital puede perderse si no se realiza ninguna intervención. Para esto resulta necesario el desarrollo de estrategias de actuación así como el empleo de normas y directrices nacionales e internacionales para asegurar la vida de la colección.

Con este trabajo presentamos las etapas del proceso de digitalización así como la necesaria planificación y estudio de viabilidad del proyecto como principales medidas para garantizar una gestión adecuada.

2. LA CADENA DE DIGITALIZACIÓN

El proceso completo de conversión se denomina cadena de

digitalización, y comprende tres operaciones o fases: la creación de la imagen, gestión de archivos y la presentación y difusión de la imagen. Cada eslabón de la cadena requiere un equipo técnico cuyas características y configuración van a depender de las necesidades documentales a las que deban responder. El esquema del que partimos¹ es una representación simplificada ya que en la práctica real la cadena de digitalización puede no desarrollarse de una forma lineal sino que puede tener distintas ramificaciones.

2.1 Creación de la imagen

Todo proceso de conversión digital de un fondo fotográfico tiene como primer paso determinar el material a digitalizar y el establecimiento de los criterios a seguir. Para garantizar la calidad del nuevo registro, la captura de imágenes debe tener en cuenta los procesos técnicos que comprende la conversión de analógico a digital y los atributos de los documentos: dimensiones y presentación, nivel de detalles, color, etc. Algunos factores que afectan a la calidad de la imagen son: resolución², profundidad de bits y tonalidad³, formato de archivo⁴, compresión⁵ y el rendimiento del sistema.

2.1.1 Captura o conversión de la imagen

Mediante esta operación se procede a la captura o conversión inicial de las imágenes en formato analógico (papel, película...) a un formato digital legible por ordenador mediante un escáner o cámara digital obteniéndose un duplicado de la imagen⁶. Algunas cuestiones a tener en cuenta son la compatibilidad de los dispositivos de captura con los documentos, es decir, si permiten trabajar con distintos tamaños o tipos de documentos teniendo en cuenta las características de los originales; la calidad y prestaciones que ofre-

¹ Nos hemos apoyado en el trabajo realizado por Cornell University Library, *Moving theory into practice: digital imaging for libraries and archives*. California, Research Group Library, 2000

² Si la imagen se va a utilizar como referencia visual en pantalla la resolución puede ser de 72 ppp, en cambio, si el producto digital va a ser sustituto del original y su uso va a estar orientado a la edición es recomendable una resolución de 300 ppp

³ Si el objetivo es la consulta en Internet (web) o en CD puede ser suficiente una profundidad de 8 bits/píxel (256 colores), en cambio si el producto digital va a sustituir al original se recomienda que sea de 24 bits/píxel (16,7 millones de colores)

⁴ El formato de archivo para las imágenes originales deberá aceptar la resolución, profundidad de bits, información de color y metadatos que se ajusten a nuestras necesidades. También debe permitir ser guardado en forma descomprimida o comprimida utilizando técnicas con pérdida o sin ella.

⁵ Se recomienda utilizar una técnica de compresión estándar y ampliamente compatible en lugar de una patentada que, con mejor calidad de compresión, puede no prestarse a un uso y preservación a largo plazo.

⁶ A la hora de seleccionar los dispositivos de captura o conversión de la imagen (escáner, cámaras digitales, chasis digitalizadores) debemos tener en cuenta que, de todos los productos que ofrece el mercado, muy pocos están diseñados para el uso específico en archivos y bibliotecas, por esta razón el objetivo será encontrar aquellos que mejor se adapten a nuestras necesidades.

cen; las necesidades a las que deben responder de acuerdo a nuestro archivo y a los servicios que vamos a prestar; rendimiento, capacidad, mantenimiento y presupuesto de conversión, entre otros.

2.1.2 Procesamiento de imágenes/archivos

A la imagen resultante se le pueden aplicar en cualquier punto de la cadena de digitalización uno o más pasos de procesamiento para alterar, extraer o agregar datos. Algunas de las operaciones de procesamiento más comunes son la edición, retoque y mejora, compresión, conversión de formato de archivo, OCR y creación de metadatos.

La creación de metadatos consiste en la incorporación de texto para describir varios atributos de los documentos, dotándolos de significado, contexto y organización⁷. Los metadatos pueden clasificarse en tres categorías atendiendo a sus funciones: metadatos descriptivos, estructurales y administrativos. No obstante estas categorías no están claramente definidas y a menudo se prestan a una cierta superposición. Los enunciamos brevemente:

- *Metadatos descriptivos* (MARC, Dublin Core). Su objetivo es la descripción e identificación de los recursos de información bien en un sistema local, bien en aplicación web para facilitar a los usuarios la localización de recursos.
- *Metadatos estructurales* (XML, EAD). Facilitan la navegación y presentación de los recursos electrónicos proporcionando información sobre la estructura interna de los mismos (página, sección, capítulo, índices); la relación entre los materiales y unen los archivos de imagen y texto relacionados.
- *Metadatos administrativos* (Administrative Metadata Elements). Facilitan la gestión y procesamiento de las colecciones digitales incluyendo datos técnicos sobre la creación y el control de calidad, gestión de derechos, control de acceso y utilización y condiciones de preservación.

2.2 Gestión de archivos

Las ventajas de un sistema de gestión que integre texto e imagen son innegables, sin embargo hay que tomar conciencia de la complejidad tanto técnica como de tratamiento documental, consubstancial a la imagen. Por ello debe ser un proceso totalmente valorado y ponderado por el profesional con argumentos sólidos para apoyar la toma de decisiones que conlleva.

- a) *Decisiones de carácter documental*, que afectan a la organización, descripción, almacenamiento y mantenimiento de las imágenes y metadatos relacionados. Al-

gunos de los aspectos a considerar desde el punto de vista documental son: política y nivel de descripción⁸, análisis de contenido, empleo de un lenguaje controlado y política de indexación.

- b) *Decisiones de carácter técnico*, principalmente la elección del software de gestión y bases de datos. Podemos optar por el diseño de un sistema formado por una base de datos textual y otra de imágenes, o bien, el empleo de un programa de gestión específico como los sistemas de Gestión Electrónica de Documentos (GED) que ofrecen todas las prestaciones de las bases de datos documentales tradicionales y además permiten determinar la presentación y selección de los resultados de búsqueda así como una serie de herramientas para la manipulación de las imágenes. Algunas de las prestaciones que debe ofrecer un programa para la gestión de imágenes son: posibilidad de integración, captura, almacenamiento, capacidad, formato y calidad de las imágenes que soporta Importación y exportación de datos e imágenes, visualización e impresión, integración de funciones de gestión, rapidez en las búsquedas y sencillez de manejo.
- c) *Organización y estructura de archivos*. Un buen sistema de organización de archivos y directorios puede ayudar a minimizar el desorden y la dispersión, especialmente en el caso de colecciones muy grandes. Con frecuencia, la naturaleza del material escaneado sugiere los criterios de organización (nº de página de las monografías, nº carpeta o de acceso para manuscritos y fotografías, etc.). En otros casos la organización vendrá determinada por criterios de carácter interno como el contenido temático. A la hora de asignar nombre a los archivos y directorios es importante emplear un sistema estándar que sea compatible con otros sistemas. También es recomendable emplear extensiones estándar para los distintos tipos de archivos.
- d) *Almacenamiento*. Con frecuencia, satisfacer todas las necesidades relacionadas con el almacenamiento puede requerir el uso de más de una tecnología (cd-r, dvd, zip, etc). Dependiendo de las condiciones de manipulación y almacenamiento algunos medios pueden tener una vida útil más o menos limitada y pueden estar sujetos a una mayor obsolescencia. Los criterios para seleccionar el almacenamiento masivo apropiado para una iniciativa de digitalización son de un modo general la velocidad (lectura, escritura, transferencia) para agilizar el acceso a los datos; capacidad; fiabilidad y adecuación para la función que va a desempeñar; y la estandarización y compatibilidad con

⁷ La teoría y práctica de los metadatos tiene sus raíces en la catalogación de documentos impresos. En el mundo digital se han incorporado para facilitar la gestión de los archivos y la navegación en red.

⁸ En estos casos resulta especialmente interesante establecer los campos de acuerdo a los metadatos definidos por la EAD (Encoded Archival Description) para preveer su integración en Intranet/Internet.

otros medios y costo⁹. Además, todas las tecnologías de almacenamiento sacrifican cierta cantidad de espacio, cuestión que también habrá que considerar a la hora de estimar las necesidades de almacenamiento.

2.3 Entrega y difusión

Comprende el proceso final que consiste en hacer llegar las imágenes digitales y los archivos auxiliares al usuario. Los componentes más importantes son las redes y los dispositivos de visualización, principalmente impresoras y monitores. Es evidente que las decisiones acerca de los formatos de archivo, las relaciones de compresión y la aplicación de escalas van a tener un impacto sobre la entrega.

En esta etapa resulta fundamental conocer las características y necesidades del usuario. A menos que las imágenes digitales sean de uso interno, algunos de los componentes de la entrega van a estar fuera de nuestro control y van a depender de la tecnología que disponga el usuario para hacer la consulta o impresión de las imágenes. Siempre que los recursos lo permitan la mejor solución es ofrecer varias versiones y calidades de presentación que se adapten a las distintas infraestructuras técnicas de manera que sea satisfactoria tanto para los usuarios como para el personal interno.

La utilización de la web para facilitar el acceso y la difusión de los recursos a un gran número de usuarios plantea una serie de cuestiones de cara al público receptor¹⁰. Esto nos lleva a considerar una serie de aspectos que afectan a la visualización como son el formato de archivo y las técnicas de compresión utilizadas, capacidades del navegador web, conexiones de red y velocidad de acceso y calidad de la imagen en la pantalla, entre otras cuestiones.

3. GESTIÓN. EL CICLO DE VIDA DE UN PROYECTO DE DIGITALIZACIÓN

El ciclo vital no es otra cosa que la definición de los objetivos y fases de actuación, la temporalización, administración de los recursos, la coordinación de los equipos de trabajo así como la evaluación y prospectiva de futuro de un proyecto de digitalización.

El concepto de ciclo de vida del material digital fue desarrollado en un importante proyecto que ha resultado clave para explorar los desafíos asociados con su conservación¹¹. En cada una de las fases del ciclo los intereses y prioridades son distintos, pero lo esencial es que el aspecto de la conservación debe tenerse presente en todas y cada una de las etapas y no sólo al final. El reconocimiento del ciclo vital del proyecto implica una planificación en donde las fases y la evolución del trabajo deben estar claramente definidas. Al

mismo tiempo se deben conseguir los recursos necesarios para obtener los resultados en el tiempo previsto y con el presupuesto disponible.

3.1 Planificación y recursos

Esta fase previa comprende una serie de cuestiones de carácter general que van a influir de manera decisiva en el establecimiento de los objetivos y fases posteriores. Las instituciones interesadas en iniciar un programa de digitalización de imágenes deben abordar tres aspectos esenciales como son la finalidad del proyecto, recursos necesarios y la metodología de trabajo, cuestiones entre las que se establece una relación de interdependencia. El núcleo del proyecto lo constituyen los objetivos y los resultados en los que se traducen dichos objetivos. A partir de aquí se determinan los recursos de la institución para llevar a cabo el proyecto, es decir, selección del material, colección o colecciones a digitalizar, personal técnico, capacitación y formación, tiempo de ejecución, recursos económicos y espacio físico de trabajo. En función de los recursos se establecerán los pasos y procesos a seguir en la digitalización de imágenes. Algunas de las cuestiones a tener en cuenta en esta fase son:

a) Identificación de objetivos, servicios y tipo de usuario, es decir, cuál es el objetivo de la digitalización del fondo (difundir, conservar o ambas), uso de las imágenes y cuáles son los usuarios finales del sistema (personal interno, investigadores, diseñadores gráficos, editoriales, público en general, etc).

b) Características de la documentación y criterios de selección. Tamaño del fondo y posibilidades de crecimiento, naturaleza del material: soportes, tamaño, color, contenido temático y criterios de selección en función de los usos previstos, protección de los derechos de autor, preservación, organización del material, capacidad de la institución para la gestión, difusión y mantenimiento de los materiales digitalizados.

Del mismo modo se deben establecer los recursos económicos y humanos así como la infraestructura técnica necesaria para llevar a cabo la conversión.

3.2 Metodología

Resulta fundamental desarrollar unas pautas metodológicas que orienten la manipulación del material, la descripción, el proceso de escaneado, la creación de metadatos, el mantenimiento y control de calidad que garanticen la coherencia y uniformidad del tratamiento. También han de aplicarse las normas y requisitos legales establecidos según el caso.

La evaluación y el control de calidad del proceso debe po-

⁹ En este sentido no es recomendable adquirir mayor cantidad de tecnología de almacenamiento de la que necesitamos en previsión del crecimiento futuro ya que podemos pagarla a un precio mucho más elevado que decrecerá a medida que pase el tiempo.

¹⁰ Los estudios que se han realizado en este sentido ponen de manifiesto que el usuario espera una recuperación rápida, la visualización completa de la imagen y con una calidad aceptable.

¹¹ Concretamente en el estudio *JISC/NPO sobre la conservación del material electrónico*, cuyo objetivo es elaborar un marco de políticas estratégicas para crear y conservar el material digital.

ner de manifiesto que lo realmente importante no es llevar a cabo la conversión digital sino la metodología necesaria para garantizar la calidad de los procesos así como los productos y servicios resultantes. En la medida que se valoren y analicen los factores y procedimientos que intervienen en el proyecto, podremos definir con más precisión nuestros objetivos y los servicios que queremos desarrollar.

3.3 Preservación digital

Se trata de un término amplio empleado para definir tanto el mantenimiento como la salvaguarda del material digital a corto y a largo plazo. De hecho, la estrategia de preservación debe considerarse como uno de los aspectos más importantes de un proyecto de digitalización que requiere una cuidadosa planificación sobre todo en lo que se refiere al empleo de normas y directrices nacionales e internacionales y la utilización de formatos de archivo y sistemas de almacenamiento estándar para asegurar la vida de la colección. Por esta razón la IFLA, junto a otras organizaciones están tratando de hacer conscientes de los problemas inherentes a la conservación de los materiales digitales a corto y a largo plazo. Los actuales estudios sobre el tema hacen patente la necesidad de estrategias y pautas de conservación digital promoviendo la conciencia sobre los problemas y particularidades de los archivos digitales¹².

a) Estrategias técnicas

Entre las primeras medidas citamos el *mantenimiento*, entendido como una estrategia continua para garantizar el buen estado de los recursos digitales (almacenamiento de las imágenes y archivos en medios seguros y fiables; manipulación adecuada para optimizar su expectativa de vida y la realización de verificaciones y copias de seguridad de forma periódica y sistemática).

La *actualización*, aunque no es una estrategia de preservación en sentido pleno debe entenderse como una parte esencial de la política de conservación a largo plazo. Significa la copia del contenido de un sistema de almacenamiento que ha quedado obsoleto a otro más actual.

Otras estrategias son la *migración*, es decir, el proceso de transferencia de información digital de una configuración a otra o de una generación de ordenadores a otra¹³ y la *emula-*

ción, que comprende la recreación del entorno técnico requerido para ver y utilizar la colección digital.

Finalmente citamos la *preservación de la tecnología*, es decir, el entorno que hace funcionar el sistema (sistemas operativos, programas, aplicaciones originales, etc) así como la *arqueología digital* que incluye métodos y procedimientos para rescatar contenidos dañados de programas o entornos obsoletos o afectados por cualquier anomalía.

b) Estrategias de organización y gestión

Implican la contratación de personal, programas de formación y la dotación de recursos económicos pero sobre todo, una gestión efectiva de las colecciones digitales requiere que las instituciones desarrollen y sigan un plan de actuación para evaluar los requisitos de la preservación y el acceso a largo plazo, identificando costos y beneficios y valorando los riesgos. Ante la pregunta de si la digitalización puede considerarse un modo de preservación podemos decir que en una pequeña medida, la digitalización reduce la necesidad de manipular los originales gracias a la consulta electrónica y permite la creación y empleo de archivos maestros con fines editoriales. Sin embargo la ausencia de normas y pautas de actuación y la excesiva dependencia tecnológica tanto en el proceso como en los productos resultantes hacen que de momento, la digitalización no se considere una buena fórmula para la preservación.

En la medida en que se adopten estas estrategias y se haga una apuesta clara por el desarrollo de normas y prácticas metodológicas se podrá ir consolidando la digitalización como un verdadero principio de preservación¹⁴.

La digitalización resulta muy útil para facilitar la consulta y evitar manipulaciones innecesarias de los originales. En este sentido, puede resultar interesante considerar la digitalización parcial de un archivo o fondo porque su interés y valor así lo requiera. Pero un proyecto de conversión retrospectiva de archivos, colecciones o fondos fotográficos en soporte tradicional no debe hacernos olvidar que el principal objeto de nuestra protección no es tanto la versión digital de la colección sino el material original y que, salvo en las tareas de consulta, la mejor copia digital nunca podrá sustituir a la fuente primaria, que deberá ser conservada permanentemente.

¹² En este sentido la National Preservation Organization (NPO) creó un grupo de trabajo sobre archivos digitales que dio como fruto el proyecto *Preservation Management of Digital Materials* (Gestión de la conservación de materiales digitales) para tratar de definir orientaciones y pautas para la conservación digital, recursos y la colaboración de servicios externos para la preservación del material.

¹³ También puede aplicarse para mover archivos de imágenes de un formato de archivo obsoleto a otro más reciente o para aumentar su funcionalidad.

¹⁴ En esta línea están en marcha una serie de proyectos y enfoques prácticos que tratan de respaldar la gestión organizativa de las colecciones digitales y pueden ser un modelo de referencia para iniciativas de este tipo. En el Reino Unido, *Arts and Humanities Data Service* (Servicio de Datos sobre las Artes y las Humanidades), dedicado al análisis costo-beneficio y la toma de decisiones que intervienen en la preservación digital (<http://ahds.ac.uk>); en California, el proyecto de la Universidad de Cornell *Risk and Management of Digital Information* (Gestión de Riesgo de la Información Digital), encargado de desarrollar una herramienta para la evaluación de los riesgos implicados en la migración de datos (<http://www.clir.org/pubs/reports/pub93/contents.html>). Otro proyecto que ofrece un marco para la preservación y el acceso digital a largo plazo es *Open Archival Information System - OAIS*. Incluye terminología y conceptos para comparar distintas arquitecturas de archivo (<http://ssdoo.gsfc.nasa.gov/nost/isoas/us/overview.html>); *Cedars 1* (CURL Exemplars in Digital Archives - Modelos CURL en los Archivos Digitales) tiene como objetivo desarrollar estrategias en la política de gestión de las colecciones digitales y promover métodos adecuados para la preservación a largo plazo de distintos recursos digitales así como la creación de los metadatos (<http://www.leeds.ac.uk/cedars>); la *Biblioteca Europea de Depósito en Red*, proyecto de colaboración de las bibliotecas nacionales europeas para la creación de una Biblioteca de depósito en red, ha desarrollado procedimientos de mantenimiento de archivos, requisitos de enlace entre metadatos y estrategias de preservación (<http://www.kb.nl/coop/nedlib>).