

## LA DIGITALIZACIÓN DE MATERIALES BIBLIOTECARIOS EN LA BIBLIOTECA NACIONAL<sup>1</sup>

XAVIER AGENJO  
FRANCISCA HERNÁNDEZ  
*Biblioteca Nacional. Madrid*

### 1. LA DIGITALIZACIÓN: DEFINICIONES GENERALES<sup>2</sup>

Con el término digitalización denominamos habitualmente tanto al proceso concreto de transformación de imágenes (y también sonidos), etc. en códigos binarios, como a los sistemas y procesos desarrollados a partir de los años 80 para la captura, tratamiento, almacenamiento y transmisión de imágenes o sonidos digitalizado. Estrictamente es la representación y almacenamiento de la información, sea ésta icónica o sonora, en códigos binarios, es decir, en secuencias de 0 y 1 (*bits*).

El proceso de digitalización comienza en el momento de la captura de la imagen, para lo cual el documento se divide en líneas horizontales que a su vez son subdivididas en pequeñas áreas de escaneo, *pixels* o *puntos*, mediante

---

<sup>1</sup> Una versión abreviada de este artículo fue presentada en la Conferencia anual de LIBER celebrada en Göttingen los días 4 a 9 de julio de 1994.

<sup>2</sup> Puede usarse la definición que proporciona M. Stuart Lynn y el Technology Assesment Advisory Committee de la Commission on Preservation and Access: *Preservation and Access Technology: The Relationship between Digital and Other Media Conversion Processes: A Structured Glossary of Technical Terms*. Publicado inicialmente en *Information Technology and Libraries*, December 1990, pp. 309-336

Las *Tecnologías Digitales Electrónicas* son las tecnologías utilizadas para capturar, almacenar, transformar, distribuir o presentar información en forma electrónica (normalmente como una secuencia de 0 y 1 denominada bits).

La *Digitalización Electrónica* se refiere a la captura de documentos en forma electrónica por medio de un proceso de escaneo y digitalización. La imagen escaneada se almacena electrónicamente, normalmente en un medio de almacenamiento óptico o magnético. La imagen almacenada electrónicamente puede ser transformada posteriormente para su compresión o interpretación; y consecuentemente puede ser seleccionada para ser usada por tecnologías de acceso, distribuida por medio de tecnologías de distribución o visualizada por medio de tecnologías de presentación.

un sistema periférico denominado scanner<sup>3</sup>. Los escáneres están provistos de componentes fotosensibles que miden la cantidad de luz reflejada por estos puntos y transmiten la correspondiente señal eléctrica a la unidad de proceso de imágenes que realiza la codificación digital. En el caso más sencillo los elementos de imagen (puntos) son codificados como 0 para el blanco y 1 para el negro.

De forma resumida las principales ventajas de la digitalización son: preservación, facilidad de acceso, velocidad en la recuperación de la información, transmisión en línea y de forma remota y elevadísima capacidad de almacenamiento<sup>4</sup>.

## 2. DIGITALIZACIÓN Y PRESERVACIÓN

Las bibliotecas han ido incorporando a lo largo de los años distintas tecnologías para detener el deterioro de los materiales que conservan. Podemos agrupar las diferentes tecnologías disponibles en tres categorías:

— las que se dirigen a la preservación del contenido y a la conservación del objeto original (p.e. restauración, desacidificación<sup>5</sup>).

— las que preservan el contenido y reproducen el soporte original del documento de la manera más fiel posible (microfilmación, fotocopia de preservación<sup>6</sup>, digitalización).

— las que preservan el contenido únicamente (grabación masiva, uso de cintas de fotocomposición).

Hay que señalar que si bien la digitalización entra dentro de las tecnologías aplicadas a la preservación que se dirigen a salvaguardar el contenido y reproducir el soporte físico original, esto sólo ocurre cuando se trata de documentos en soporte papel o pergamino, en el caso de documentos antiguos, ya que cuando hablamos de sonido o imágenes en movimiento el soporte original no forma parte de la reproducción digital.

---

<sup>3</sup> Del inglés "to scan": explorar, registrar.

<sup>4</sup> Existe una nueva norma: ISO 10090 Information processing systems - 90mm rewritable read only optical disc cartridge for information interchange. Esta norma va en la línea de asegurar la portabilidad de la información digitalizada.

<sup>5</sup> *Panorama general de las técnicas de desacidificación masivas* / Francisca Hernández Carrascal // Boletín de la ANABAD. ISSN 0210-4164. XLII : 2 (abril-junio 1992), 123-132.

<sup>6</sup> Fotocopias en papel permanente. Véase: *Manual of archival reprography* / edited by Lajos Kórmendy ; compiled by James A. Keen ... [et al.]. - München : Saur, 1989. - ISBN 3-598-20277-6.

La digitalización puede usarse para la preservación de todo tipo de información existente en formato no digital.

Es importante remarcar que en la actualidad lo que hace unos años eran nuevos materiales bibliotecarios han ido ingresando en gran cantidad en nuestras bibliotecas y que estos materiales están sufriendo asimismo un proceso de deterioro cuya solución conlleva problemas diferentes a los planteados por el papel<sup>7</sup>. Nos estamos refiriendo a:

— Tecnologías analógicas vigentes: casetes, fotografías, películas, cintas de video VHS, etc.

— Tecnologías analógicas *obsoletas* o semiobsoletas: discos de pizarra o vinilo de 33, 45 y 78 rpm, cintas de video Beta, casetes audio de 8 pistas, cintas magnetofónicas, etc.

— Unidades de almacenamiento para microcomputadores obsoletas o semiobsoletas: adaptadores para casetes audio, disquetes flexibles de 5 1/4 pulgadas.

— Unidades de almacenamiento para microcomputadores vigentes: disquetes de 3 1/2 pulgadas, cintas DAT (digital audio tape), CD-ROM, CDI, DVI, VCR<sup>8</sup>.

La mayor parte de estos materiales y de las tecnologías correspondientes fueron diseñadas sin ánimo alguno de perdurabilidad. En muchos casos los aparatos lectores necesarios para su consulta han desaparecido o tienen una corta duración.

Para todos estos materiales la preservación no estriba en la permanencia del soporte físico, sino en la copia o reformateado de la información. Para ello la digitalización ofrece una vía cada vez más económica y eficaz.

Los problemas que plantean los nuevos medios en cuanto a su preservación son:

La obsolescencia tecnológica de algunos medios es un problema mayor que el de su perdurabilidad o deterioro. La evolución del mercado ha provocado, como se ha mencionado, la desaparición de determinados productos o de sus dispositivos de lectura.

---

<sup>7</sup> Un breve, pero clarísimo estudio sobre la preservación de estos materiales lo ofrece Lesk, Michael: *Preservation of new technology: a report of the Technology Assessment Advisory Committee to the Commission on Preservation and Access*. - Washington, The Commission on Preservation and Access, 1992.

<sup>8</sup> Para una introducción a las tecnologías de disco compacto véase: *Tecnologías de Disco Compacto* // Boletín informativo UDT (Transmisión Universal de Datos y Telecomunicaciones). ISSN 1021-3333. 24 (Primavera 1994), 5-6.

El estado físico del objeto no es un elemento definitivo sobre el que basar un análisis del estado de la información que contiene. Hay una gran interrelación entre formatos, soporte físico y soporte lógico, puede ser que la información esté preservada, pero se necesita salvaguardar también el soporte lógico preciso para buscar, recuperar, editar e imprimir.

Por estas razones, la supervivencia de la información digital no depende de la perdurabilidad del objeto particular si no de que sea refrescada periódicamente en una nueva tecnología. Para este tipo de materiales el reformateo de la información constituye la mejor medida de preservación.

En palabras de Patricia Battin<sup>9</sup> hay que considerar "...la capacidad de la tecnología digital para la preservación por varias razones importantes: 1) la necesidad de conservar los documentos científicos retrospectivos impresos en papel que se deteriora; 2) las enormes desventajas de las microformas para difusión y uso científico; 3) la necesidad de asegurar un acceso permanente a la información futura almacenada en una vía de transmisión de soportes electrónicos que cambian rápidamente".

## 2.1 *Microfilmación o digitalización*

Para muchos bibliotecarios el dilema se produce cuando se trata de adoptar tecnologías diferentes que se dirigen a un mismo fin y con resultados similares. Este es el caso de la microfilmación y la digitalización. Por supuesto, lo que se debate es la aplicación a documentos impresos o manuscritos o en soporte papel, ya que para sonido o imágenes en movimiento la digitalización es la alternativa más razonable. Compararemos a continuación las ventajas y desventajas que ofrecen ambas tecnologías.

— Una de las críticas más utilizadas en contra de la implantación de la digitalización es la perdurabilidad del soporte de almacenamiento.

El microfilm es un soporte duradero, siempre que se mantengan las condiciones necesarias de almacenamiento y siempre que su producción se haya efectuado bajo la normativa adecuada<sup>10</sup>. Básicamente diremos que para

---

<sup>9</sup> *From preservation to access : paradigm for the nineties* / Patricia Battin. // 59th IFLA Council and General Conference : Barcelona, Spain, 22-28 August 1993. - Booklet 0, pp. 81-84.

<sup>10</sup> Normativa disponible ISO para microfilmación:

— 435-1975: Documentary reproduction - ISO conventional typographical character for legibility tests (ISO character)

— 446-1975: Microcopying - ISO No. 1 Mire - Description and use in photographic documentary reproduction.

— 1116-1975: Microcopying - 16 mm. and 35 mm. microfilms, spools and reels.

que se cumplan las condiciones de perdurabilidad citadas, han de realizarse al menos dos copias del original, revisiones periódicas de una muestra de las mismas, copias periódicas, etc. Esta situación óptima no es la se encuentra en las bibliotecas, donde, en el caso de que existan medios para microfilmear o programas de microfilmación, una obligada reducción de gastos limita frecuentemente el número de copias a 1, que se utiliza para lectura, obtención de reproducciones en papel, duplicación, etc.

Se sitúa la duración de los soportes ópticos<sup>11</sup> entre 10 y 20 años. Algunos fabricantes de WORM llegan a garantizar una perdurabilidad de 100 años, aunque esta situación está cambiando rápidamente. Sin embargo, con ser este un punto importante a la hora de la decisión, el problema más grave es la rapidez con la que la tecnología, tanto de lectura como de almacenamiento,

---

— 2707-1980: Micrographics - Transparent A6 size microfiche of uniform division - Image arrangements No. 1 and No. 2.

— 2708-1980: Micrographics - Transparent A6 size microfiche of variable division - Image arrangements A and B.

— 2803-1974: Microfilmes tipo gelatina de plata. Procesado y almacenamiento para archivo permanente

— 3272/1-1983: Microfilming of technical drawings and other drawing office documents - Part 1: Operating procedures.

— 3272/2-1978: Microfilming of technical drawings and other drawing office documents - Part 2: Quality criteria and control.

— 3272/3- 1975 Microfilming of technical drawings and other drawing office documents - Part 3: Unitized 35 mm. microfilm carriers.

— 3334-1976: Microcopying - ISO Test Chart No. 2 - Description and use in photographic documentary reproduction.

— 4087-1979: Microfilming of newspapers on 35 mm. unperforated microfilm for archival purposes.

— 4331-1986: Photography - Processed photographic black-and-white film for archival records - Silver-gelatin type on cellulose ester base - Specifications.

— 4332-1986: Photography - Processed photographic film black-and-white film for archival records - Silver-gelatin type on poly(ethylene terphthalate) base - Specifications

— 5123-1984: Documentation - Headers for microfiche of monographs and serials.

— 5126-1980: Micrographics - Computer output microfiche (COM) - Microfiche A6.

— 6196/1-1980 Micrographics - Vocabulary - Section 01 : General terms.

— 6197/1-1980 Microfilming of press cuttings - Part 1 : 16 mm Silver-gelatin type roll microfilm.

— 6200-1979 Micrographics - Density of silver-gelatin type films.

— 6343-1981 Micrographics - Unitized microfilm carrier (aperture card) - Determination of adhesion of protection sheet to aperture adhesive.

— 6829-1983 Flowchart symbols and their use in micrographics.

— 8126-1986 Micrographics - Diazo and vesicular films - Visual density - Specifications.

<sup>11</sup> Se rechaza, sin embargo, el uso de las nuevas tecnologías de almacenamiento: *Dual Alloy writeonce chemistry*, por su falta de longevidad.

se vuelve obsoleta. La solución está a la orden del día en cualquier centro de proceso de datos y tiene una respuesta clarísima: refrescado regular de los datos. Por otro lado, una vez digitalizados los documentos pueden obtenerse copias de los mismos en papel permanente, del mismo modo que existen escáneres en el mercado que ofrecen la posibilidad de microfilmarse y digitalizar en una sola operación<sup>12</sup>.

— El coste de la microfilmación es menor<sup>13</sup>, aunque la mayor parte de los datos en este sentido provienen de las bibliotecas americanas, datos que no son fácilmente extrapolables a la realidad española. No son muchas las bibliotecas españolas que disponen de presupuestos para microfilmación o incluso de medios técnicos para poder realizarlo en casa. En España ha ocurrido que las empresas de microfilmación no han tenido tiempo de estabilizarse cuando se ha introducido la digitalización, mercado más competitivo y dinámico. Por todo ello no pueden ofrecerse datos de costes estables, incluso por parte de las empresas de microfilmación se está produciendo un trasvase a las nuevas tecnologías, desde luego sin abandonar su ocupación inicial. En general el mercado de la digitalización está cambiando rápidamente, mejorando constantemente su capacidad tecnológica, calidad, rapidez y precios.

Hay que añadir que los costes no deben contemplarse únicamente en función del dinero inmediato que hay que desembolsar si no en relación al uso futuro y a las posibilidades y funciones que debe ofrecer una biblioteca moderna. En cualquier presupuesto técnico, tanto de microfilmación como de digitalización, hay que incluir el coste del mantenimiento que en el segundo de los supuestos puede llegar al 20 %.

— La microfilmación es una tecnología estable, suficientemente contrastada y no es probable que sufra cambios significativos en el futuro. Existe toda una colección de normas ISO, que se ha citado anteriormente, que lo apoyan y un grupo de trabajo especializado dentro de la ISO, por el contrario la normativa sobre digitalización está comenzando y tecnológicamente está en constante avance.

— El microfilm tiene sin embargo sus límites en la consulta y distribución. El lector está obligado a usar un equipo especial de proyección en un lugar específico y es incómodo de ojear y a veces difícil de leer: se pierde la facilidad existente en el original, consulta de tabla de contenidos, sumarios, prefacio, notas a pie de página, índices, etc. Existe por ello una cierta resistencia a su uso, lo que representa un inconveniente a pagar por el beneficio de la preservación. La digitalización permite un acceso más rápido y eficaz a la información ya sea localmente o a través de redes.

---

<sup>12</sup> Véase nota 4

<sup>13</sup> Puede situarse el coste medio por fotograma en 35 pesetas.

## 2.2 *Compatibilidad de sistemas*

A pesar de esta polémica entre la microfilmación y la digitalización, tanto por parte de las instituciones o entidades que deben definir y financiar los programas, como por parte de las firmas comerciales dedicadas a ambos campos, existe un punto de vista, cada vez más extendido, que conjuga ambos sistemas<sup>14</sup>, tomando de cada uno de ellos las características más convenientes según el uso:

- microfilmación para documentos de bajos niveles de utilización y para preservación, y
- digitalización de originales o de microfilm para un acceso rápido y remoto de documentos con un nivel elevado de uso.

## 3. LOS PROYECTOS DE DIGITALIZACIÓN EN LA BIBLIOTECA NACIONAL

Durante 1993, coincidiendo con el inicio del año presupuestario, la Biblioteca Nacional comenzó un proceso de redefinición de su política de reproducción de fondos. Hasta ese momento, y salvo algún programa concreto que analizaremos más adelante, todos los esfuerzos presupuestarios en este sentido estaban destinados a la microfilmación.

Se trató en ese momento de la viabilidad de iniciar programas de digitalización, de la continuidad de los de microfilmación, así como la relación entre ambas tecnologías en el marco general de la preservación y el acceso a los fondos de la Biblioteca Nacional.

En los últimos años, la mayor parte del esfuerzo global de la Biblioteca Nacional, tanto en medios humanos como económicos, ha estado destinado a potenciar la accesibilidad de sus colecciones. Por un lado se ha dotado a la base de datos de la Biblioteca Nacional de los módulos apropiados para la catalogación en línea de los diferentes tipos de materiales bibliográficos que ingresan. Por otro lado se ha pretendido la difusión del fondo conservado, mejorando el acceso a todo tipo de fondos para atender las necesidades de los usuarios actuales y futuros.

---

<sup>14</sup> *A hybrid Systems approach to Preservation of Printed Materials (Part One)* / Don Willis. // *Microform Review*. - ISSN 0002-6530. - 22 : 4 (1993), p.168-179. Este artículo ofrece además una visión muy concisa sobre las ventajas y desventajas tanto del microfilm como de la digitalización.

Para una completa visión de la cuestión consútese también la segunda parte del artículo publicada en *Microform Review*, 23 : 1 (1994), pp. 18-25

En esta discusión se definieron como puntos claves una serie de cuestiones que afectan a todas las bibliotecas: entre ellas la discusión entre las ventajas de la microfilmación y la digitalización y la integración de los diferentes posibles proyectos en la estructura general de información y servicios de la Biblioteca Nacional: reprografía, acceso al documento, acceso a la información y control bibliográfico. A partir de todo ello la Biblioteca Nacional adoptó una política que hiciera compatible ambos sistemas.

### 3.1 *Descripción y alcance de los diferentes proyectos*

Los proyectos que a continuación se describen han ido iniciándose desde 1992 como proyectos piloto, en el caso de ADMYTE, o bien como respuesta a situaciones de demanda especiales como el caso de la Base de datos Heráldica. A partir de 1993 la Biblioteca Nacional ha dotado a estos proyectos de una línea de acción común y los ha integrado, redefiniendo el proceso de reproducción de fondos.

#### 3.1.1 *Admyte (Archivo Digitalizado de Manuscritos y Textos Españoles)*<sup>15</sup>

Este proyecto surgió inicialmente como uno de los subproyectos a realizar dentro del marco de las conmemoraciones del V Descubrimiento de América. Conformaba una colección de CD-ROM que incluye la digitalización de textos españoles básicos para el estudio de la Literatura española y que se ha iniciado con la digitalización de 63 incunables. Además las imágenes digitalizadas van acompañadas de sus respectivas transcripciones en ASCII, realizadas por expertos filólogos de distintos centros universitarios españoles y norteamericanos, por lo que la localización de términos antiguos, apellidos o topónimos se realiza en pocos segundos. Todo ello permite reducir la consulta de originales en un 95% de los casos.

ADMYTE está configurado como una colección de CD-ROMs que irán apareciendo a medida que crezca la demanda y venta del producto, es decir tiene un incremento constante.

La elección del CD-ROM como medio de difusión estuvo motivada por las siguientes razones:

- Sistema de almacenamiento estandarizado
- bajo coste de los lectores de CD-ROM

---

<sup>15</sup> *Admyte (Archivo Digitalizado de Manuscritos y Textos Españoles)*: [archivo de ordenado]. - Madrid: Sociedad Estatal Quinto Centenario: Micronet, 1993.



- el equipo necesario para su manejo es barato, popular y por tanto muy accesible
- permite el desarrollo de un proyecto de forma incremental.

La realización concreta estuvo basada en la microfilmación y posterior digitalización del microfilm y en el escaneado directo

### 3.1.2 *Base de datos heráldica*

Se trata de un proyecto de un sistema integrado de información heráldica, que se vertebra en torno a una base de datos genealógica, y que se ha desarrollado sobre una doble perspectiva:

En primer lugar, supone la satisfacción de una demanda muy frecuente por parte de los usuarios de la Biblioteca Nacional: contar con información sobre los propios apellidos, así como con reproducciones de escudos heráldicos. Las peticiones de estos escudos, mediante fotocopias, transparencias, diapositivas o fotografías, son muy numerosas.

El diseño y preparación de esta base de datos ha tenido las siguientes fases:

1. Constitución de un archivo de información textual sobre cada apellido, según campos definidos por la Biblioteca Nacional
2. Constitución de un archivo de imágenes en color con los escudos correspondientes a los apellidos
3. Vinculación de los apellidos a las imágenes correspondientes
4. Visualización en pantalla de color, de las informaciones textuales y gráficas de modo simultáneo.
5. Mantenimiento simultáneo de varias consultas
6. Obtención de copias láser en color.

Sobre esta base, en años sucesivos, se incrementará el contenido con la digitalización de los grandes repertorios genealógicos y se preparará la consulta en red y en remoto.

### 3.1.3 *Digitalización de la colección de reproducciones fotográficas*

La Biblioteca Nacional cuenta con una colección de obras microfilmadas que alcanza los 20.000 rollos de 35 mm. fundamentalmente prensa española del siglo XIX y monografías anteriores a 1830, con un incremento anual de 2.500 rollos. A partir de esta colección, y sin interrumpir los trabajos de microfilmación para preservación, se decidió iniciar la digitalización de los rollos existentes.

Igual situación es la de la colección de transparencias y fotografías, obte-

nidas a partir de solicitudes para reproducción a lo largo de los años. La frecuencia de las solicitudes nos ha inducido a incluir esta colección entre los documentos a digitalizar, ya que la digitalización nos permitiría una mayor agilidad en el servicio de este tipo de reproducciones.

En total la cantidad de las reproducciones fotográficas con la que se inicia la digitalización en la Biblioteca Nacional es la siguiente:

- B) 2.000 transparencias en color de material cartográfico.
- C) 20.000 transparencias de grabados en escalas de 256 grises
- D) 2.000.000 de fotogramas de microfilm en blanco y negro.
- E) 20.000 páginas de códigos en microfichas de color.

### 3.1.4 *Digitalización de originales*

Del mismo modo y como ya se ha comentado hay que iniciar la digitalización directa de materiales originales, comenzando por los dibujos originales en color y sonido, de modo que conjuntos documentales completos puedan recibir este tratamiento:

- A) 10.000 dibujos originales en color.
- B) 10.000 minutos de sonido.

## 3.2 *Requisitos para la digitalización*

### 3.2.1 *Descripción de los documentos*

La conversión de cada lote comprende la producción de cada uno de los ficheros y su vinculación a su descripción bibliográfica y a los datos locales de los distintos ejemplares de la misma existentes en la Biblioteca Nacional. El formato USMARC así como el formato IBERMARC tienen previstos un conjunto de códigos que permiten la descripción de los contenidos de los archivos de ordenador y, naturalmente, la descripción bibliográfica de los mismos está basada en las ISBD (CF)<sup>16</sup>.

Partiendo de estos datos es posible llevar a cabo una indexación dentro del sistema que permita referenciar ambas entidades, siendo una el archivo digital que residirá en una memoria auxiliar ad hoc y la otra, la descripción del referido archivo, que formará parte de la base de datos bibliográfica.

---

<sup>16</sup> ISBD (CF) : *Descripción Bibliográfica Internacional Normalizada para Archivos de Ordenador* / recomendada por el Grupo de Trabajo de la Descripción Bibliográfica Internacional Nor-

### 3.2.2 Niveles de resolución

La calidad de la imagen viene determinada por el tono o número de bits utilizados para representar cada punto y por la resolución o número de líneas y puntos por línea a escanear.

Los escáner usan esquemas de codificación de diferente complejidad, desde el esquema de dos tonos hasta el color real. Con una codificación de 4 bits se puede representar 16 gamas de grises ( $2^4$ )<sup>17</sup>, un esquema de 8 bits permite representar 256 niveles de gris ( $2^8$ ). La digitalización del color puede requerir todavía más bits por punto, 36 bits por punto<sup>18</sup> (12 bits para cada uno de los colores básicos: rojo, azul y verde), a lo que se denomina "true colour" y es esencial a la hora de reproducir en artes gráficas.

El número de líneas y puntos por línea a escanear es otro factor determinante de la calidad de la imagen. La resolución expresa el nivel de detalle de las imágenes digitalizadas y por tanto es otro factor de su calidad, varía en función de la densidad de los puntos con la que se trabaja. El poder de resolución de los escáneres se expresa normalmente en puntos por pulgada (ppp.) (o dpi.: dots per inch). Los escáneres actuales permiten resoluciones entre 75 y más de 1.000 ppp. Podemos definir los niveles de resolución como sigue:

- Muy baja, hasta 150 ppp., adecuada para documentos muy claros y sin detalles o textos pequeños y consultas cortas.
- Baja, en torno a los 300 ppp., adecuada para una gran variedad de documentos, es la resolución de la mayor parte de las impresoras láser y fotocopiadoras. Las imágenes digitalizadas a 400 ppp. se aproximan mucho al original y se aplican a artículos de revista, contratos, formularios impresos, etc.
- Media, en torno a los 600 ppp., proporciona muy buena presentación, resuelve los detalles más finos.
- Alta, en torno a los 1.000 ppp.
- Muy alta, por encima de los 1.000 ppp., usada cuando se requiere una absoluta fidelidad en la reproducción de escala de grises y color, p. e., fotografías.

---

malizada para Archivos de Ordenador ; establecida por el Comité de Catalogación de la IFLA ; traducción de María Luisa Martínez-Çonde, María Jesús Santurtún y Xavier Agenjo ; revisión de Justo García Melero y Félix Moya. - Madrid : ANABAD : Arco/Libros, 1994.

<sup>17</sup> Este sería el caso del proyecto de digitalización del Archivo General de Indias.

<sup>18</sup> La Biblioteca Vaticana ha iniciado un gran programa de automatización, en el que se incluye la creación de un banco de datos de imágenes de libros, manuscritos y miniaturas. El esquema de escaneo será de 36 bits, con una resolución de 3072 x 4000. A lo largo del desarrollo del proyecto decidirán la forma de acceso: CD-ROM, Internet.

Véase: *CIDOC Multimedia Working Group Newsletter*. - 1 : 1 (1994), p. 3.

No en todos los casos el uso de una resolución alta es lo más conveniente, en ocasiones es necesario un tratamiento cuidadoso, del mismo modo que una fotocopiadora degrada el papel ocurre lo mismo en el escaneado. En otras ocasiones, la combinación de una resolución adecuada y la escala de grises permite obtener resultados apropiados a los documentos a tratar y a los objetivos del sistema de imágenes establecido. Los niveles de resolución con los que trabajar hay que definirlos en función del documento y del servicio que se pretende dar (copia de preservación, acceso al documento local o remoto, información, etc.)<sup>19</sup>, además cuanto mayor es la resolución, mayor es la cantidad de puntos a codificar y por tanto las necesidades de almacenamiento son también mayores.

Los niveles de resolución a los que trabajará la Biblioteca Nacional serán los siguientes:

Dibujos originales en color: 1.200 ppp.

Transparencias de material cartográfico: 1200 ppp.

Transparencias de grabados: 600 ppp.

Fotogramas de microfilm en blanco y negro: 300 ppp.

Fotogramas de microfichas en color: 2.400 ppp.

El planteamiento general es, que aunque estos niveles de resolución no permitan la visualización en pantalla, resultan imprescindibles para la reproducción. La visualización y consulta se realizarán a través de una memoria tampón con unos niveles de resolución más bajos, observables en la pantalla. Sin embargo, para la impresión se extraerán los ficheros, con los niveles de resolución mencionados anteriormente, se transmitirán a las diferentes salidas: impresora de sublimación, disquetes, discos ópticos magnéticos (ISO 10090) o ektachrome.

La digitalización de sonido se realizará con una frecuencia mínima de muestreo de 32 KHz. La reproducción del audio se puede efectuar con calidades muy dispares, la máxima calidad es la usualmente conocida como *calidad CD*<sup>20</sup> por corresponder con la implementada en la grabación de discos compactos musicales (Compact Disc) que corresponde con una frecuencia

---

<sup>19</sup> Don Willis, *ob. cit.*, define los diferentes niveles de resolución de la siguiente manera (la traducción es nuestra):

“Resolución para archivo permanente” es definida como la resolución necesaria para capturar una réplica exacta del original independientemente de su coste.

“Resolución óptima para archivo permanente” es la resolución más baja que puede satisfacer completamente los objetivos definidos por un sistema para el archivo de imágenes.

“Resolución adecuada para el acceso”, del orden de 300 ppp. binario, se define como la resolución suficiente para capturar el 99,9 por ciento de la información contenida en una página.

<sup>20</sup> 176 Kbytes por segundo en señal estéreo.

de muestreo de 44,4 Khz y resolución de 16 bits por muestra en cada uno de los dos canales del estéreo. La implementación de esta calidad es la deseable en reproducción de piezas orquestales, pues supone una perfecta calidad para todo el rango de frecuencias audible. Deberá tenerse en cuenta el uso de técnicas de compresión, dado el elevado número de muestras resultante del uso de calidad CD. El formato de almacenamiento deberá soportar el algoritmo de compresión seleccionado.

### 3.2.3 *Codificación y compresión*

Una vez escaneada y tratada una imagen se obtiene un fichero que contiene la imagen digitalizada. Los ficheros de imágenes son muy grandes por lo que antes de grabarlos o transmitirlos es necesario efectuar un proceso de compresión que reduzca su tamaño original. Sin este procedimiento las necesidades de memoria de almacenamiento serían mucho mayores y el funcionamiento en general del sistema mucho más lento y costoso. Por ejemplo, un documento tamaño A-4 digitalizado a 200 ppp. de resolución y con un esquema de codificación de 2 bits necesitaría 0,5 Mbytes de memoria de almacenamiento. Comprimiéndolo puede reducirse en una relación de 15:1, es decir ocuparía 30 Kbytes con lo que no sólo se reduce el espacio de almacenamiento sino el tiempo de transmisión.

Existen diferentes métodos de compresión/descompresión que se aplican a los sistemas de gestión electrónica de documentos. La mayor parte de los sistemas utilizan los algoritmos desarrollados por los grupos 3 y 4 del Comité Consultivo Internacional Telegráfico y Telefónico (CCITT) de la Organización de Naciones Unidas que fueron desarrollados para reducir la duración de la transmisión de imágenes a través de los circuitos telefónicos.

El algoritmo de compresión del grupo 3 emplea el método de codificación unidireccional de Huffmann. Según este método la imagen es codificada línea a línea, el algoritmo de Huffmann atribuye códigos a secuencias cortas de puntos predeterminadas, por ejemplo la secuencia de 5 puntos blancos seguidos de 5 puntos negros puede ser transmitido por la traducción del código Huffmann correspondiente. Las unidades de recepción y de emisión poseen un diccionario que transforman las secuencias de puntos en códigos Huffmann que son los que únicamente se transmiten. Ofrece índices de compresión de 10:1 en documentos de oficina y de 15:1 en dibujos y planos.

El algoritmo de compresión del grupo 4 se basa en el hecho de que los puntos blancos o negros de una línea dada están a menudo rodeados de puntos de idéntica tonalidad. La mayor parte de las imágenes se modifican gradualmente cuando se pasa de una línea digitalizada a la siguiente. Por ello, según las recomendaciones del grupo 4 se registran las pequeñas modificaciones entre líneas, no las líneas enteras, y comprime las imágenes digitali-

zadas tanto horizontal como verticalmente. Ofrece índices de compresión de 15:1 para documentos de oficina y de 20:1 para planos y dibujos.

Como se ha dicho la mayoría de los sistemas de archivo óptico soportan las técnicas de ambos grupos. Sin embargo, para hacer un uso más efectivo de las capacidades de disco es preferible la compresión del grupo 4. A medida que la complejidad y la densidad de la imagen aumenta la eficacia de la compresión decrece. La compresión del grupo 4 es generalmente más eficaz que la compresión del grupo 3, pero presenta problemas a la hora de comprimir imágenes con una escala de grises grande o color<sup>21</sup>.

Las recomendaciones definidas por el CCITT conllevan también informaciones complementarias que sirven para describir los documentos y las técnicas de compresión. Determinados códigos son utilizados para caracterizar los límites de los datos numéricos que representan la imagen, la técnica de compresión utilizada, la velocidad de transmisión y la resolución de la imagen.

El tipo de algoritmo de compresión depende del tipo de documento, de su tamaño y de la resolución. Mucho más utilizables para el tratamiento de imágenes de alta calidad son los algoritmos de compresión:

- MPEG (Motion Picture Experts Group)
- JPEG (Joint Photographic Experts Group)

Este último, está especialmente indicado para las imágenes en color, sobre todo permite la digitalización en 16.8 millones de colores, lo que resulta imprescindible no sólo para la visualización de la imagen por pantalla sino para posteriores usos de impresión, pues permite llegar a obtener copias ektachrome.

La codificación y digitalización se realizará de acuerdo con los formatos TIFF y algoritmo de compresión CCITT G.4 para imágenes monocromas, JPEG (ISO/IEC 10918-1) para imágenes en color y escalas de grises, y formato AIFF y algoritmo de compresión CCITT G.7111 y G.722 para audio.

### 3.2.4 Almacenamiento

El formato TIFF (Tagged Image File Format), proporciona el nivel más alto de normalización necesario para permitir el intercambio de imágenes.

La adquisición digital de un manuscrito o de una pieza musical supone un elevado volumen de información para poder ser reproducido con un nivel de calidad aceptable. Los requerimientos de calidad CD para el audio y color real para imágenes son los exigidos en la actualidad para este tipo de sistemas.

---

<sup>21</sup> El Joint Bi-level Group (JBIG) ha preparado un tipo de compresión para imágenes de dos tonos que mejora el método del CCITT, denominado con las abreviaturas del nombre del grupo, JBIG. Está formado por miembros de ISO y de CCITT.

Si analizamos, de un modo orientativo, los volúmenes de almacenamiento para una imagen o una pieza musical, obtenemos:

— Imagen de 1280 x 1024 puntos en color real supone aproximadamente 3,75 Mbytes.

— Una pieza musical de 3 minutos supone 30,5 Mbytes.

Considerando los fondos de la Biblioteca Nacional estimados en:

22.580 manuscritos,

500.000 grabados,

400.000 postales,

250.000 discos,

100.000 cassettes,...

se puede calcular un volumen total de información que superará los cientos de terabytes. Utilizando técnicas de comprensión, como JPEG, pueden ser reducidas en una proporción de 30 a 1, lo cual nos llevaría a decenas de terabytes, es decir, a decenas de millones de megas.

Dado el elevado volumen de información y que ésta no va a sufrir modificaciones, es lógico considerar la utilización de medios ópticos de almacenamiento tipo WORM.

La tecnología óptica actual permite la utilización de discos de varios gigabytes; para abordar capacidades de terabytes se pueden buscar soluciones que mecanicen el acceso a colecciones de discos ópticos, soluciones del tipo *juke box*.

La optimización del almacenamiento será una de las prioridades que se establecerán al efectuar el diseño del sistema, así como la elección del algoritmo de compresión óptimo permitirá una reducción sustancial de las dimensiones del sistema de almacenamiento.

El controlador de almacenamiento proporciona una interfaz entre el almacenamiento en disco y el servidor de imágenes. Traduce las solicitudes de imágenes en comandos que localizan, recuperan las imágenes de los dispositivos de almacenamiento (*jukebox*, disco magnético, etc.). Para las peticiones de almacenamiento de imágenes el controlador debe localizar espacio, almacenar apropiadamente los archivos y actualizar el directorio de archivos. El archivo de estructura del documento contiene información sobre la indicación y el servidor de imágenes enlaza el archivo de la estructura del documento con el controlador del almacenamiento.

### 3.2.5. Comunicaciones

Se están realizando intentos para la transmisión de imágenes a niveles de resolución mucho más bajas de las consideradas por la Biblioteca Nacional a

través de Internet, que como se sabe utiliza el protocolo TCP/IP. De hecho algunos de los experimentos que se han llevado a cabo con la digitalización realizada en el Archivo de Indias con escalas de 16 gamas de grises. Parece evidente que el futuro está encaminado a la transmisión de imagen y sonido a través de la red digital de banda ancha (RDSI) que permite velocidades de transmisión muchísimo más altas, y que por lo tanto pueden enfrentarse con los enormes volúmenes que pueden alcanzar determinados ficheros. En lo que se refiere a protocolos de redes está claro que es necesario producir un "upgrade" de los actuales sistemas Ethernet a FDDI, protocolo que como es sabido se inscribe en la arquitectura ISO/OSI.

### 3.2.6 Acceso

En cuanto a la reproducción, deberían evaluarse impresoras de sublimación de cera, como la ya adquirida para el sistema Heráldico, para las copias de papel y monitores de alta resolución en RGB analógico (con resoluciones de máquinas para diseño gráfico, 1280 x 1024 puntos en 17-19 pulgadas).

La elección del terminal vendrá determinada por tres aspectos que deberán ser evaluados:

- Medios que se presentan: información textual, gráfica, audio.
- Calidad de la presentación: resolución y colores en imágenes, resolución y muestreo en audio.
- Usuario: usuario externo, operador entrenado.

Según el tipo de servicio el terminal deberá posibilitar la presentación de información textual, gráfica y/o audio. El tipo de terminal más sencillo corresponderá con aquellos que presentan información solamente textual; mientras que el más complejo corresponde con los que implementan audio+vídeo.

Respecto a la calidad de la presentación, un terminal de gama alta con posibilidad de presentar imágenes en alta resolución y color real, además de audio en calidad CD, puede ser seleccionado en el mercado de entre una extensa gama de máquinas comerciales. Dependiendo del software básico utilizado, la solución existe desde la gama PC hasta las estaciones de trabajo Unix.

## 4. ARQUITECTURA

La realización de las especificaciones del desarrollo de la arquitectura necesaria para cubrir los servicios de almacenamiento y distribución de la in-



formación, así como los mecanismos de impresión, grabado y presentación, requieren de una labor de consultoría en la que se determinará la plataforma necesaria para la implementación de dichos servicios y además se establecerán las normas de comunicaciones que se precisen.

El diseño se realiza en torno a una red de fibra óptica. Esto abre el camino hacia la implementación de modelos cliente-servidor para todos aquellos servicios de consulta que se realicen al sistema. Será necesario establecer el tipo de servicios de comunicaciones necesario. Dependiendo del entorno será necesario un protocolo de comunicación interproceso (LU 6.2. en entorno SNA, RPC en entorno UNIX, etc.).

Elementos servidores serán aquellos que almacenen un determinado tipo de información del sistema u ofrezcan un servicio al resto de elementos. Un centro servidor será la máquina Fujitsu, destinada a aquellas consultas sobre los fondos bibliográficos de la Biblioteca Nacional, así como para las referencias necesarias para la recuperación de imágenes y audio en otros elementos servidores<sup>22</sup>.

El sistema de gestión de las imágenes de los grabados y de las digitalizaciones musicales será uno de los elementos a definir conjuntamente con los terminales de consulta, pues representan una de las funcionalidades más relevantes del sistema.

Una posible solución podría ser la siguiente:

Utilizar estaciones de trabajo Unix de bajo coste funcionando como X-terminal. Se utilizarían frame-buffers de 24 bits y sonido en calidad CD con interfaz gráfica Open Look o Motif. Otra estación, o una de las anteriores, operaría como concentrador de las peticiones de todos los terminales.

Los accesos serían canalizados como peticiones al gestor Adabas mediante servicios de comunicaciones interproceso (LU 6.2). Si la consulta requiere la reproducción de la imagen de un grabado o la reproducción de una pieza musical, se obtendrá de la base de datos las direcciones de almacenamiento de éstas en el sistema servidor de las imágenes y el sonido. En el caso de haber optado por una solución juke box de discos ópticos será necesario obtener el número del disco y su posición dentro del disco.

Cuando una consulta involucre reproducción impresa o grabación en cinta, se realizará una petición de dicho servicio al equipo controlador en cada caso.

Habrà que definir servicios de acceso remoto así como las comunicaciones con otras entidades. Servicios, cuya especificación, determinará la utilización de la red digital Iberpac X.25 y/o la red telefónica conmutada. X.25

---

<sup>22</sup> En el momento de realizarse la corrección de pruebas de este artículo, la Biblioteca Nacional está inmersa en un proceso de cambio de su sistema informático, que supondrá la migración a un entorno UNIX.

ofrece la posibilidad de implementar soluciones de muy alto nivel, como correo electrónico mediante norma OSI X.400. Sobre la red telefónica conmutada se pueden establecer servicios de consulta de bajo coste, con sencillas emulaciones de terminal (Figura n.º 4: esquema completo de lo que podría ser el acceso global a la información y al documento).

Por último, queremos añadir a todo lo expuesto, que la Biblioteca Nacional está dando los primeros pasos en las técnicas de tratamiento de los documentos en texto completo, utilizando las fórmulas del SGML, para reutilizar estructuradamente la información que recibe de periódicos, revistas y monografías, vía Depósito Legal, no ya en formato papel, sino a partir de las cintas y ficheros de fotocomposición, lo que redundará en unas enormes bases de datos textuales accesibles remotamente. Este procedimiento permitirá obviar tratamientos de digitalización o ICR para los ingresos en curso, reduciendo al mismo tiempo los recursos económicos y humanos a emplear.

La experiencia adquirida por la Biblioteca Nacional en el desarrollo de nuestro videodisco interactivo, los puestos de autoservicio de información, los CD-ROM de referencias o de texto completo, o nuestro sistema integrado de información heráldica, nos ha permitido ahondar en la aplicación de las técnicas multimedia. Apoyándonos en ellas será posible que la información tradicional, que en tan gran número almacena la Biblioteca Nacional, se interrelacione de tal modo que el usuario final pueda navegar por la información de forma tal que se logre nuestro objetivo de hacer que la información esté disponible para todo el mundo en cualquier sitio.

A continuación se señalan algunas normas relativas a la digitalización:

*ISO 12653: Electronic Imagery - Test Targets for the Black and White Scanning of Office Documents*, esta norma especificará el método para establecer la calidad en el proceso de escaneado. Su objetivo es definir tanto las cartas de pruebas, como los métodos de medida de la calidad de los escáneres; sin embargo, la norma no establecerá niveles a alcanzar.

La norma TWAIN, desarrollada en 1992, por un grupo de productores (Aldus, Caere, Kodak, HP y Logitech) va dirigida a establecer los requisitos que debe cumplir la interfaz entre los escáneres y los programas de aplicación.

*British Standard 7768: 1994 Recommendations for Management of optical disk (WORM) systems for the recording of documents that may be required as evidence*, define los procedimientos normativos para el escaneado y almacenamiento de imágenes en discos WORM que garanticen su aceptación como pruebas documentales.

Otra norma de interés es la que está desarrollando el Grupo de trabajo ISO (TC171/WG5) para establecer un vocabulario común sobre imágenes, digitales o micrográficas.