

*DEL MICROFILME
A LA IMAGEN
DIGITAL*

*Informe de la
Yale University Library
por DONALD J. WATERS*

*BIBLIOTECA NACIONAL
DE VENEZUELA
CENTRO NACIONAL
DE CONSERVACIÓN
DE PAPEL
CENTRO REGIONAL IFLA-PAC
PARA AMÉRICA LATINA
Y EL CARIBE
COMISIÓN DE PRESERVACIÓN
Y ACCESO
COUNCIL ON LIBRARY
AND INFORMATION RESOURCES*

Caracas, Venezuela

**BIBLIOTECA NACIONAL
DE VENEZUELA**



**CENTRO NACIONAL
DE CONSERVACION DE PAPEL
CENTRO REGIONAL IFLA-PAC
PARA AMERICA LATINA Y EL CARIBE**

Edificio Rogi, Piso 1
Calle Soledad con Calle Las Piedritas
Zona Industrial de La Trinidad
Caracas, Venezuela
Telefax: (58-2)-941.4070
Central: (58-2)-941.8011 (x 203, 218)

CONSERVAPLAN

**Documentos para Conservar
Nº 9, 1998**

*Del microfilme a la imagen digital
por Donald J. Waters*

**Derechos reservados por
Commission on Preservation and Access
Washington, D.C. 1991**

**Para los países de habla hispana,
por la Biblioteca Nacional de Venezuela
1998.**

2

Este programa recoge y disemina en traducción al español documentos significativos de la literatura de conservación aparecida en otros idiomas y cuya lectura es recomendada en los programas de formación. La ausencia de publicaciones actualizadas en español, sobre conceptos, historia y técnicas, ha frustrado el nivel y calidad de la conservación en países hispanoparlantes. **Conservaplan** ha sido creado para proporcionar apoyo bibliográfico en temas fundamentales. Los interesados en suscribirse y en realizar propuestas para la serie podrán dirigirse al Editor de Conservaplan, a la dirección arriba señalada.

©Instituto Autónomo Biblioteca Nacional 1998

Hecho el depósito de ley
Depósito legal LF227199802516
LF227199802516.9

ISSN 1315-3579 (Conservaplan)
ISBN 980-319-145-4

PRESENTACIÓN

La Biblioteca Nacional de Venezuela, en su carácter de Centro Regional IFLA-PAC para América Latina y El Caribe, y como promotora y responsable del curso de "Conservación de obras gráficas", dirigido a empleados de las bibliotecas nacionales y archivos de Latinoamérica, ha percibido la enorme importancia de contar con información técnica actualizada que oriente a los conservadores y responsables de bibliotecas y archivos de la región en su constante esfuerzo por preservar en el tiempo sus diversas, y muchas veces valiosísimas, colecciones de material bibliográfico y audiovisual.

Hasta hace poco menos de un lustro, casi nada de la información existente sobre preservación de materiales de bibliotecas y archivos publicada por reconocidas instituciones archivísticas, centros de investigación y especialistas en la materia se encontraba en español. Actualmente, aparte de UNESCO, muchas organizaciones están realizando aportes en este sentido. En el marco de este esfuerzo, el Centro Nacional de Conservación de Papel de la Biblioteca Nacional de Venezuela publica desde 1987 *Conservoaplan*, un instrumento de divulgación dirigido a profesionales y técnicos hispanohablantes en el área de la conservación. El presente número de *Conservoaplan* forma parte de un proyecto de traducción, y de disseminación en dieciocho fascículos, de ocho títulos en inglés sobre preservación de material bibliográfico y no bibliográfico, iniciado en 1996 y desarrollado con la coparticipación de la Comisión de Preservación y Acceso, programa internacional del Council on Library and Information Resources con sede en Washington D.C. Este proyecto se complementa con uno similar recientemente culminado en Brasil, y que pone a disposición estos temas en portugués para profesionales en conservación y responsables de colecciones de ese país.

En este logro ha sido fundamental el apoyo de Hans Rütimann, responsable del Programa Internacional de la Comisión de Preservación y Acceso, en quien, desde su primera visita a Latinoamérica en 1989, hemos encontrado una receptividad y un empeño excepcionales en beneficio de proyectos orientados hacia este objetivo.

Este documento, *Del microfilme a la imagen digital*, contribuye a facilitar nuestra comprensión sobre la capacidad de uso de la tecnología digital para preservar el acceso a materiales valiosos, inestables o en proceso de deterioro. El análisis presentado permite evaluar la pertinencia de emplear el cúmulo de materiales ya preservados a través del microfilme para aprovechar las ventajas, en cuanto a disseminación y acceso, que ofrecen los sistemas de digitalización, los cuales hacen posible al usuario de una biblioteca, e incluso, al de otra institución remota, hojear, seleccionar, imprimir o almacenar información desde su computador personal, gracias a las redes que interconectan los sistemas automatizados y las bases de datos de la biblioteca con los de otras instituciones.

Centro Nacional
de Conservación de Papel
de la Biblioteca Nacional de Venezuela

Datos de la versión original en inglés:

*From Microfilm to Digital Imagery :
on the feasibility of a project to study
the means, costs and benefits of
converting large quantities of preserved
library materials from microfilm to
digital images / a report of the Yale
University Library to the Commission
on Preservation and Access
by Donald J. Waters, Head, Systems
Office, Yale University Library.
New Haven, Connecticut 06520*

Copyright ©1991 por Commission on
Preservation and Access
Todos los derechos reservados

Edición en español :

*Del microfilme a la imagen digital:
sobre la factibilidad de un proyecto para
estudiar los medios, costos y beneficios
de convertir a imágenes digitales
grandes cantidades de materiales
bibliográficos preservados en
microfilme / informe de la Yale
University Library para la
Commission on Preservation and
Access por Donald J. Waters,
Jefe, Oficina de Sistemas,
Yale University Library*

Biblioteca Nacional de Venezuela
con la autorización de la
Comisión de Preservación y Acceso
del Council on Library and Information
Resources
Caracas, 1997-1998

ISSN 1315-3579 (Conservaplan)
ISBN 980-319-145-4

Coordinación y revisión:
**Centro Nacional de Conservación de Papel
Centro Regional IFLA/PAC
para América Latina y el Caribe**
Calle Soledad con Calle Las Piedritas
Edificio Rogi, 1er. piso
Zona Industrial de La Trinidad
Caracas, Venezuela
Telefax: (582)-941.4070

Comité Editor:
Virginia Betancourt, Lourdes Blanco,
Aurelio Álvarez

Comité Coordinador:
Pedro Hernández, Adelisa Castillo V.,
Ramón Sánchez, Pía Rodríguez

Traducción:
Ana Margarita González

Composición electrónica:
Adelisa Castillo V.

Impresión:
Editorial EX-LIBRIS, Caracas

Waters, Donald J.

Del microfilme a la imagen digital / informe de la
Yale University Library para la Comisión de
Preservación y Acceso por Donald J. Waters, Oficina
de Sistemas, Yale University Library ; coordinación y
revisión técnica, Centro Nacional de Conservación de
Papel/Centro Regional IFLA/PAC para América Latina
y el Caribe. — Ed. en español. — Caracas : Biblioteca
Nacional de Venezuela, 1998.

42 p. : il. ; 28 cm. — (Conservaplan. Documentos
para conservar ; n° 9)

Proyecto financiado por la Commission on
Preservation & Access, Council on Library and
Information Resources.

Traducción de: From Microfilm to Digital
Imagery.
ISBN 980-319-145-4

1. Digitalización de imágenes. 2. Digitalización de
información. 3. Microfilms. I. Yale University Library.
Systems Office. II. Biblioteca Nacional (Venezuela).
Centro Nacional de Conservación de Papel. III. Título.

Edición
de la versión
original en inglés
de junio de
1991

Donald J. Waters
Jefe, Oficina de Sistemas
Yale University Library

Biblioteca Nacional
de Venezuela
Centro Nacional de
Conservación de Papel
Centro Regional
IFLA/PAC
para América Latina
y el Caribe

Comisión de
Preservación y Acceso
Council on Library
and Information
Resources

Caracas, 1998

Del Microfilme a la Imagen Digital

*Sobre la
factibilidad de un
proyecto para estudiar los
medios, costos y beneficios de
convertir a imágenes digitales
grandes cantidades de
materiales bibliográficos
preservados en
microfilme*

*Informe
de la Yale
University Library
para la Commission on
Preservation and Access
por Donald J. Waters*

PREFACIO DE LA COMISIÓN

En un informe anterior de la Comisión, Michael Lesk, miembro del Comité Asesor sobre Evaluación de Tecnologías, señala que la conversión de microfilme a imágenes digitales para la preservación y consulta de materiales bibliográficos en condición de deterioro es aconsejable a largo plazo, ya que en la actualidad es técnicamente posible y relativamente barato de realizar (*Image Formats for Preservation and Access*, July 1990, p. 8). Sin embargo, la tecnología de conversión de documentos en imágenes aún está en desarrollo, las normas en cuanto a los equipos y programas de computación continúan evolucionando y la conversión del microfilme al formato de imágenes digitales todavía no ha sido comprobada en verdaderos ambientes de bibliotecas.

En este momento, pareciera decisivo que la investigación y las inversiones fueran orientadas a desarrollar sistemas para manejar la conversión digital, el almacenamiento y el acceso que puedan ser utilizados por las bibliotecas universitarias y de investigación. Para desarrollar esos sistemas la Yale University Library orientó sus esfuerzos hacia este estudio de planificación.

El estudio, realizado bajo contrato con la Comisión, explora la factibilidad de un proyecto para estudiar los medios, costos y beneficios de convertir a imágenes digitales grandes cantidades de materiales bibliográficos preservados en microfilme. Esta labor señala la necesidad de que se lleve a cabo un proyecto multianual, de gran envergadura, para convertir dichos materiales del microfilme al formato digital, dando acceso tanto intrainstitucional como interinstitucional a las imágenes almacenadas, así como para investigar las amplias implicaciones de brindar un mayor acceso intelectual a los materiales de investigación digitalizados. La Comisión espera que los resultados constituyan una importante contribución a nuestra capacidad de utilizar las nuevas tecnologías de digitalización con fines de

preservación. Ejemplares complementarios de este informe han sido distribuidos de acuerdo con la lista de correos de la Comisión.

Publicado por
Commission on Preservation and Access
1785 Massachusetts Ave, NW Suite 313
Washington, DC 20036
Junio 1991

CONTENIDO

Introducción	8
Visión	8
El valor del acceso	9
La biblioteca accesible	10
* Acceso intelectual y físico	11
* Colecciones y renovación de colecciones	11
Preservando el acceso mediante la preservación	12
La digitalización de imágenes	14
Alternativas económicas	16
La conversión desde el microfilme como objetivo	17
Un modelo de inversión progresiva	18
● El sistema básico	18
● Impresión por solicitud	19
● Revisión en línea	20
● Acceso remoto	20
● Revisión a nivel de página	21
● Revisión a nivel de la estructura del documento	21
Arquitectura del sistema	22
Normas	23
Componentes del sistema	24
● Conversión	24
● Almacenamiento	27
● Acceso	28
● Reproducción	29
Aspectos relacionados con la factibilidad	30
Plan de trabajo	32
Metas del proyecto	32
Alcance del proyecto	32
Fase 1: Organización	33
Fase 2: Preparación	34
Fase 3: Conversión	35
Fase 4: Acceso a la biblioteca	36
Fase 5: Acceso remoto	38
Fase 6: Finalización	39
Conclusión	39
Notas	40

Tablas e ilustraciones

Tabla 1	
Uso de los modos para obtener información	10
Ilustración 1	
Modelo de biblioteca orientada al acceso	11
Ilustración 2	
Proceso de la renovación de colecciones	12
Ilustración 3	
Preservación en microfilme	13
Ilustración 4	
Digitalización de imágenes en la biblioteca	15
Ilustración 5	
Conversión de microfilme de preservación a imágenes digitales	17
Ilustración 6	
Arquitectura de los sistemas de digitalización de imágenes	24
Ilustración 7	
Fases del proyecto	33

DEL MICROFILME A LA IMAGEN DIGITAL

INTRODUCCIÓN

8 Durante las últimas tres décadas, el microfilme se ha convertido en el medio por excelencia para preservar materiales bibliográficos susceptibles de deterioro. El microfilme es duradero, siempre y cuando sea almacenado en un ambiente controlado. Las bibliotecas y los archivos siguen procedimientos y especificaciones estándares, e incluso se sirven de instalaciones comunes en el proceso de microfilmación. Los costos del proceso son predecibles, variando entre 10 y 15 centavos de dólar por página, sin incluir los costos generales de administración y de selección de materiales a ser microfilmados. Por otra parte, la tecnología para tener acceso al microfilme es estable y no es probable que sufra grandes cambios en el futuro. En pocas palabras, por la solución que brinda al problema de los "libros quebradizos", que actualmente azota a los repositorios de obras impresas en papel ácido, el microfilme esencialmente está exento de riesgos.¹

Michael Lesk, Gerente de la División de Investigación en Ciencias Informáticas en Bellcore, en un reciente informe presentado a la Commission on Preservation and Access, ratificó el uso del microfilme para fines de preservación hoy en día. Sin embargo, asumió una posición relativamente cautelosa con respecto a su papel en el futuro, alegando que, con seguridad, el microfilme no será el final del proceso de preservación. Sostuvo, por el contrario, que es "un paso intermedio razonable para llegar a la digitalización de imágenes", el cual consiste en un proceso de escanear libros y almacenarlos en la memoria de un computador.²

Aunque Lesk reconoció que la tecnología para digitalizar imágenes es costosa y "todavía no se ha afianzado", sostuvo con firmeza que la tecnología encierra una promesa sumamente importante para la misión bibliotecaria de generar, preservar y mejorar el acceso al conocimiento registrado. Lesk hizo notar que

en la actualidad la conversión del microfilme a la imagen digital es técnicamente posible y argumentó que los usuarios de las bibliotecas demandarán, a largo plazo, los resultados de dicha conversión. Predijo que la tecnología para digitalizar imágenes, incluida la capacidad de conversión a partir del microfilme, "estará al alcance de la mayoría de las bibliotecas dentro de una década", e instó a que se invirtiera e investigara hoy en el "desarrollo de sistemas que puedan ser utilizados por bibliotecas comunes" en el futuro.³

La Yale University Library, una de las pioneras en desarrollar las prácticas de preservación en microfilme, actualmente está comprometida en las primeras etapas de un exhaustivo esfuerzo a largo plazo para desarrollar sistemas prácticos para convertir materiales microfilmados a imágenes digitales y, por consiguiente, llegar a comprender la debida relación entre las dos tecnologías para las futuras prácticas de preservación. Este informe resume los resultados de la fase inicial de planificación de un proyecto de mayor envergadura. En la siguiente sección, se comienza a articular una visión del lugar que tiene la tecnología de digitalización en una biblioteca de investigación como la de Yale. Luego, a grandes rasgos, se indican las alternativas económicas y de organización, en las cuales la conversión de materiales microfilmados a imágenes digitales pudiera surgir como una opción de preservación viable. El informe describe la arquitectura de un sistema de digitalización de imágenes e identifica puntos críticos referentes a la factibilidad. Finalmente, presenta un plan de trabajo para un proyecto de demostración, que investigará en forma exhaustiva los costos y beneficios que implica la conversión en el proceso de conformar una biblioteca digitalizada de 10.000 libros a partir de materiales microfilmados.

VISIÓN

Para tener una visión a largo plazo del papel que desempeñan las bibliotecas de investigación en la universidad, se debe analizar de cerca y cuidadosamente las diversas formas en que los investigadores trabajan

para crear y diseminar el conocimiento. Numerosos estudios recientes enfatizan que la calidad y la cantidad de la productividad académica en todas las disciplinas dependen enormemente del acceso rápido y oportuno a la información pertinente. La biblioteca del futuro, para tales investigadores, no será necesariamente una "biblioteca electrónica", como tampoco será únicamente su misión la de preservar el conocimiento registrado. La biblioteca del futuro se orientará a generar, preservar y mejorar para sus usuarios el rápido acceso -tanto intelectual como físico-, del conocimiento registrado. Las tecnologías emergentes, como la imagen digital, encontrarán allí un lugar crítico, en la medida en que ayuden a respaldar la misión general de las bibliotecas.

El valor del acceso

A principios de 1985, el Research Libraries Group estableció un Programa especial sobre el Manejo de la Información de Investigación (PRIMA, por su nombre en inglés, *Program for Research Information Management*). El programa dio como resultado una evaluación detallada de las necesidades de información en más de veinte disciplinas en tres amplios grupos: las humanidades, las ciencias sociales y las ciencias. Entre las diversas necesidades que surgieron de la evaluación, muchas fueron específicas para las disciplinas que estaban siendo estudiadas. Un tema, sin embargo, aparece en todos los estudios y se aplica a todas las disciplinas. Los investigadores, según las palabras del informe sobre las necesidades de información en humanidades, "prefieren los métodos familiares y no sistemáticos de consultar a colegas y notas a pie de página".⁴ No se aportan evidencias en los informes para respaldar la caracterización de estos métodos como "no sistemáticos", pero sí queda muy claro que, para los investigadores, el valor de consultar las notas a pie de página y a sus colegas radica en que estos métodos son rápidos y de fácil acceso.

Un estudio patrocinado por el Faxon Institute for Advanced Studies in Scholarly

and Scientific Communication y presentado en una conferencia celebrada en abril de 1991, señala el mismo punto con respecto a los científicos de manera todavía más precisa. El estudio examinó, en una muestra considerable de profesionales del área científica, su comportamiento al localizar y utilizar información en su trabajo. Se concentró en la información utilizada por profesionales noveles, medios y de más alto rango en las disciplinas de química, genética y ciencia informática en instituciones académicas, privadas y gubernamentales. Basándose en los recuentos diarios de los llamados encuentros de información, que fueron definidos como "cualquier uso o adquisición de información, ideas o datos relacionados con la sustancia de su profesión", el estudio del Faxon Institute contiene muchos datos valiosos y reveladores sobre la "competencia por información" percibida entre los profesionales estudiados: el tipo de fuentes que utilizan para recabar información, la frecuencia con la cual usan esas fuentes, el tiempo que dedican a recabar información, el uso que le dan a ésta y, particularmente, sus métodos o modos de recabar la información que necesitan.⁵

La **Tabla 1** resume algunos de los resultados obtenidos en el estudio del Faxon Institute e indica claramente hasta qué punto los investigadores del área científica dependen, para su trabajo productivo, de fuentes de información que estén fácilmente al alcance de su mano. Los participantes del estudio encontraron que la biblioteca es tan útil como consultar sus bibliotecas personales y entablar discusiones cara a cara. Sin embargo, utilizan las bibliotecas con menos frecuencia que los otros modos, presumiblemente porque consume más tiempo llegar hasta ellas y usarlas. Por otra parte, la consulta de los materiales que se tienen a mano en la biblioteca personal apareció en muchos más "encuentros de información" que cualquier otro método y fue, junto con el correo electrónico, el método usado con mayor frecuencia. Además, si se combinan todos los métodos para consultar a sus colegas -discusiones cara a cara, llamadas telefónicas, comunicaciones

Modos de Información	Porcentaje del total de "encuentros de información" en que el modo fue usado	Número promedio de veces usado la semana pasada	Utilidad percibida del modo (clasificación media en una escala de 10 puntos)
Archivo/Biblioteca personal	45,0%	9,7%	7,9%
Biblioteca	31,0%	2,8%	7,9%
Discusión cara a cara	29,0%	7,7%	7,9%
Teléfono	15,0%	9,5%	7,2%
Comunicación por escrito	7,0%	4,4%	5,9%
Bases de datos en línea	6,0%	2,6%	6,8%
Correo electrónico	5,0%	14,7%	6,2%
Fax	4,0%	2,6%	6,8%
CD-Rom	2,0%	1,6%	5,1%
Boletines computarizados	0,3%	3,6%	4,3%

Tabla 1. Uso de los modos para obtener información⁶

por escrito, correo electrónico y fax-, el estudio del Instituto Faxon sugiere que la red personal de colegas ocupa el primer lugar y la biblioteca está en un distante tercer puesto en el porcentaje de veces que es utilizada para resolver un problema concerniente a la búsqueda de información.

10

La biblioteca accesible

Los resultados de los estudios del Research Libraries Group y del Faxon Institute refuerzan en general los hallazgos hechos en casi dos décadas de investigación sobre el comportamiento de los investigadores en la búsqueda de información.⁷ Todavía hay mucho campo para trabajos adicionales que refinan y pongan a prueba estas conclusiones en forma rigurosa entre disciplinas y grupos disciplinarios. En particular, dentro del contexto de estos estudios se necesita información más detallada sobre las distintas formas en que los investigadores de varias disciplinas usan realmente las bibliotecas. No obstante, si, como estos estudios sugieren, los investigadores tienen en alta estima y tienden a favorecer la información que está fácilmente a mano, entonces un parámetro crítico de éxito para las bibliotecas que proyectan su rumbo hacia el futuro será determinar con qué facilidad pueden poner la información en manos de sus usuarios.

Un modelo de biblioteca orientada al

acceso pudiera estar representado esquemáticamente en la **Ilustración 1**. Este modelo es, por supuesto, una simplificación. No representa adecuadamente todas las características de la biblioteca y, ciertamente, no es la única forma de caracterizar el sistema. Sin embargo, y para los fines de la discusión, resalta los elementos funcionales claves y sus interrelaciones.

En este modelo, las colecciones y los servicios son los dos componentes primarios de la biblioteca. La administración de la biblioteca es el tercer gran elemento. La administración respalda y apunala la colección y las funciones de servicios de acceso, organizando y prestando servicios dentro de la biblioteca, tales como instalaciones, presupuesto, personal y gerencia tecnológica, y representando a la biblioteca ante sus usuarios y otras agencias externas. También es trabajo de la administración formular y articular la misión de la biblioteca en el contexto institucional local.

La misión de la biblioteca orientada al acceso es generar, preservar y mejorar el acceso a las colecciones de conocimiento registrado. Esta misión sirve de directriz para la relación fundamental entre los servicios de acceso y las colecciones de la biblioteca. Los servicios de acceso ofrecen los puntos de contacto entre la biblioteca y sus usuarios, dividiéndose en dos tipos:

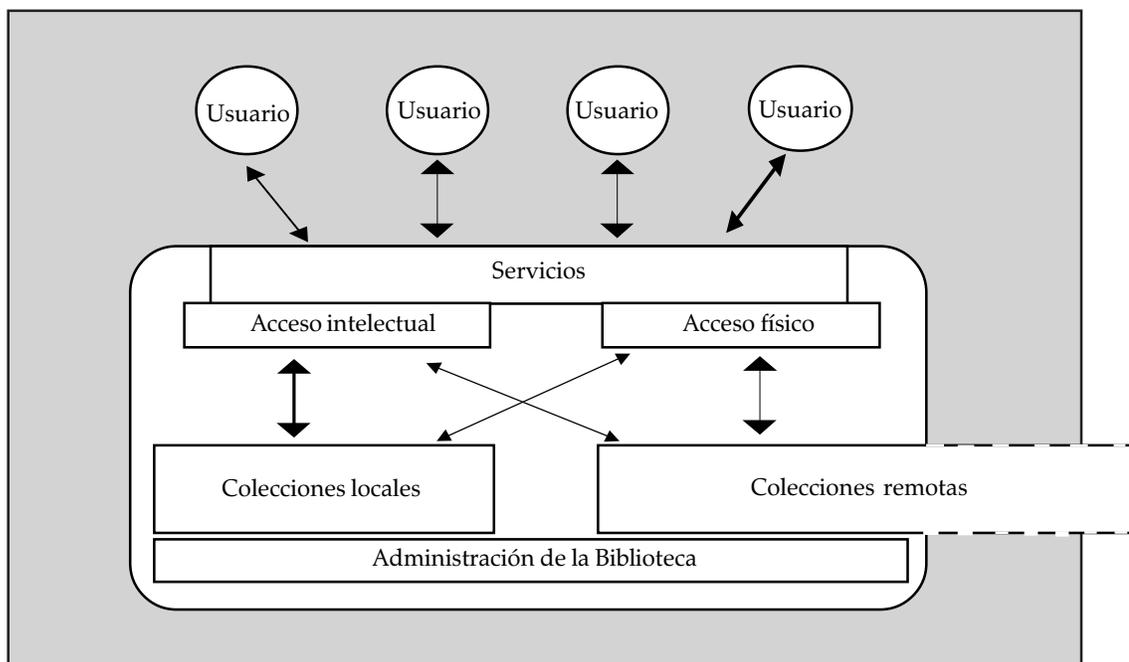


Ilustración 1. Modelo de biblioteca orientada al acceso

- **Acceso intelectual y físico**

Los servicios de acceso intelectual, como los de catalogación, describen y organizan para los usuarios de la biblioteca la información y el conocimiento registrado que se guardan en las colecciones locales y en las de otros sitios. Las descripciones de una obra generalmente refieren, en formas normalizadas a características como autor, título, lugar y fecha de publicación, editorial y formato; generalmente la organización se hace por materias. Dada esta organización por materias, la ubicación de materiales en la estantería pone a disposición de los usuarios de una biblioteca diversos materiales relacionados. Éste es un servicio de acceso intelectual que se aprecia particularmente en los estudios humanísticos e interdisciplinarios, donde los métodos asociativos hacen de los resultados obtenidos por causalidad (al revisar las estanterías de la biblioteca) un aspecto crítico de investigación.⁸ Otras dimensiones del acceso intelectual incluyen la disponibilidad del contenido estructural del material, a menudo representado en la tabla de contenido, y la disponibilidad de índices

por palabras para todo el texto del material. Los servicios de acceso físico, tales como los préstamos circulantes e interbibliotecarios, se refieren a los métodos de distribución que utilizan las bibliotecas para poner la información y el *conocimiento registrado*, existente tanto en colecciones locales como remotas, directamente en las manos de sus usuarios.

- **Colecciones y renovación de colecciones**

El concepto de colecciones de biblioteca se refiere a las disposiciones y modos de almacenar la información seleccionada y el *conocimiento registrado*. Para los usuarios de la biblioteca, las colecciones locales son las más relevantes. Por supuesto, las colecciones de una sola biblioteca de investigación representan únicamente una pequeña fracción del reservorio total de información y *conocimiento registrado*. De allí que la responsabilidad por las colecciones locales incluya necesariamente responsabilidades tanto para distinguirlas de las colecciones ubicadas en otros sitios como para relacionarlas con ellas.

Dentro de la universidad, las colecciones de las bibliotecas representan una gran inversión de capital, quizás la más grande de este tipo de inversiones. Como sucede con todos los capitales, la colección de una biblioteca está sujeta a distintos y constantes procesos de depreciación y renovación (ver **Ilustración 2**). Los materiales se pierden o extravían y deben ser reemplazados. Las nuevas adquisiciones aumentan la colección. Las colecciones son sometidas a un proceso de revisión y selección en el cual se descartan materiales. El material existente se deteriora y debe ser reparado o conservado de otro

al tiempo, la acidez o el uso intensivo limita tanto el acceso físico como el acceso intelectual al mismo. Cuando se selecciona un material deteriorado para preservarlo, comúnmente se prepara para microfilmarlo (ver **Ilustración 3**). Por ser un proceso fotográfico, la microfilmación reproduce fielmente el material impreso original, incluyendo manchas, decoloraciones, tinta desvanecida, notas de lectores, y los desagradables bordes de las páginas superpuestas.

El proceso genera una película de 35 mm, imágenes en blanco y negro, altamente contrastadas y en una relación de reducción

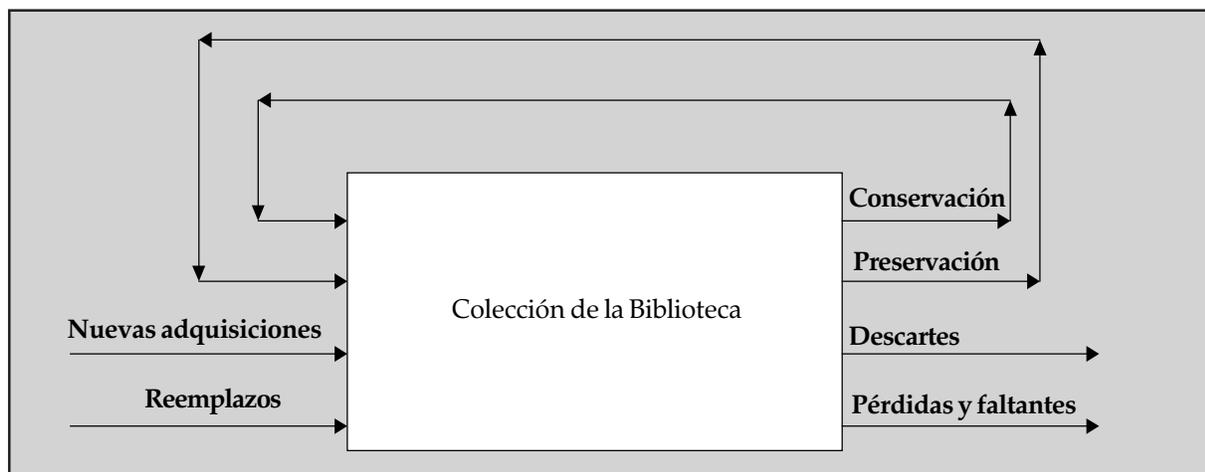


Ilustración 2. Proceso de la renovación de colecciones

modo, o se preserva copiándolo o convirtiéndolo a otro formato, por ejemplo del papel a microfilme.

En una biblioteca orientada al acceso, cada uno de los diferentes procesos de renovación de colecciones está sometido a una rigurosa evaluación frente a la misión de servicio de la organización: ¿Cómo generan, preservan o mejoran el acceso intelectual y físico a la colección? A los fines de esta discusión, nos concentraremos en uno de los procesos de renovación: la preservación en microfilme.

Preservando el acceso mediante la preservación

El deterioro de un documento debido

determinada, dependiendo del tamaño del original. Los límites prácticos de utilizar un medio de alto contraste en blanco y negro significan, por supuesto, que la microfilmación no puede reproducir adecuadamente el material con su contenido de colores, las ilustraciones en escalas de grises o los detalles impresos muy finos. Con las actuales tecnologías para evitar la pérdida de información, el material con ese tipo de contenido debe ser, en la medida de lo posible, conservado en su forma original.⁹

Después de hacer la microfilmación, se guarda una copia maestra (*master*) de la película lejos de la biblioteca, en un ambiente controlado en el que, de acuerdo con algunos estimados, durará hasta más de 500 años sin sufrir un deterioro considerable. Una copia

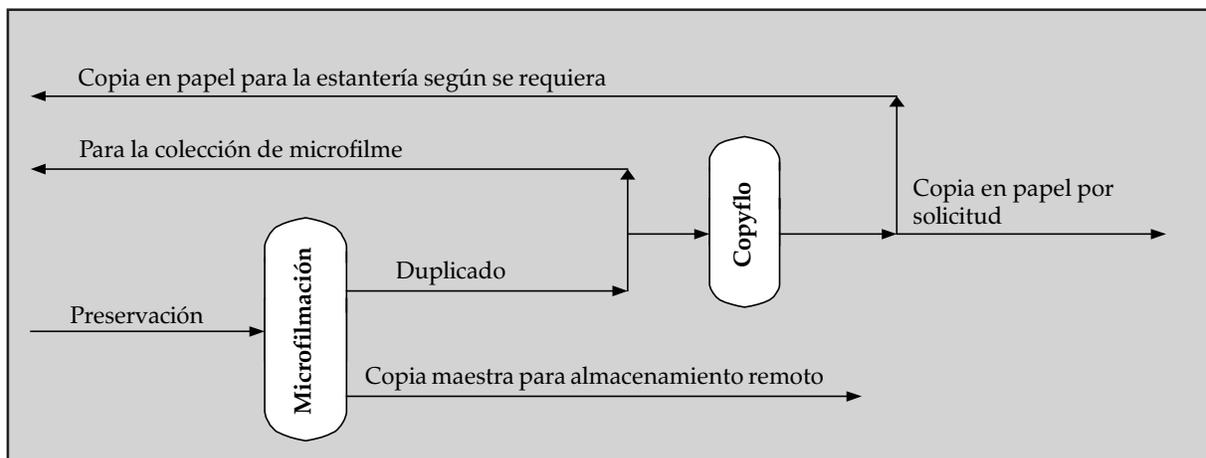


Ilustración 3. Preservación en microfilme

positiva y una negativa del microfilme se almacenan en la colección de microfilmes de la biblioteca. La película puede duplicarse a partir de la copia negativa con relativa facilidad, a un costo aproximado de 20 dólares por rollo, y el proceso *Copyflo* puede generar una copia en papel a un costo de aproximadamente 0,15 dólares por página. Para los materiales de mayor uso, la biblioteca puede decidir si coloca una copia en papel en las estanterías. Los lectores de la biblioteca pueden obtener una copia en papel si lo solicitan, o pueden imprimir ciertas páginas desde el lector/impresor de microfilme.

Ante la pérdida total del contenido intelectual de un documento deteriorado, la biblioteca se da cuenta de que la preservación en microfilme ofrece ventajas únicas. El contenido se salva para la colección en una forma duradera y compacta, usando una tecnología que es relativamente sencilla y está bien establecida. No obstante, en el cambio, los lectores pierden un considerable grado de acceso intelectual y físico al material.

A diferencia de un libro, que uno puede llevar consigo y usarlo prácticamente en cualquier sitio, el microfilme obliga al usuario a utilizar equipo especial de proyección en un lugar específico. Un lector puede saber del contenido intelectual de un documento en microfilme, en parte, a través de la entrada bibliográfica creada para el mismo durante el proceso de preservación, pero, al convertir

el papel al microfilme, el proceso promueve la separación de la colección, de modo tal que ya no es posible beneficiarse intelectualmente de la cercanía física de un volumen en las estanterías con otros de contenido afín. Por otra parte, el microfilme es engorroso de revisar y difícil de leer. Con un lector de microfilme no es posible tener acceso rápido y eficiente al contenido intelectual del volumen, ya sea un hecho, una página, un capítulo o alguna combinación de éstos,¹⁰ lo cual sí lo es en el caso del ejemplar en papel, debido a su estructura interna que incluye tabla de contenido, resúmenes de capítulos, prefacios, notas a pie de página, índices, etc. Finalmente, la copia impresa a partir del microfilme, utilizando ya sea la tecnología de *Copyflo* o del lector/impresor, ha recibido escasos elogios por sus cualidades de claridad y resolución.

No resulta sorprendente que, ante estas deficiencias, los lectores se quejen insistentemente de tener que usar el microfilme. En cierto sentido, las quejas por las deficiencias de esta tecnología parecen injustas: sin la preservación en microfilme, partes considerables de colecciones de bibliotecas se perderían por completo. En este contexto, la inconveniencia de utilizar el microfilme es el bajo precio a pagar. Sin embargo, en vista del valor que le dan los investigadores al fácil acceso a la información, la biblioteca está obligada a garantizar que no puede aplicar otra alternativa tecnológica mejor y más efectiva

(desde el punto de vista de los costos) que la microfilmación para salvar el contenido intelectual de los materiales que se deterioran rápidamente y, al mismo tiempo, preservar o mejorar el acceso de los investigadores al material. Por esta razón, bibliotecas como la de Yale, en las cuales más del 80% de la colección se está actualmente deteriorando o es susceptible de deterioro, deben investigar los medios, costos y beneficios de agregar a su arsenal de preservación nuevas y prometedoras herramientas, incluyendo la tecnología de la digitalización de imágenes.

La digitalización de imágenes

La tecnología de la digitalización de imágenes brinda los medios para codificar digitalmente documentos escaneados en forma de imágenes para su almacenamiento, transmisión y recuperación en sistemas computarizados. Las imágenes digitales producidas utilizando esa tecnología contienen texto, pero el texto no es convertido a caracteres alfanuméricos y tampoco se tiene acceso a él por esa forma. Existe el potencial para este tipo de conversión, ya sea total o parcial, aplicando las tecnologías de reconocimiento de caracteres a las imágenes digitales. El uso de la digitalización de imágenes avizorado aquí está diseñado para anticipar la eventual aplicación del reconocimiento de caracteres, pero por el momento no lo incorpora.¹¹

La digitalización conlleva la promesa de mejorar enormemente el acceso a materiales preservados en comparación con el microfilme (ver **Ilustración 4**). El acceso físico a la biblioteca de imágenes digitales, al igual que el acceso al microfilme, requiere de equipo especial. Sin embargo, en comparación con los lectores de microfilme, el equipo para leer imágenes digitales -un poderoso computador con un monitor de resolución relativamente alta-, constituye una tecnología de uso general donde el acceso a la biblioteca digital sería sólo una de sus aplicaciones. Por otra parte, el investigador que está fuera, y posiblemente lejos de la sede de la biblioteca que guarda las imágenes, puede tener acceso a ellas con relativa facilidad y rapidez, porque

las imágenes digitales pueden transmitirse a través de las redes de alta velocidad que ya muchas universidades han instalado e interconectado. El investigador también tiene la oportunidad, utilizando la tecnología de la imagen digital, de agregar un documento digitalizado, o parte de éste, a su biblioteca personal solicitando una copia impresa. La forma impresa probablemente tenga mejor calidad y mayor resolución que la copia equivalente obtenida a partir de un microfilme, e incluso pudiera ser sustancialmente más barata, porque una vez más la tecnología implícita sirve a un propósito general.

La digitalización de imágenes también promete mejorar sustancialmente el acceso intelectual, en comparación con el microfilme. La capacidad de la tecnología para suministrar copias en papel, de alta calidad y relativamente baratas, a la biblioteca personal de los investigadores cuando es solicitada, hace que el aparato intelectual normalizado -la tabla de contenido, el índice y lo demás-, sea directamente accesible. Además, a diferencia de la actual práctica con imágenes en microfilme de 35 mm, las imágenes digitales pueden clasificarse por número de página y, si los recursos lo permiten, por un orden más estructurado (por ejemplo, sección, capítulo y parte). Si se dispone de un índice de este tipo y del programa para usarlo, los investigadores pueden moverse rápidamente por el documento digital y revisar su contenido intelectual de manera mucho más eficiente de lo que pueden hacerlo con documentos preservados en microfilme.

En las formas esbozadas aquí, la digitalización de imágenes no significa una mejora sustancial frente al microfilme, en lo concerniente a brindar un sustituto efectivo para la revisión de estanterías. Sin embargo, recientes estudios sugieren que agregando información suplementaria, como transcripciones de tablas de contenido, a descripciones bibliográficas en línea, se pudieran brindar oportunidades todavía más provechosas para las asociaciones intelectuales que aquellas que son posibles con los métodos tradicionales de clasificar y colocar el material por materia en las estanterías de la biblioteca.¹²

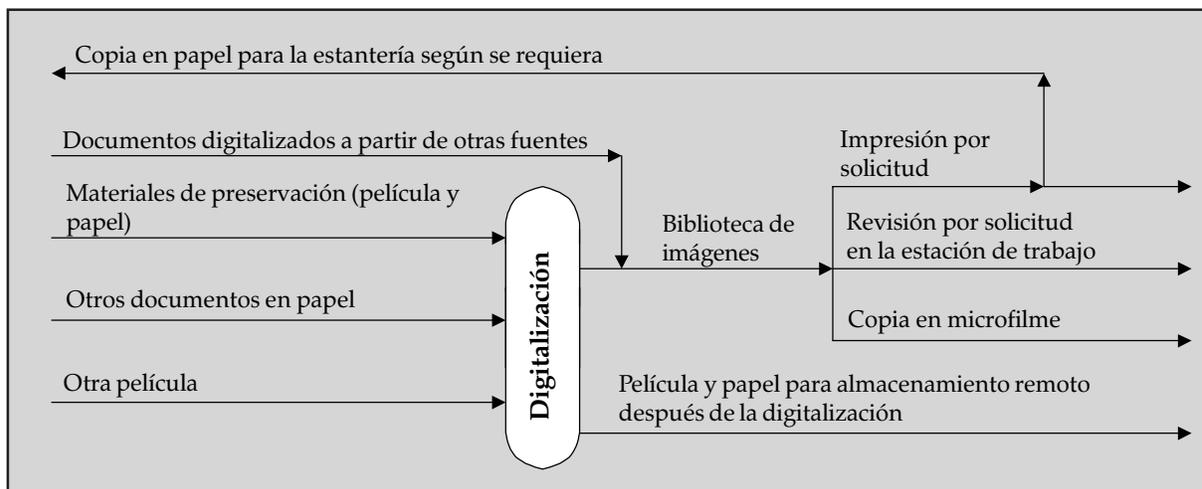


Ilustración 4. Digitalización de imágenes en la biblioteca

A medida que se añade información sobre el contenido de los documentos a los registros bibliográficos en línea, se hacen evidentes los beneficios resultantes relativos al mejoramiento del acceso intelectual a los materiales en todos los formatos, haciéndose posible sacar una considerable ventaja intelectual del material en formato de imágenes digitales, por la creación de lazos electrónicos directos entre la información adicional en el registro bibliográfico y las imágenes digitales del documento. La creación de tales lazos requerirá un considerable trabajo, particularmente en el sistema bibliográfico en línea, pero la perspectiva de su posibilidad también significa que se deben diseñar cuidadosa y creativamente el número de página y los índices estructurales del documento dentro del sistema de imágenes digitales, de modo tal que las imágenes digitales puedan ser relacionadas fácil y confiablemente con el contenido informativo almacenado en otras formas legibles mecánicamente, tales como los registros bibliográficos en línea.

Nuestra visión de la biblioteca del futuro depende de las distinciones e interrelaciones centrales que hemos identificado entre:

- las colecciones de la biblioteca, que comprenden los modos de almacenamiento para la información seleccionada

y el *conocimiento registrado*;

- el acceso físico, que comprende los modos de distribuir la colección, y
- el acceso intelectual, que comprende los principales modos de usar la colección.

Así como esperamos que la tecnología de digitalización de imágenes introduzca mayor flexibilidad y mejore los modos de utilizar y distribuir la colección, particularmente en comparación con el microfilme, también esperamos que la tecnología ejerza efectos similares en las distintas maneras de almacenar la colección. En este contexto, es muy importante distinguir entre almacenamiento de archivo y almacenamiento para uso y distribución.

En la biblioteca del futuro, suponemos que el microfilme seguirá desempeñando un papel relevante, quizás permaneciendo como el medio preferido para guardar versiones con calidad de archivo de los documentos preservados.¹³ Sin embargo, para el uso y la distribución de materiales preservados, las imágenes digitales probablemente se convertirán en el medio de almacenamiento por excelencia. Dependiendo de los costos, el microfilme pudiera seguir siendo un paso preliminar en el proceso de preservación, conduciendo eventualmente a la digitalización, o quizás la película podría ser generada a partir de la imagen digital. En cualquier caso, esperamos que la tecnología

digital acepte como insumo a su demanda la sustancial base de documentos ya preservados en microfilme que ahora sólo se consiguen en ese formato.

Finalmente, nuestra visión del uso de la tecnología para la digitalización de imágenes sugiere varios efectos más amplios. Aunque esperamos establecer la biblioteca digital por el deseo de preservar y mejorar el acceso a las inmensas cantidades de libros frágiles y en deterioro que se encuentran en las estanterías de las bibliotecas, también deseamos que la biblioteca se enriquezca en el transcurso del tiempo con documentos en forma de imágenes digitales que han sido creados por muchas otras razones dentro y fuera de la biblioteca. Además, la tecnología de la digitalización de imágenes probablemente tendrá un efecto significativo, quizás profundo, en la forma en que la biblioteca del futuro ve, organiza y valora sus espacios y su personal. Por ejemplo, en la medida en que la biblioteca digital surja como una forma de renovar y sustituir grandes cantidades de materiales en papel y microfilme, y en la medida en que la biblioteca digital sea almacenada y accesible desde lugares remotos a la propia sede, el uso del espacio central de estantería de la biblioteca obviamente cambiará y el balance de las funciones del personal tradicionalmente asociadas con las colecciones circulantes y con las almacenadas en papel en las estanterías seguramente se transformará en función de los diversos modos de distribución de los documentos que surgirán como consecuencia de las solicitudes procedentes de formatos electrónicos.¹⁴

ALTERNATIVAS ECONÓMICAS

Las bibliotecas del futuro no alcanzarán, en un solo y rápido paso, la visión articulada aquí sobre el uso de la tecnología digital para preservar y mejorar el acceso al *conocimiento registrado*. La promesa de beneficios en el servicio y de posibles reducciones en los costos hará que las bibliotecas se encaminen hacia esta tecnología en diferentes formas. Los presupuestos y otros impedimentos organizativos no les permitirán invertir en ella de muchas otras maneras. Si la promesa justifica la inversión, la mayoría incorporará la tecnología para la digitalización de imágenes de la misma forma en que adoptaron y asimilaron otras tecnologías en el pasado: a través de una serie creciente de decisiones y selecciones particulares, adaptadas al mandato y a las necesidades de sus instituciones específicas.

Quienes desarrollan la tecnología de digitalización de imágenes para ser utilizada en la preservación de materiales de biblioteca deben estar conscientes del principio de perfeccionamiento y de su poder e influencia sobre los cambios organizativos. La economía para gerenciar y administrar los recursos de una biblioteca, como en la mayoría de las grandes instituciones, es una economía de opciones progresivas, incluso las concernientes a tecnologías potencialmente revolucionarias. Por consiguiente, el trabajo de desarrollo de aplicaciones para la digitalización de imágenes necesita generar resultados en lo relativo a los medios, costos y beneficios de la digitalización de imágenes, que faciliten e informen, sin distorsionar ni confundir, la operación de tal economía.

En esta sección se intenta identificar un conjunto de opciones funcionales críticas, que posiblemente regirán la incorporación de la digitalización de imágenes en el proceso de preservación de las bibliotecas. En la próxima sección se esbozan muchos de los detallados requerimientos técnicos para la arquitectura de un sistema que pueda acomodar flexiblemente las distintas opciones que las bibliotecas probablemente requerirán. En la sección final se presenta un plan de trabajo

para desarrollar la arquitectura y generar la información que las bibliotecas necesitarán para escoger las opciones adecuadas en lo referente a los costos y beneficios de la digitalización de imágenes.

La conversión desde el microfilme como objetivo

La Cornell University, pionera en el uso de la digitalización de imágenes para la preservación en bibliotecas, en vez de microfilmar documentos optó por investigar las maneras de superar las limitaciones de la preservación a través del microfilme, mediante el escaneo directo de documentos deteriorados para obtener imágenes de los mismos. Todavía es necesario trabajar más para expandir los medios de alimentación directa, incluyendo los métodos para manejar documentos de grandes dimensiones, documentos a color y otros tipos de documentos especiales. Además, como pueden ser mejoradas electrónicamente para eliminar manchas y otras marcas y para realzar el contraste de la tinta desvanecida, las imágenes digitales, a fin de cuentas, deben resultar una fuente de mejor calidad para la copia en microfilme que la fuente original en papel. Los medios para generar microfilmes a partir de imágenes digitales también necesitan ser explorados sistemáticamente.

El papel, sin embargo, no es la única fuente posible de insumo para la creación de

una biblioteca digital de materiales preservados. La Yale University Library intenta demostrar a la amplia comunidad bibliotecaria la viabilidad de una fuente alterna para desarrollar los medios e identificar los costos y beneficios de escanear materiales preservados a partir de fuentes microfilmadas. Aunque sea alimentado con microfilmes en vez de papel, el sistema de procesamiento de imágenes desarrollado en el proyecto de Yale también proveerá las mismas capacidades funcionales básicas que el que está siendo desarrollado en Cornell. En el sistema totalmente desarrollado, los usuarios podrán hojear la biblioteca digital desde un terminal de computador (ver **Ilustración 5**). Igualmente estarán en capacidad de obtener una versión impresa del documento digital si así lo solicitan; la biblioteca pudiera incluso generar una copia en papel del documento preservado para devolverla a las estanterías. Por otra parte, la biblioteca de imágenes se almacenará en un lugar alejado de la colección ubicada en estanterías. Debido a la supuesta accesibilidad de la biblioteca digital, tanto fuera como dentro de la biblioteca, la versión en microfilme de los documentos digitales pudiera transportarse fuera de la biblioteca a un depósito más apartado y barato.

El sistema de Cornell está siendo desarrollado, inicialmente, para manejar una biblioteca de imágenes de 1.000 volúmenes. Además de desarrollar una fuente alterna de alimentación, el proyecto de Yale probará la

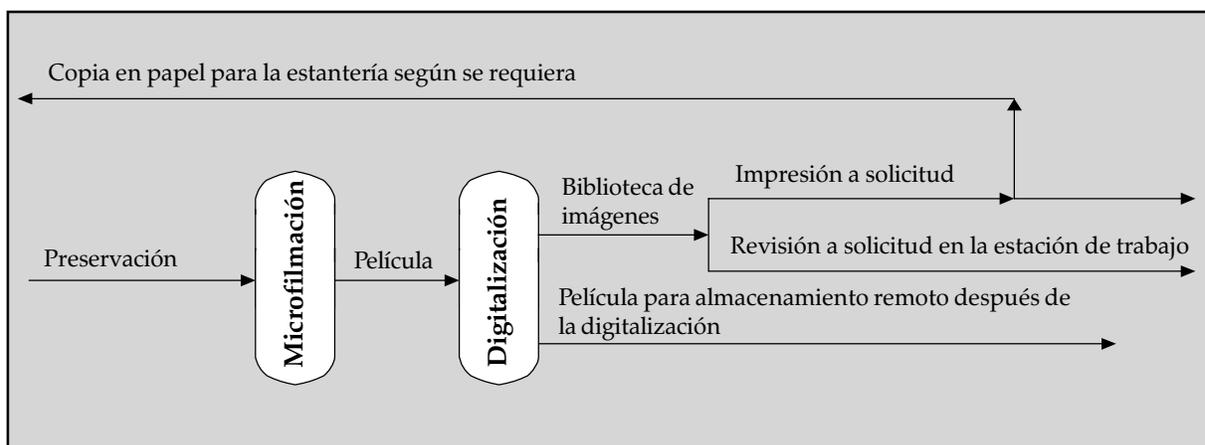


Ilustración 5. Conversión de microfilme de preservación a imágenes digitales

posibilidad de estructurar por escalas los sistemas de digitalización de imágenes, como los de Cornell, aumentando la biblioteca digital a 10.000 volúmenes. La dimensión de la biblioteca digital de Yale, comparada con la de Cornell, servirá para demostrar las economías de escala en el sistema y garantizará a la comunidad bibliotecaria que se puedan tomar medidas realistas, tanto en lo que se refiere a los costos de agregar progresivamente componentes al sistema, como en lo concerniente al mantenimiento de éstos con el paso del tiempo.

Un modelo de inversión progresiva

Los costos y beneficios de un sistema de digitalización de imágenes dependen en gran medida, aunque no exclusivamente, de los principales componentes funcionales del sistema, así como del método y el momento de su implementación. No es necesario adoptar todos los componentes posibles para crear un sistema de trabajo que funcione. Como está previsto que la misión del sistema sea la de prestar un servicio, algunos componentes son esenciales y deben ser implementados; otros son menos críticos, por lo que su instalación es opcional o puede hacerse posteriormente, según un esquema progresivo. En el siguiente análisis se distinguen los principales componentes funcionales de un sistema de preservación por digitalización, se ordenan los componentes según los propósitos del sistema y se organizan en una secuencia acumulativa de etapas. La secuencia sugiere una estrategia plausible, que ofrece opciones en cada etapa para una biblioteca que está invirtiendo en la tecnología de digitalización de imágenes. Para respaldar y dirigir las opciones, el desarrollo de un proyecto de digitalización de imágenes en Yale debe estar diseñado con miras a medir los costos y beneficios del incremento progresivo de componentes del sistema, identificados en cada etapa de la secuencia.

- **El sistema básico**

Si el objetivo de la preservación es

resguardar y mejorar el acceso físico e intelectual a materiales impresos deteriorados, y suponiendo que dichos materiales ya hayan sido microfilmados, entonces el sistema de digitalización de imágenes más simple y directo es aquel que, a partir del microfilme, produzca una copia impresa del documento original con tan alta calidad que permita a la biblioteca volver a colocarlo en sus estanterías. Los componentes esenciales de un sistema de este tipo serían:

- un escáner para microfilme,
- un dispositivo para almacenar temporalmente las imágenes digitales,
- una estación de control de calidad para revisar las imágenes, mejorarlas si es posible y volver a escanear el microfilme, en caso de ser necesario,
- una red de alta velocidad para transferir las imágenes al servicio de impresión,
- un servicio de impresión, organizado por turnos de solicitudes o cualquier otro método de control, y
- una impresora digital de alta calidad.

Por supuesto, una compañía de servicios podría cumplir con todas estas funciones o parte de ellas. Podría escanear bajo procedimientos de control de calidad e imprimir el documento; podría escanear bajo procedimientos de control de calidad y remitir el conjunto de imágenes digitales a la biblioteca para su impresión; o podría aceptar las imágenes digitales que le envía la biblioteca para su impresión. Sería importante, en el proyecto de demostración de Yale, acumular suficiente experiencia interna en el proceso de digitalización como para que se pueda suministrar a la compañía de servicios los patrones y las expectativas de conversión e indización, y para que se pueda comparar cabalmente los costos de un proceso interno con los de contratar una oficina de servicios para realizar las mismas funciones.

Como mediante el proceso de *Copyflo* ya es posible reproducir una copia impresa directamente a partir del microfilme, es probable que una biblioteca encuentre atractivo este sistema básico para la digitalización de

imágenes si y sólo si los costos de la digitalización e impresión son menores que los costos del *Copyflo*, o si el producto resultante de la impresora digital es de mayor calidad. Suponiendo, lo cual es probable, que la resolución de la imagen digital generada a partir del microfilme esté en el rango de los 300 puntos por pulgada (*dpi*), entonces la calidad de la impresión seguramente será igual o superior a la calidad del *Copyflo*. Por otra parte, Michael Lesk ha estimado que el costo de una conversión simple de microfilme a imagen digital es de aproximadamente 0,02 dólares por cuadro o fotograma.¹⁵ Estimados iniciales del proyecto de Cornell sugieren que los costos de impresión de un documento a partir del formato digital estarán en el orden de 10 dólares por un libro de 300 páginas (sueltas) o 0,033 dólares por página.

Estos estimados incluyen hipotéticos gastos laborales, de producción, de volumen y métodos de financiamiento y amortización de equipos que pudieran ser válidos o no en todos los casos. Sin embargo, reconociendo las limitaciones de los estimados, sigue siendo necesario compararlos con el costo actual de 0,15 dólares por página (suelta) del proceso de *Copyflo*. Un sistema de digitalización de imágenes limitado a sus componentes esenciales y capaz simplemente de digitalizar e imprimir imágenes del microfilme, no sólo produce un resultado de mayor calidad, sino que también reporta ahorros superiores a 60% frente a los actuales medios de impresión a partir del microfilme. Por supuesto, con componentes adicionales, un sistema de digitalización de imágenes puede cumplir con muchas más funciones que el sistema básico descrito aquí, pero los montos de la inversión ascienden de manera proporcional.

● **Impresión por solicitud**

Si un sistema básico de digitalización de imágenes brinda a la biblioteca la posibilidad de restringir el acceso a un documento deteriorado al generar una copia facsímil de alta calidad y a un costo relativamente bajo que pueda volver a colocar en sus estanterías, entonces uno tiene todo el derecho a

preguntarse a cuánto ascendería la inversión adicional para producir una copia impresa por solicitud de un investigador que desea incorporarla a su biblioteca personal. Un sistema de digitalización de imágenes capaz de ofrecer la impresión por solicitud necesitaría la incorporación de los siguientes componentes al sistema básico:

- un mecanismo para almacenar de manera permanente las imágenes de los documentos en formato digital,
- un índice por títulos para el documento en forma de imágenes, de modo que se pueda distinguir un documento de otro en el medio de almacenamiento,
- una entrada en una ficha bibliográfica nueva o ya existente, quizás registrada en el catálogo en línea de la biblioteca, que le indique al lector que el documento está almacenado en forma digital y puede imprimirse por solicitud,
- un servicio de solicitud de impresiones, y
- un servicio de procesamiento de imágenes que extraiga las imágenes según su identificación particular y las transmita a la impresora.

Debido a los mayores beneficios que reporta la impresión por solicitud, también se incurre en costos sustanciales, particularmente en lo que se refiere al sistema de almacenamiento. Almacenar imágenes digitales es un proceso técnicamente complejo y lo discutiremos más detalladamente en la siguiente sección sobre la arquitectura del sistema. Sin embargo, es importante hacer notar aquí que la noción de "permanencia" en el almacenamiento digital implica los costos de renovar periódicamente los bancos de imágenes -lo cual se hace relativamente con más frecuencia de lo que ocurre con los libros y el microfilme-, ya que los medios de almacenamiento, los formatos de estos bancos y el equipo utilizado para tener acceso a dichos bancos cambian por completo. Es posible, pero todavía no se ha demostrado, que los ahorros en términos de espacio, al contar con un método de almacenamiento más denso y compacto

y otras ventajas obtenidas del cambio de tecnología, podrían compensar los costos de renovar los bancos de imágenes y lograr así que este proceso de renovación se mantenga con sus propios recursos.

El proceso de indización en esta etapa es fácil de crear y utilizar. Consiste en aplicar un número de clasificación u otra identificación particular para la colección de imágenes que comprende el documento digitalizado. Esta identificación particular brinda la clave que conecta la entrada bibliográfica en el catálogo con el título en la biblioteca digital, de manera muy parecida a la forma en que una cota le permite al lector ir del catálogo al libro en una estantería, o a un rollo de microfilme en una gaveta específica del archivador.

Por supuesto, un investigador no siempre puede descifrar por la entrada del catálogo si un documento es relevante para el problema o la investigación a la cual está dedicado. Antes de solicitar una copia personal impresa, siempre tendrá la necesidad de revisar en la biblioteca el documento en microfilme para determinar su importancia. Por consiguiente, invertir en un sistema de digitalización de imágenes que brinde el servicio de impresión por solicitud nos hace incurrir en costos de almacenaje e indización de los documentos en formato digital, a la vez que también requiere que la biblioteca mantenga la versión en microfilme del documento para que el lector pueda revisarla.

- **Revisión en línea**

Dado el almacenamiento permanente necesario para un servicio de impresión por solicitud, otra extensión natural de un sistema de digitalización de imágenes sería permitir al lector revisar en línea un documento directamente desde el formato de imagen digitalizada. La aplicación de revisión en línea requiere, además de los ya identificados, los siguientes componentes:

- un terminal de computador y una impresora en los cuales el lector pueda recuperar, traer a la pantalla, retroceder,

adelantar e imprimir imágenes determinadas de un documento, y

- un servicio de presentación que adapte la presentación de la imagen al tipo particular de estación de trabajo para revisiones que se está usando.

La inversión progresiva en estos componentes del sistema de digitalización de imágenes permite al lector aplicaciones funcionales que, por lo menos en un aspecto, se asemejan a la revisión del documento en el formato de microfilme. Como en esta etapa el documento digitalizado sólo se indiza a nivel de título con una identificación particular, uno no puede ir automáticamente a una página o sección determinada del documento. Como sucede con el lector de microfilme, se debe, en efecto, retroceder y avanzar a lo largo del documento, imagen por imagen. La principal diferencia funcional con la utilización del microfilme en esta etapa radica en que el lector no tiene que encontrar el rollo de microfilme y cargarlo manualmente, sino que más bien depende del programa del sistema para encontrar el documento y cargarlo en forma automática. Además, después de la revisión, el lector puede solicitar que el sistema genere una copia impresa de alta calidad para su uso personal.

Es tarea de quienes desarrollan el sistema garantizar que las aplicaciones de revisión e impresión por solicitud incorporadas al sistema de digitalización de imágenes en este momento de la inversión sean atractivas y suficientes para que los investigadores las utilicen en lugar del microfilme. Suponiendo que los haya, la biblioteca puede comenzar al menos a acumular al menos ahorros modestos en espacio y costos de servicio, eliminando la copia *in situ* del documento microfilmado. La biblioteca también puede contemplar la posibilidad de llevar a cabo considerables extensiones de sus aplicaciones de revisión.

- **Acceso remoto**

La biblioteca podría extender las aplicaciones de revisión contenidas en el sistema

de digitalización de imágenes, facilitando su acceso remoto. Dentro de la biblioteca, la presentación de las funciones de revisión puede ser controlada limitando el tipo de estación de trabajo utilizado. Sin embargo, el acceso remoto requeriría un servidor de presentación mejorada para respaldar las diversas estaciones de trabajo con aplicaciones para el procesamiento de las imágenes que los investigadores utilizan fuera de la biblioteca. El servidor de imágenes y la red local del campus también deberían estar en capacidad de soportar este incremento en el tráfico. La posibilidad de que la función de revisión pueda extenderse realmente más allá de la red del campus es un tema que suscita un considerable interés y que requerirá investigaciones adicionales y numerosas pruebas.

- **Revisión a nivel de página**

La biblioteca también puede ampliar la aplicación de revisión en el sistema de digitalización de imágenes, permitiendo que se pueda acceder directamente a las imágenes de un documento por el número de la página. La inversión progresiva en la revisión a nivel de página depende directamente de los costos de generar, mantener y respaldar en el programa de revisión un índice a nivel de páginas del documento digital. No se puede ingresar un comando para ir a una página específica del documento a menos que alguien deliberadamente dedique un tiempo para crear un índice que relacione la imagen con un número de página del documento. Las páginas pueden estar ordenadas en números romanos o arábigos. Por otra parte, algunas páginas quizás ni siquiera estén asociadas con un número específico en ninguna forma de numeración. No obstante, el sistema de indización y recuperación debe reflejar estas distinciones, variaciones y peculiaridades.

- **Revisión a nivel de la estructura del documento**

La biblioteca puede extender todavía más la aplicación de revisión en el sistema de

digitalización de imágenes, permitiendo que se pueda acceder a las imágenes de un documento directamente desde sus propias divisiones estructurales internas, tales como portada, tabla de contenido, parte, capítulo, sección e índice. Es decir, que sin ir pasando página por página del documento, el lector puede ingresar un comando para ir inmediata y automáticamente al tercer capítulo, a la tercera parte o a la portada. Una vez más, la opción progresiva de implementar un sistema de digitalización de imágenes con este nivel de funcionalidad depende directamente del interés de la biblioteca por crear un índice que relacione una imagen o un conjunto de imágenes con un elemento o elementos estructurales particulares del documento. Una ligera variación del tema ofrecería al lector la posibilidad de estructurar temporalmente el documento durante el proceso de revisión con indicadores o marcapáginas. Una variación más ambiciosa le permitiría al lector estructurar el documento por completo o parcialmente según sus propios términos, y guardar el índice estructural resultante bajo su nombre para usarlo posteriormente.

El análisis presentado aquí sobre un sistema de digitalización de imágenes para una biblioteca, orientado a preservar y mejorar el acceso a materiales deteriorados derivado de la conversión de documentos microfilmados al formato digital, no agota todos los usos posibles de un sistema de este tipo. Por ejemplo, se podría preferir digitalizar toda una colección de materiales partiendo de la base de que la coherencia intelectual del material justifica el esfuerzo por hacerlo más accesible. Por otra parte, para ayudar a mitigar los costos de almacenamiento e indización, se pudiera optar por digitalizar sólo los materiales más utilizados, como un paso intermedio, o incluso, como una alternativa, colocar en las estanterías una copia en papel. O también, se podría decidir digitalizar los materiales a medida que son solicitados para su uso, dejando las partes no utilizadas de la colección en microformas.

Este análisis tampoco agota todas las posibles características de un sistema de

preservación mediante la digitalización de imágenes. Por ejemplo, cuando la tecnología para el reconocimiento de caracteres esté suficientemente madura, se harán posibles mejoras adicionales al sistema de digitalización de imágenes. La totalidad o parte de un documento en formato de imagen digitalizada pudiera ser convertida en un archivo de texto de caracteres alfanuméricos, y las imágenes del documento posiblemente podrían relacionarse con una parte del archivo de texto, tal como la tabla de contenido, que esté almacenada en la descripción bibliográfica del documento digitalizado o con el índice de palabras claves generado a partir de todo el archivo de texto. Avances de este tipo podrían enriquecer enormemente el valor intelectual de la biblioteca de imágenes.

Sin embargo, el análisis presentado en esta sección sí esboza las principales dimensiones funcionales de un sistema de preservación mediante la digitalización de imágenes. También sugiere un esquema plausible de opciones progresivas que pudieran motivar a una biblioteca a hacer buenas inversiones en un sistema de este tipo. Por supuesto, se debe distinguir la motivación y la necesidad funcional tanto de la arquitectura técnica intrínseca del sistema de digitalización de imágenes, que sirve a los propósitos de la biblioteca y satisface sus requerimientos, como del plan de trabajo para un proyecto de demostración. Dicho plan está diseñado para desarrollar un sistema completamente articulado dentro de la arquitectura, y para generar suficiente información acerca del sistema con el propósito de que las bibliotecas puedan tomar decisiones prudentes y prácticas en lo que se refiere a costos y beneficios.

ARQUITECTURA DEL SISTEMA

La arquitectura -el diseño y la estructura generales-, de un sistema para crear, almacenar, recuperar e imprimir documentos bibliográficos en formato digital ha evolucionado gradualmente durante la última década. Las primeras investigaciones y el desarrollo de proyectos piloto, particularmente en la Library of Congress y la National Medicine Library, han probado, refinado y validado ideas claves de diseño a medida que ha ido surgiendo la tecnología para el procesamiento de imágenes: las estaciones de trabajo realizan las funciones de digitalizar imágenes; se distribuyen en una red de alta velocidad, utilizan recursos que se necesitan en la red e interactúan entre sí bajo la modalidad cliente/servidor.¹⁶ En su proyecto de preservación digital, Cornell ha incorporado un diseño con estaciones de trabajo distribuidas, bajo la modalidad cliente/servidor, en su arquitectura de sistema, y Yale también lo hará.

Otros principios de diseño más generales también se aplican a la arquitectura de los sistemas de digitalización de imágenes. Varios de estos principios regirán el sistema que se está desarrollando en Yale, y es necesario que se articulen, no que se den por sentado. En primer lugar, tanto las imágenes como los índices que la integran conformarán el principal activo generado con el proyecto de Yale. Los programas y los equipos de computación necesarios para crear, almacenar y utilizar la información cambiarán y serán reemplazados. Sin embargo, la información permanecerá y debe seguir siendo rápida y fácilmente convertible de un sistema a otro. En segundo lugar, basándonos en el argumento presentado en la sección anterior, referente a la necesidad de que haya opciones, los componentes del sistema deben desarrollarse en forma de módulos, de modo que puedan ser rápida y fácilmente agregados, eliminados o modificados según sea necesario. En tercer lugar, para garantizar la flexibilidad cuando se agreguen, eliminen o modifiquen componentes del sistema de digitalización de imágenes en Yale, dichos

componentes deben cumplir las normas oficiales o *de facto* de la industria, o bien ser elaborados según interfaces normalizadas.

Invocando estos principios en orden inverso, esta sección identifica los patrones pertinentes al sistema de digitalización de imágenes en Yale, revisa los componentes del sistema y los aspectos técnicos relacionados con ellos, y analiza la factibilidad general del diseño, resaltando aquellas áreas donde la preocupación por la información pudiera requerir una especial atención a determinadas características del sistema.

Normas

Las normas pertinentes que rigen para el diseño y la construcción del sistema de digitalización de imágenes en Yale se dividen en tres grupos: normas de comunicación, de almacenamiento y de aplicación.¹⁷ Tal y como ha hecho la mayoría de las principales universidades de investigación en Estados Unidos, la Yale University acepta diversos protocolos de conexión en red, pero se ha concentrado en las llamadas normas *ethernet*, que incluyen TCP/IP, como la suite de protocolos por excelencia para las comunicaciones en red. Los componentes del sistema pueden operar en una subestación de la red del campus universitario, utilizando un patrón diferente de protocolo para la serie, por ejemplo un *token ring* en el nivel de enlace de datos. Sin embargo, la subestación debe ofrecer una puerta de entrada confiable y operativa para que la columna vertebral, que es esa red basada en *ethernet*, pueda comunicarse con otros componentes del sistema.

Para el almacenamiento de los datos de imágenes digitales, el sistema de Yale utilizará el *Tagged Image File Format (TIFF)*, que brinda la normalización a nivel de encabezados necesaria para permitir el intercambio de imágenes. Como los archivos de imágenes son tan extensos, también se necesita una norma para comprimirlos a una fracción de su tamaño original, a fin de almacenarlos y transmitirlos económicamente. El patrón de compresión CCITT Grupo 4 presenta algunas deficiencias porque no cubre páginas que

sean mayores al tamaño carta y tampoco maneja bien las imágenes con amplias gradaciones de grises o de color. Actualmente, los grupos de normalización están cooperando para desarrollar normas de compresión más apropiadas. Mientras tanto, la norma CCITT Grupo 4 debería satisfacer las necesidades de este proyecto.

En el nivel de las aplicaciones, la información bibliográfica referente a un documento en forma de imágenes digitales se incluirá en el catálogo en línea de la biblioteca bajo el formato normalizado MARC (catalogación para lectura automatizada). Las convenciones MARC existentes en este momento no son completamente adecuadas para describir la versión en imágenes digitales de un documento, pero la organización responsable de la normalización ha preparado una revisión aceptable de esta norma. Yale adoptará una solución temporal hasta que la norma en revisión sea aceptada.

Otras dos normas también son apropiadas a nivel de aplicaciones. Primero, los índices que abren un documento en forma de imágenes para su revisión a nivel de título, a nivel de página y a nivel de estructura del documento dependen del trabajo que describe el contenido de las imágenes, y que define y hace posible intercambiar y presentar un conjunto determinado de imágenes como un grupo de entidades de orden superior (por ejemplo, un libro, un capítulo, un grupo de páginas). Las normas que cubren el contenido del documento deben aplicarse (o ser debidamente ampliadas de modo de poder aplicarlas) a la creación, almacenamiento y uso de los diversos niveles de índices estructurales. Varios proveedores, entre ellos IBM, han promulgado sus propias normas internas para el contenido de documentos. Sin embargo, la norma ISO 8613, *Information Processing -Text and Office Systems -Office Document Architecture (ODA) and Interchange Format (ODIF)*, es la principal norma internacional, y el proyecto de Yale tratará de cumplir con sus requerimientos.

Segundo, los archivos de estructura, que describen el contenido de los documentos con imágenes digitales conforme a la norma ISO

8613, deben ser almacenados y recuperados de un archivo con base de datos conectado referencialmente a los archivos de imágenes digitales de los documentos. Esperaríamos que la base de datos fuese relacional. También esperaríamos que incorporase y cumpliera con la norma SQL (*Structured Query Language*).¹⁸

Componentes del sistema

En la **Ilustración 6** aparece una representación general de la arquitectura del sistema para el proyecto de digitalización de imágenes en Yale. Todos los componentes están conectados con la red *ethernet* de alta velocidad del campus, la cual tiene un núcleo de fibra óptica y velocidades de transmisión de hasta 10 *megabytes* por segundo. Los componentes claves del diseño son un subsistema de conversión, un subsistema de reproducción (o impresión), un subsistema de almacenamiento y un subsistema de acceso.

imágenes en fotogramas de microfilme que constituyan un documento (libro, folleto o publicación periódica), y tomar una "fotografía digital" de cada cuadro, de modo tal que la imagen quede almacenada en un archivo de computador y pueda verse en un monitor de computador. El proceso de conversión garantiza la calidad física de la digitalización, lo cual incluye el control de la resolución, así como el contenido de las imágenes, contraste, posición y secuencia. El proceso también sirve para caracterizar, o indizar, el conjunto de archivos de imágenes, a fin de que conserven su identidad como un solo documento para su posterior almacenamiento y recuperación. El paso final del proceso consiste en consignar el conjunto de archivos como una entidad documental para su almacenamiento permanente.

El subsistema de conversión incluye una digitalizadora de microfilme y papel y, como mínimo, una estación de trabajo de control

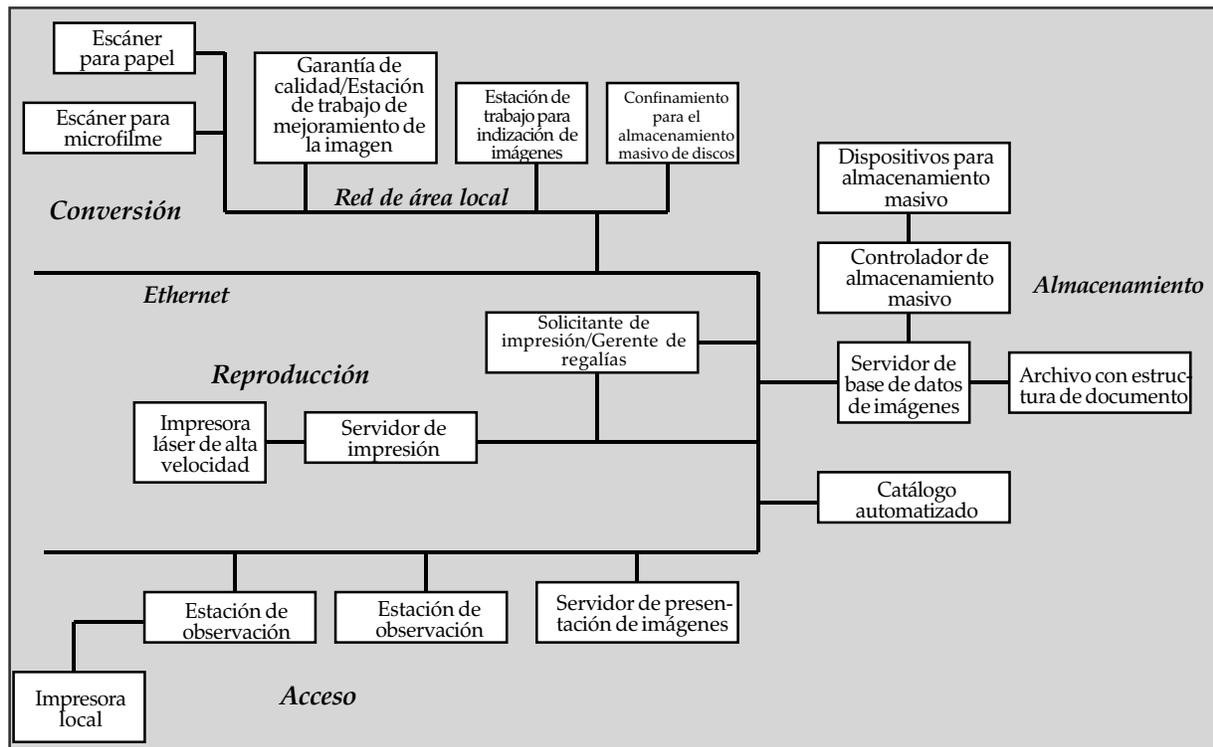


Ilustración 6. Arquitectura de los sistemas de digitalización de imágenes

- **Conversión**

El propósito del subsistema de conversión es aceptar como insumos una serie de

en la cual un operador pueda realizar las funciones de control de calidad, indización y consignación. Para brindar máxima flexibilidad, la estación de trabajo y la digitalizadora

deberían estar conectadas a través de una red local de alta velocidad. Con la interconexión de redes, se pueden agregar estaciones de trabajo adicionales para dividir las tareas, pero cada función deberá ser realizada en cada una de las estaciones. Alternativamente -y esto pudiera ser necesario para alcanzar la máxima eficiencia-, la digitalizadora debe estar conectada directamente a la estación de trabajo de control, en la que el operador realiza las funciones de control de calidad. Al finalizar su tarea, el operador puede enviar los archivos, utilizando una red con mecanismo para compartir archivos (o *sneakernet*), a otra estación de trabajo a fin de continuar con el proceso de conversión.

La digitalizadora debe aceptar rollos de película negativa o positiva de 35 mm, con fotogramas dentados o no, y en variadas proporciones de reducción. Debería estar provista de un mecanismo automático de alimentación que permita su funcionamiento sin supervisión. Debería además estar equipada con sensores para detectar el borde de comienzo y de salida de cada fotograma y corregir automáticamente la posición de encuadre. De ser posible, debería detectar el borde de comienzo y de salida de cada página en los fotogramas microfilmados con dos páginas por cuadro. Dado el esfuerzo invertido en la digitalización, el proyecto de Yale buscará un resultado final con la más alta resolución posible desde el punto de vista técnico. Sin embargo, la digitalizadora debería ser capaz de escanear a diferentes resoluciones y, para evitar cuellos de botella en los trabajos, de generar su más alta resolución de producción a razón de un fotograma por cada 2 segundos o menos. Finalmente, la digitalizadora debería presentar un conjunto abierto de interfaces que le permitan trabajar con estaciones de trabajo DOS, UNIX o Macintosh. Debería estar al alcance una digitalizadora de alta velocidad con un conjunto similar de características para permitir comparaciones controladas de velocidad y calidad en aquellos casos en que la copia en papel del material microfilmado siga estando disponible.

Para trabajar con las imágenes digitales

en el proceso de conversión, cada estación de trabajo requiere la capacidad de procesamiento para leer y expandir una imagen en papel comprimida, así como para comprimir y guardar en disco una imagen en papel exhibida en el monitor de la estación de trabajo. Todas las estaciones de trabajo deben tener la posibilidad de retroceder y adelantar a lo largo de la secuencia de archivos de imágenes y saltar a una imagen específica relativa a la imagen actual. Cada estación de trabajo también requiere un monitor de alta resolución capaz de mostrar una imagen en papel legible. Así mismo debe brindar funciones generales de manipulación de imágenes, como acercamiento, reducción, toma panorámica, embobinado y rotación. Finalmente, todas las estaciones de trabajo en el proceso de conversión requieren acceso a una impresora láser local para imprimir copias de determinadas páginas.

La estación de trabajo que controla la digitalizadora y la calidad de su resultado final debería permitir al operador tanto iniciar un proceso de escaneado completamente automático, como intervenir en ese proceso limpia y vigorosamente. Debería permitirle también ajustar y determinar la resolución de escaneado, produciendo quizás un conjunto de imágenes a alta resolución para impresión, y otro grupo a una resolución menor para revisión. La estación de trabajo debería utilizar técnicas de mejoramiento automático de imágenes, incluyendo detección fija y dinámica de umbrales y picos, las cuales ayudarían a mejorar el débil contraste de la impresión con respecto al fondo. Además, debería permitir al operador recortar los bordes de las imágenes y realizar otras funciones de mejoramiento de las mismas en tiempo real, así como también volver a escanear completamente la imagen si fuese necesario. Finalmente, la estación de trabajo debería crear automáticamente archivos desde el digitalizador bajo TIFF, comprimir las imágenes escaneadas en la técnica bidimensional del CCITT Grupo 4, y asignar nombres secuenciales a los archivos de imágenes.

La aplicación de indización en el proceso de conversión debería guiar al operador a

través del proceso de suministrar información sobre ciertos aspectos claves de la estructura y el contenido del conjunto de imágenes que conforma el documento convertido. En una ventana, debería mostrar el conjunto de imágenes y permitir al operador retroceder y adelantar a lo largo de la secuencia de archivos de imágenes e ir directamente a una imagen específica. En otra ventana, la aplicación de indización debería llevar al operador a verificar que la secuencia de las imágenes es correcta y suministrar una identificación única para el documento, bajo la cual el grupo de imágenes puede ser almacenado como una sola entidad. Al suministrar tal identificación, la aplicación debería revisar el carácter único contra la lista en línea de las identificaciones existentes.

En una tercera ventana, la aplicación debería dar al operador acceso al catálogo local en línea, crear una entrada apropiada para el documento en su forma digital e indicar la identificación única mediante la cual se puede encontrar el documento digital. Cuando se ha concluido el trabajo bibliográfico, el operador podría tener la opción de suministrar información que relacione la secuencia de imágenes con una enumeración o secuencia de páginas. El operador también podría indizar las divisiones estructurales significativas dentro del documento. Entonces, este operador estaría en capacidad de guardar la información referente a la estructura y al contenido en un archivo de base de datos, el cual es distinto, pero apunta al conjunto de imágenes que constituye el documento preservado. Si existen dos copias de un documento en diferentes resoluciones, el archivo de estructura debería apuntar correctamente hacia las copias como instancias diferentes del mismo documento con la misma estructura.

Una vez creado el archivo de estructura, e incluso después de que éste haya sido almacenado en forma permanente, el operador de indización debería tener la posibilidad de recuperarlo y modificarlo. Más aún, cuando el operador ha creado un índice por número de página para un documento específico, debería estar en capacidad de ingresar en

forma inmediata un comando para ir a una página específica. Igualmente, cuando el operador ha creado un índice de la estructura interna del documento, la aplicación debería responder inmediatamente al operador cuando éste solicite ir a la imagen inicial de cada división estructural. La interfaz tendría también que dar al operador una función especial de "marcapágina", que marcaría una imagen específica y autorizaría el regreso directo a ella en cualquier momento; no obstante, la colocación de marcapáginas por parte del operador no debería guardarse en la versión permanente del archivo de estructura.

Al aplicar la función de consignación, el operador verifica que el proceso de conversión ha concluido y que el conjunto resultante de imágenes digitales ya puede ser almacenado. La función da por sentado, como un asunto de seguridad y conveniencia del usuario, que los documentos digitalizados y la correspondiente base de datos con el archivo de estructura están almacenados en un sitio céntrico y accesible a la red, y no, por analogía con el microfilme, en archivadores cerca de las estaciones de trabajo de revisión. En caso de que el propósito de la conversión no sea crear una copia permanente del documento en forma digital sino imprimirla, la función de consignación pudiera cerrar el proceso al transferir las imágenes, por medios removibles o a través de la red, a la impresora digital de alta velocidad.

Con respecto al almacenamiento en discos en el proceso de conversión, cada estación de trabajo que controla la digitalizadora requiere suficiente espacio en el disco magnético para que el operador almacene y manipule las imágenes escaneadas no comprimidas de al menos dos libros (aproximadamente 600 *megabytes*). En cada estación de trabajo que realiza las funciones de indización y consignación, es necesario que haya espacio en el disco magnético para almacenar las imágenes comprimidas de, por lo menos, dos libros (aproximadamente 100 *megabytes*). En las estaciones de trabajo que realizan la función de consignación, debería

haber un medio de almacenamiento portátil, como una unidad para disco óptico WORM - *write once, read many* ("escribir una vez, leer muchas") que tenga un disco, sobre el que las imágenes de los documentos digitalizados comprimidos puedan escribirse, y luego transferirse físicamente al sitio de almacenamiento permanente en el disco. Alternativamente, es necesario que exista la posibilidad de escribir a un medio de almacenamiento masivo a través de la red de alta velocidad.

● Almacenamiento

El propósito del subsistema de almacenamiento es brindar un medio (o combinación de medios) con gran capacidad, confiable y de pronta respuesta, en el cual almacenar una colección de documentos que ha sido preservada en forma digital. Un operador de conversión debe estar en capacidad de convertir un documento digital al subsistema de almacenamiento y, desde allí, debe ser posible la recuperación de documentos, ya sea para imprimirlos o para revisarlos en una estación de trabajo. Los componentes del subsistema de almacenamiento incluyen dispositivos para el almacenamiento masivo, un controlador de almacenamiento masivo, un archivo de estructuras de documento y el servidor de base de datos de imágenes.

Supongamos que el proceso de conversión crea cada archivo de imágenes con una resolución de 400 *dpi* y que en la forma comprimida cada imagen ocupa 150 kilobytes de almacenamiento. Asumamos también que cada libro convertido consta de 300 páginas. A partir de estas suposiciones, y por simple multiplicación, tenemos que una biblioteca digital para los 10.000 libros previstos para su conversión en el proyecto de Yale requerirá aproximadamente 450 gigabytes de almacenamiento masivo.

Uno da por sentado que hoy en día el disco óptico WORM es el medio más efectivo, desde el punto de vista de los costos, para almacenar este volumen de información.¹⁹ No obstante, si se creara una segunda copia de cada documento con menor resolución para fines de revisión, los requerimientos de

almacenamiento pudieran ser considerablemente más bajos para la segunda copia (alrededor de 125 GB) que para la primera, y su utilización pudiera ser lo bastante frecuente como para justificar los costos de almacenar los archivos de imágenes en disco magnético, el cual ofrece tiempos de acceso más rápidos que el medio óptico. Los cartuchos WORM, que contienen las versiones en más alta resolución para impresión, pudieran instalarse en un dispensador automático, o en una línea de espera, que requiera un operador para instalarlos cuando sean solicitados. La información guardada en medios magnéticos de almacenamiento deberían respaldarse (*backup*) regularmente en cintas como medida de precaución contra la pérdida de datos producto de fallas de los discos. Los discos almacenados en medios WORM deberían copiarse periódicamente en nuevos medios ya que los formatos o equipos de acceso se vuelven obsoletos y deben ser reemplazados.

El controlador del almacenamiento masivo en el subsistema ofrece una interfaz entre el almacenamiento en disco y el servidor de imágenes. Traduce las solicitudes de imágenes en comandos que las localizan y recuperan desde el dispositivo de almacenamiento magnético o el dispensador automático para discos ópticos, según corresponda. En caso de que el documento esté almacenado fuera de línea, el controlador generará una solicitud al operador para que instale el cartucho correspondiente. Para las solicitudes de almacenamiento de imágenes digitales de documentos, en vez de solicitudes para recuperarlos, el controlador localizará el espacio, almacenará adecuadamente los archivos y actualizará el directorio de archivos.

El archivo de estructura del documento contiene información que indiza los aspectos relacionados con la estructura y el contenido de cada documento digitalizado. Es un componente integral de la biblioteca de imágenes y toda solicitud de un documento debe ir acompañada de información sobre él, y viceversa. Cualquier documento digitalizado que está siendo almacenado por primera vez

debe llevar un componente de índice para guardarlo en el archivo de estructura. El archivo es una base de datos relacional que cumple con la norma SQL. Por ser un archivo en línea, requiere respaldo (*backup*) regular.

El servidor para la base de datos de imágenes conecta al archivo de estructura de documentos con los archivos de documentos digitalizados a través del controlador de almacenamiento. Es accesible por la red y, en respuesta a las solicitudes de imágenes, recupera la información correspondiente desde el archivo de estructura y los archivos pertinentes de la biblioteca de imágenes, para luego llevarlos hasta el subsistema de acceso o la impresora, según corresponda. El servidor también responde a las solicitudes donde se requiere almacenar documentos digitalizados, garantizando que el archivo de estructura esté actualizado y que los archivos de imágenes sean convertidos y transferidos al correspondiente dispositivo de almacenamiento masivo. El servidor necesita una cantidad sustancial de memoria caché para guardar los documentos que entran y salen. Una función importantísima del servidor en este proceso consiste en dejarle saber a sus usuarios la condición de los datos, la información referente al progreso de la transmisión y los errores en el almacenamiento o la recuperación.

- **Acceso**

El propósito del subsistema de acceso es ofrecer un sitio para que las personas recuperen y usen la biblioteca de documentos preservados en forma de imagen. Los componentes del subsistema incluyen un servidor de presentación de imágenes y las estaciones para visualizar las imágenes. En el caso más sencillo del subsistema, las imágenes de los documentos digitalizados requieren ser presentadas para su revisión en un solo tipo de estación de trabajo, que suministra la biblioteca y cuya selección está controlada por la misma. El caso más complicado ocurre cuando la biblioteca da acceso a las imágenes de los documentos digitalizados a los investigadores del campus, quienes utilizan

diferentes tipos de estaciones de trabajo con diferentes capacidades para mostrar imágenes. En cualquier caso, las funciones mínimas provistas por una estación de revisión deberían ser las mismas. La estación debería permitir al lector recuperar la imagen de un documento digital por su identificación única, y mostrar y revisar el documento dependiendo de la forma en que está descrito en el archivo de estructura.

Para que pueda ser “capaz de procesar imágenes”, cada estación de revisión necesita tener la capacidad de procesamiento para leer, expandir y mostrar una página digitalizada comprimida. La estación requiere un monitor de alta resolución capaz de mostrar una imagen de la página legible. Cada estación de trabajo debería brindar funciones generales de manipulación de imagen, como acercamiento, reducción, toma panorámica, embobinado y rotación. Finalmente, si cuentan con el equipo necesario, las estaciones de revisión deberían ser capaces de transferir imágenes a las impresoras láser locales para imprimir copias de páginas específicas.

Desde la estación de trabajo de acceso, el lector debe estar en capacidad de buscar en el catálogo en línea para determinar la identificación única de un documento. Al tener la identificación, el lector debería entonces ser capaz de solicitar el documento a un servidor de imágenes y llevarlo a su monitor para revisarlo. Si no hay información adicional en el índice, el lector debería tener la posibilidad de retroceder y adelantar a lo largo de la secuencia de los archivos de imágenes, e ir directamente a cualquier imagen específica relacionada con la imagen actual.

No obstante, si existe un índice por número de páginas, el lector debería estar en capacidad de usar un comando para ir a una página específica. Igualmente, si existe un índice de la tabla de contenido, las partes y los capítulos del documento, así como de otras partes importantes de su estructura interna, el programa debería brindar acceso inmediato a la imagen que marca el inicio de cada división estructural. La interfaz también

debería permitir al lector utilizar una función especial de marcapágina, la cual marcaría una imagen determinada y brindaría un retorno inmediato y directo a ella en cualquier momento; sin embargo, la colocación de estos marcapáginas por parte del lector sólo debería tener vigencia mientras dure su sesión de revisión.

El servidor de presentación de imágenes separa estas diversas funciones de acceso, que todas las estaciones de revisión deben brindar en conjunto, de las diferentes formas en que los sistemas operativos de las diversas plataformas de estaciones de trabajo (DOS, UNIX, Macintosh) permiten presentarlas al lector de manera eficaz. Al comenzar la sesión de revisión, el servidor consultará una estación de revisión particular para determinar su tipo y sus capacidades. Pudiera ser que la estación no estuviese debidamente equipada para revisar imágenes y que, en consecuencia, el servidor respondiera conforme a ello. Por el contrario, cuando el servidor de imágenes responde a la solicitud de un documento, el servidor de presentación comienza a trabajar. Tomará el archivo de estructura del documento, el cual describe los aspectos críticos del documento y, utilizando las convenciones de la *Office Document Architecture* (ODA), suministrará un conjunto de indicaciones de alto nivel sobre cómo debe presentarse el contenido del documento. El servidor traduce estas indicaciones en un formato y en un conjunto de instrucciones de programación que son adecuadas para esa estación de trabajo en particular. Utilizando un controlador que obtiene del servidor, la estación de revisión, a su vez, procesa las instrucciones programadas y genera un formato de presentación con el cual el lector pueda revisar fácilmente la imagen del documento digital.

● Reproducción

El propósito del subsistema de reproducción es generar copias impresas, de muy alta calidad, de una parte o la totalidad de los documentos seleccionados que están almacenados en forma de imágenes digitales.

Los componentes del subsistema incluyen la impresora, un servidor de impresión y una unidad para solicitar impresiones. La impresora debe estar en capacidad de aceptar como insumos archivos de imágenes con muy alta resolución (600 puntos por pulgada o más). Debe generar un resultado impreso con la misma resolución. Para una mayor economía, debería funcionar a velocidades muy altas (100 páginas por minuto o más, según un criterio de referencia utilizado actualmente), y brindar capacidades incorporadas de cotejo y encuadernación.

El servidor de impresión maneja la cola de espera para la impresora, suministrando diversos controles de operación que incluyen la posibilidad de establecer prioridades de trabajo, averiguar la condición de una tarea y cancelar labores. Debería brindar una considerable capacidad de transferencia y de aceptar archivos comprimidos TIFF, de acuerdo con la norma CCITT Grupo 4. Pudiera necesitar reformatear tales archivos a un formato interno, pero debería poder hacerlo sin que haya pérdida o menoscabo en la calidad de la información. Para dar cabida a las solicitudes de impresión directamente desde el subsistema de conversión, el servidor de impresión quizás necesite tener una unidad compatible WORM para manejar el insumo directo del documento a imprimir.

Los usuarios que desean imprimir una parte o la totalidad de un documento determinado, generalmente lo harán invocando un servicio de red para presentar solicitudes de impresión. Una persona puede solicitar una copia impresa de la imagen de un documento digital por la identificación única especificada en el catálogo en línea. El servicio de solicitudes recuperará la información de estructura referente a ese documento y le pedirá a la persona que determine cuáles porciones o qué páginas imprimir, y si desea que las encuadernen. El servicio de solicitudes de impresión debería incorporar aplicaciones de contabilidad para poder facturar a los usuarios las copias que solicitan. El uso general del servicio de solicitud en la red para fines de facturación da por sentado que hay una manera de identificar a los usuarios y verificar

que son quienes dicen ser. Suponiendo que se contempla el aspecto de la identificación, una extensión natural del servicio de solicitudes de impresión sería utilizarlo para cobrar los montos correspondientes a la reproducción de materiales cubiertos por derechos de autor, tales como artículos recientes de diarios, que eventualmente pudieran sumarse a la biblioteca digital.

Aspectos relacionados con la factibilidad

Como hemos visto, está surgiendo una serie de normas para regular la arquitectura de los sistemas para la digitalización de imágenes. Más aún, se está difundiendo una opinión generalizada acerca de la naturaleza y la interrelación de los componentes del sistema en la arquitectura. Juntos, estos factores generan confianza en nuestra habilidad para lograr la visión de un sistema de digitalización de imágenes de preservación a gran escala en Yale, en el cual el material, convertido de microfilme a imagen digital, se hace más accesible y valioso para la comunidad de investigadores.

Sin embargo, el optimismo en cuanto a la factibilidad general de crear un sistema viable sobre la base de una biblioteca de imágenes de 10.000 volúmenes no debería opacar los riesgos inherentes. No todos los aspectos relacionados con los componentes del sistema y sus interrelaciones son tan bien comprendidos como otras partes; todavía hay muchas incógnitas e incertidumbres con respecto a la tecnología y su funcionamiento. Invocando un principio de diseño que citamos anteriormente -que la información creada y almacenada es el principal activo de una biblioteca en un sistema automatizado-, y preguntando de qué manera los datos correrían mayores riesgos en esta arquitectura en particular, podemos identificar por lo menos cuatro áreas de interés que pudieran limitar o impedir el desarrollo de un sistema, y requerir, por lo tanto, especial atención.

En primer lugar, el área de principal interés es la integración de dispositivos de almacenamiento masivo. La tecnología del

almacenamiento óptico en WORM generalmente se comprende bien, pero todavía no han surgido formatos normalizados para los medios. Además, la industria todavía no cuenta con una rica experiencia en lo que respecta al verdadero uso de la tecnología y, particularmente, sobre la forma de optimizar una combinación de almacenamiento óptico y magnético para aplicaciones particulares y, por consiguiente, para mejorar el tiempo de acceso y aligerar la competencia de las unidades en un ambiente con múltiples usuarios.

En vista de eso, la aplicación que la biblioteca espera desarrollar pareciera ser especial. Las aplicaciones para la digitalización de imágenes han ganado terreno en los ambientes de oficina, donde el documento modelo es una carpeta de archivo que contiene unos cuantos artículos de consulta rápida, para luego pasar a otro punto. No obstante, el documento modelo en una biblioteca es el libro con varios cientos de páginas, sobre el cual uno puede clavar la mirada por algún tiempo, retrocediendo y avanzando ocasionalmente en el texto para consultar puntos relacionados, citas y cosas por el estilo. Pareciera razonable esperar que los diferentes patrones de uso de documentos en la biblioteca enfatizarán nuevas y diversas clases de mecanismos de almacenamiento, los cuales requerirán soluciones que quizás todavía no estén disponibles. La interrogante radica en si la experiencia confirmará o no esta hipótesis.

El segundo aspecto que sigue siendo motivo de preocupación es el de refrescar periódicamente los datos almacenados en un medio óptico. Hasta ahora los centros de computación y los proveedores de computadores han atravesado por numerosos ciclos de cambios técnicos en los medios magnéticos, haciendo necesario que la información tenga que volver a copiarse periódicamente. Sin embargo, los ciclos de cambio en la tecnología óptica apenas han comenzado a afectar los grandes centros de producción de datos. Como la tecnología es diferente y el volumen de datos almacenados en un disco óptico es, en orden de magnitud, mucho mayor que el de los medios

magnéticos, los familiares paradigmas de renovación de datos pudieran no ser válidos. Los proveedores indudablemente desempeñarán un papel clave en la promulgación de soluciones apropiadas y prácticas; pero, si se les pregunta al respecto, pocos de ellos son capaces de articular tales soluciones.

La tercera gran área de interés es el llamado archivo de estructura del documento. En la medida en que la biblioteca indice sus documentos digitalizados por número de página o por la estructura interna del documento, una gran parte de su trabajo e inversión estará vinculada a este archivo de estructura. Las normas ODA (*Office Document Architecture*) y SQL parecen apropiadas para gobernar su naturaleza y organización, así como también para hacer que la información contenida en ellas sea convertible a otro sistema o implementación, según sea necesario. No obstante, ya es evidente que este modo de indizar el contenido de las imágenes requerirá algunas revisiones y extensiones en la norma ODA. La inversión en la biblioteca de imágenes y en los diversos índices incorporados en el archivo de estructura de documento, sin duda alguna garantizará la participación activa de la biblioteca de Yale, no sólo en la implementación de la ODA, sino también en sus extensiones y desarrollo.

Finalmente, un cuarto punto de potencial preocupación emana del proceso de control de calidad. La comunidad bibliotecaria ha esperado durante años que, existiendo la tecnología correcta, el microfilme fuese fácilmente convertible al formato digital. El proyecto de Yale generará con seguridad información sumamente útil acerca de la facilidad (o dificultad) técnica del proceso de conversión. Pero la calidad del producto resultante de ese proceso dependerá no sólo de los mecanismos y la magia técnica del propio proceso de conversión, sino también de la calidad del microfilme, calidad necesaria para que las imágenes digitales sirvan como insumos para los consiguientes procesos técnicos -tales como el reconocimiento óptico (OCR) de caracteres-, y de los patrones de calidad que hagan valer los usuarios de bibliotecas que utilizarán las

imágenes digitales en su trabajo de investigación. Por consiguiente, en este proyecto será importante captar y evaluar las diferentes maneras por las cuales la preservación en microfilme pueda o deba modificarse para generar una mejor imagen digital. Igualmente, sería importante encontrar formas de evaluar la calidad del proceso de digitalización desde el punto de vista de los procesos técnicos relacionados, tales como el reconocimiento de caracteres, y será absolutamente esencial validar el proceso desde la perspectiva de la satisfacción de los usuarios.

PLAN DE TRABAJO

El proyecto de preservación digital de Yale se basará en trabajos afines realizados en la Library of Congress, la National Medicine Library, la Cornell University y otras instituciones bibliotecarias y archivísticas, concentrándose en la conversión del microfilme a imágenes digitales, e intentando conformar una biblioteca digital de 10.000 volúmenes en el proceso. Buena parte del equipo y de la aplicación necesarios para construir un sistema en Yale ya existen en el mercado de proveedores o se encuentran en avanzadas etapas de desarrollo. La excepción principal es el administrador de presentación de imágenes descrito en la sección anterior, del cual Yale tiene experiencias especiales para desarrollar, y que será generado como un producto específico del sistema de trabajo en Yale.

Por supuesto, las preocupaciones en torno a varios aspectos ambiguos e inciertos de la arquitectura de los sistemas de digitalización de imágenes destacan los riesgos de invertir en tecnología relativamente nueva y costosa, en la escala contemplada en el proyecto de Yale. Sin embargo, tales preocupaciones también concentran la atención en las áreas de oportunidades que ofrece un proyecto del alcance y la envergadura de éste en Yale, para entender el potencial y los límites de la tecnología, para desarrollar ciertos aspectos de la tecnología y, en general, para reducir considerablemente los riesgos de inversión para aquellas bibliotecas que siguen la senda de la digitalización de imágenes. A fin de mitigar los riesgos del proyecto y mejorar sus perspectivas de producir un rendimiento valioso, una de las soluciones consiste en delinear un plan de implementación deliberado y cuidadoso que aclare metas, resultados medibles y un enfoque por fases, en el cual no todo esté perdido, ni siquiera necesariamente en peligro, si uno encuentra un problema u otros obstáculos.

Metas del proyecto

El proyecto de preservación mediante digitalización de imágenes planificado para

la Yale University Library tiene tres amplias metas:

- Identificar, desarrollar y evaluar los medios para crear y almacenar una biblioteca digital de materiales preservados, a través de la conversión de imágenes microfilmadas al formato digitalizado a una escala de producción de 10.000 volúmenes.
- Identificar, desarrollar y evaluar los medios para suministrar, conservar y mejorar la distribución o el acceso físico a los materiales preservados en forma de imagen digital, tanto dentro como fuera de la biblioteca.
- Identificar, desarrollar y evaluar los medios para suministrar, preservar y mejorar el acceso intelectual a materiales de la biblioteca en forma de imágenes digitales.

Alcance del proyecto

El proyecto estará dividido en seis fases en el transcurso de tres años (ver **Ilustración 7**). La primera fase se extenderá del primer al cuarto mes (1-4) y establecerá el marco organizativo general del proyecto. Durante la segunda fase, que abarcará del quinto al octavo mes (5-8), se hará la instalación inicial del equipo y del programa. En la tercera fase se preparará el proceso a nivel de producción para convertir el microfilme en imágenes digitales. Comenzará a partir del noveno (9) mes y terminará en el décimo octavo (18). La cuarta fase del proyecto también comenzará en el noveno mes, pero concluirá en el vigésimo primero (21), cuando los documentos en forma de imágenes digitales sean accesibles para los lectores dentro de la biblioteca de Yale. La quinta fase irá del vigésimo segundo al trigésimo tercero (22-33) y brindará el acceso a los documentos en forma de imágenes digitales a los lectores fuera de la biblioteca de Yale. El proyecto finalizará en la sexta fase, desde el trigésimo cuarto al trigésimo sexto mes (34-36), al concluir la

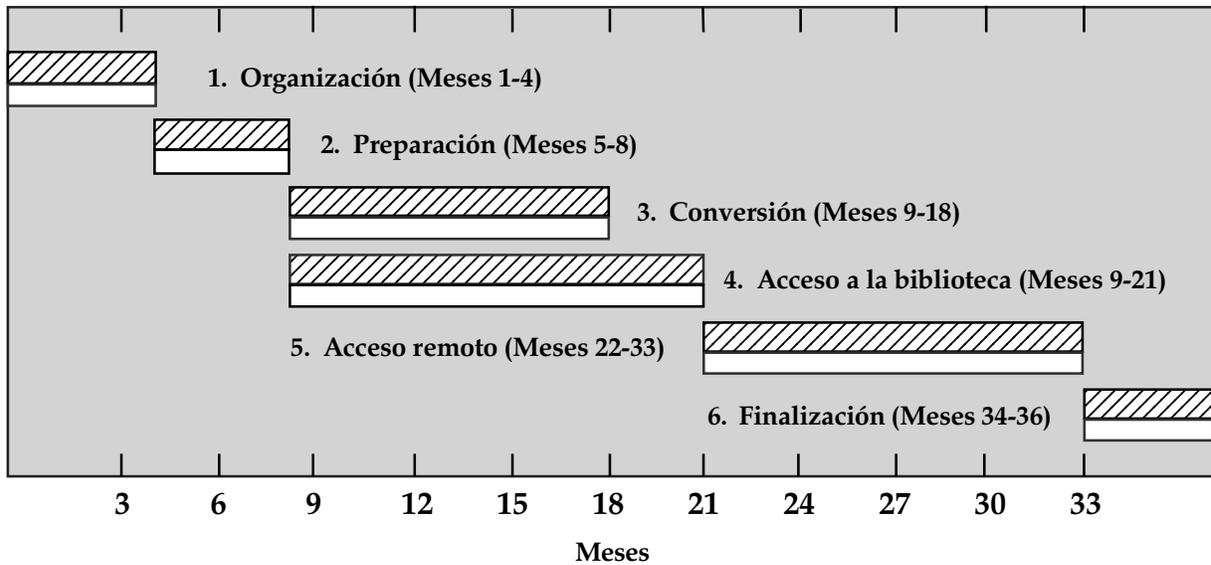


Ilustración 7. Fases del proyecto

conversión de los 10.000 volúmenes, además de dar los toques finales y hacer la última evaluación a todos los componentes del sistema.

En cada etapa, se preparará un informe para resumir los resultados obtenidos hasta la fecha y evaluar los méritos generales del proyecto. Debido a que la tecnología de digitalización de imágenes es objeto de investigación y demostración en este proyecto, todavía sujeta a rápidos cambios y progresos, es posible, aunque poco probable, que los resultados buscados aquí surjan en cualquier otra parte o que las aplicaciones de la tecnología misma prueben las suposiciones sobre las cuales se basa el proyecto. Por lo tanto, cada informe tratará el final de esa fase como un hito, en el cual formulará y responderá la siguiente pregunta: ¿vale la pena continuar con este proyecto?

Fase 1: Organización

Objetivo: El objetivo de la fase 1 es crear un marco organizativo general para el proyecto y preparar el trabajo de las siguientes fases.

Lapso: Del primer al cuarto mes.

Actividades: Durante la Fase 1, se creará

un Comité Directivo que se encargue de la supervisión y dirección general del proyecto. El comité estará compuesto, al menos inicialmente, por el director de la University Library, el director adjunto de la University Library, la Asociación de Bibliotecarios de la Universidad para servicios técnicos, servicios públicos y desarrollo de colecciones, el jefe del Departamento de Preservación, el director de Computación y Sistemas de Información de la Universidad, el director académico de Computación, el gerente del proyecto y cuatro miembros del cuerpo académico de la facultad. El comité puede ajustar el número de sus miembros según considere necesario.

La agenda inicial del comité incluirá los siguientes puntos:

- establecer los criterios para identificar qué porciones de la colección de documentos preservados en microfilme serán convertidas en imágenes digitales;
- brindar asistencia al gerente del proyecto y a los miembros del equipo del proyecto para establecer relaciones de asociación adecuadas con proveedores de equipos y aplicaciones para la digitalización de imágenes;

- revisar el presupuesto del proyecto y brindar asistencia para identificar y garantizar las fuentes de financiamiento.

El gerente del proyecto será el jefe de la Oficina de Sistemas de la biblioteca. El equipo del proyecto estará conformado inicialmente por el jefe del Departamento de Preservación de la biblioteca y el director adjunto de Sistemas, Tecnología y Planificación del Departamento de Computación y Sistemas de Información. Durante esta fase, el equipo del proyecto:

- identificará el equipo y los programas específicos que serán adquiridos para el proyecto;
- establecerá relaciones de asociación adecuadas con los proveedores de equipos y aplicaciones para la digitalización de imágenes que serán adquiridos para el proyecto;
- preparará un plan de contratación de personal para todas las fases siguientes del proyecto;
- preparará un presupuesto para todo el proyecto y trabajará con el Comité Directivo y otras fuentes pertinentes para conseguir financiamiento.

Requisitos: Este documento, "Del microfilme a la imagen digital" brindará los antecedentes y las bases de planificación para el trabajo inicial del Comité Directivo y del Equipo del Proyecto.

Productos:

- Un presupuesto de proyecto.
- Un plan de financiamiento.
- Un informe que resuma los resultados del trabajo en esta fase.

Fase 2: Preparación

Objetivo: El objetivo de la fase 2 es comenzar a preparar el ambiente laboral necesario para el proyecto.

Lapso: Del quinto al octavo mes.

Actividades: El equipo del proyecto:

- aplicará los criterios establecidos por el Comité Directivo y comenzará a seleccionar las porciones de la colección microfilmada que serán convertidas;
- creará y comenzará a implementar un plan para evaluar si los criterios utilizados para seleccionar el material para la conversión generan una biblioteca digital, que los usuarios de la biblioteca consideren intelectualmente valiosa;
- contratará el personal necesario;
- adquirirá, instalará y probará el equipo y el programa para el subsistema de conversión;
- adquirirá, instalará y probará la impresora digital de alta velocidad y el servidor de impresión;
- adquirirá, instalará y probará el componente del archivo de estructura de documento del subsistema de almacenamiento, y comenzará a adquirir, instalar y probar los otros componentes del subsistema de almacenamiento; y
- preparará el diseño y las especificaciones iniciales para el servidor de presentación de imágenes.

Además, el equipo contratará una compañía de servicios para convertir del microfilme al formato de imágenes digitales, hasta un máximo de diez volúmenes que serán seleccionados aleatoriamente. El contrato establecerá una línea base de costos, con la cual comparar los costos del procesamiento

por parte de la propia biblioteca. Esto también brindará experiencia para formular y aplicar criterios de aceptación para el proceso interno de control de calidad. Supuestamente, en esta etapa, los documentos serán devueltos en forma de imágenes digitales sin indización .

Requisitos: Para que esta fase pueda comenzar, debe existir un presupuesto aprobado con los fondos requeridos para cubrir gastos de personal, la adquisición de equipo y programa, y los costos de la compañía de servicios.

Productos:

- 10 volúmenes disponibles en formato digital, listos para el proceso de indización dentro del subsistema de conversión.
- Un informe que resuma los resultados del trabajo en esta fase, incluyendo un análisis de la experiencia con la compañía de servicios y un plan para evaluar sistemáticamente, durante el lapso de duración del proyecto, los méritos de la conversión interna frente al uso de una compañía de servicios.

Fase 3: Conversión

Objetivo: El objetivo de la fase 3 es establecer el proceso a nivel de producción para convertir el microfilme a imágenes digitales.

Lapso: Del noveno al decimoctavo mes.

Actividades: Entre las principales tareas para el equipo del proyecto en esta fase se contemplan:

- desarrollar procedimientos y capacitar a los operadores en el uso de la digitalizadora y la estación de trabajo para el control de calidad, incluyendo sus características para mejorar imágenes;
- seleccionar para su digitalización, a

principios de esta fase, una muestra de los rollos de microfilme que presenten variaciones en la calidad del proceso de microfilmación. Diseñar un método a fin de utilizar esta muestra tanto para realizar el proceso de control de calidad como para comenzar a formular conclusiones sobre las formas en que las técnicas de microfilmación afectan la calidad del proceso de digitalización;

- diseñar e implementar un patrón de control de calidad que tome en cuenta, particularmente, la futura posibilidad de aplicar tecnología para el reconocimiento de caracteres a las imágenes generadas en este proceso de conversión;
- crear un proceso continuo para medir la satisfacción de los usuarios con los procesos y normas de control de calidad;
- revisar el diseño del índice por número de páginas y el índice por estructura del documento; dicho diseño debe garantizar que en el futuro se puedan crear lazos entre estos índices y notas adicionales sobre el contenido, las cuales se pudieran almacenar en otra parte bajo un formato de lectura automática;
- desarrollar procedimientos y capacitar a los operadores en el uso de cada una de las cuatro funciones para la indización de documentos. Los operadores deberían estar en capacidad de (1) asignar una identificación única al conjunto de imágenes convertidas que constituyen una entidad documental única, (2) indizar el documento digitalizado por número de página, (3) indizar el documento digitalizado según la estructura interna del documento, y (4) crear una entidad bibliográfica en el catálogo en línea de Yale, que describa el documento digitalizado y lo identifique en la biblioteca digital por su identificación única;
- diseñar métodos para estructurar y

modificar el flujo de trabajo de la indización, a fin de aislar los costos de cada una de las cuatro formas de indizar un documento;

- desarrollar procedimientos y capacitar a los operadores para consignar el documento digitalizado y los índices del documento en su almacenamiento en archivo permanente, utilizando en esta etapa una unidad WORM portátil;
- revisar los criterios para seleccionar los materiales de preservación de mayor uso, a los que se hará una copia impresa, devolviéndose luego a las estanterías;
- desarrollar procedimientos y capacitar a los operadores para convertir las imágenes de determinados documentos directamente a la impresora de imágenes de alta calidad. El modo preferido de transmisión es a través de una red de alta velocidad;
- diseñar un método para estructurar el flujo de trabajo a fin de aislar los costos de la digitalización, del control de calidad y de la impresión, sin la indización;
- diseñar un método y un procedimiento continuo para comparar los costos de servicio interno versus los costos de una compañía de servicios para digitalización, control de calidad e indización ;
- indizar y consignar en almacenamiento los 10 volúmenes digitalizados por la compañía de servicios en la fase 2; y
- digitalizar, mejorar, indizar y consignar en almacenamiento permanente 490 volúmenes.

Requisitos: Para que esta fase comience, el subsistema de conversión, la impresora, el servidor de impresión y el componente del archivo de estructura de documento en el subsistema de almacenamiento deben haber sido instalados y probados.

Puntos que deben investigarse: El trabajo en esta fase está diseñado para producir parámetros críticos de calidad y costos en varios aspectos, incluyendo:

- los posibles efectos de la calidad del microfilme en la calidad de la digitalización;
- las implicaciones que tendría la futura tecnología para el reconocimiento de caracteres en el proceso de control de calidad en la digitalización de microfilmes;
- la satisfacción del usuario con los procesos y normas del control de calidad;
- los costos relativos del servicio interno en comparación con los costos de una compañía de servicios para la digitalización, el control de calidad y la indización;
- los costos variables de los diferentes niveles de indización; y
- los costos de simplemente digitalizar e imprimir sin indizar.

Será importante recabar y supervisar estos diferentes parámetros de manera continua a lo largo del proyecto.

Productos:

- Funcionamiento del subsistema de conversión en los niveles de producción.
- 500 volúmenes digitalizados y completamente indizados.
- Un informe que resuma los resultados del trabajo en esta fase, incluyendo un análisis de costos.

Fase 4: Acceso a la biblioteca

Objetivo: El objetivo de la fase 4 es que la creciente biblioteca de imágenes digitales

sea accesible para los lectores en las estaciones de trabajo ubicadas dentro de la biblioteca de Yale.

Lapso: Del noveno al vigésimo primer mes.

Actividades: Las principales tareas para el equipo del proyecto en esta fase incluyen:

- completar la adquisición, instalación y prueba del subsistema de almacenamiento;
- especificar, adquirir, instalar y probar las estaciones de trabajo de acceso en la biblioteca;
- desarrollar y probar un prototipo de servidor para la presentación de imágenes, capaz de interactuar con las estaciones de trabajo de acceso de la biblioteca;
- elaborar e implementar un estudio para medir el valor que tiene para los lectores de la biblioteca cada una de las herramientas de indización disponibles para los documentos digitales, en comparación con el acceso en microfilme;
- crear e implementar un plan para respaldar los datos almacenados (*backup*) en medios magnéticos y para renovar periódicamente los datos almacenados en medios ópticos y magnéticos;
- transferir todos los volúmenes digitalizados al sistema de almacenamiento permanente;
- diseñar un método para evaluar los costos de añadir el subsistema de almacenamiento y de darle mantenimiento en forma continua;
- establecer e implementar métodos y procedimientos continuos para evaluar los efectos de la carga de acceso en el rendimiento del subsistema de almacenamiento de imágenes;

- diseñar e implementar métodos y procedimientos continuos para poner a punto el rendimiento del subsistema de almacenamiento de imágenes;
- establecer e implementar métodos y procedimientos continuos para supervisar y ajustar los efectos de la transmisión de imágenes en el rendimiento de la red;
- diseñar métodos para evaluar los costos de crear y mantener un subsistema de acceso;
- adquirir, instalar y probar la unidad para solicitar copias;
- diseñar métodos para evaluar los costos de crear y mantener el subsistema de impresión; y
- después de concluir la fase 2, y del decimosexto al decimooctavo mes, digitalizar, mejorar, indizar y consignar en almacenamiento 1.300 volúmenes.

Requisitos: Para poder iniciar esta fase, la impresora, el servidor de impresión y el componente del archivo de estructura de documento en el subsistema de almacenamiento deben estar instalados y haber sido probados. El diseño inicial y las especificaciones para el servidor de presentación de imágenes también deben estar completos.

Puntos que deben ser investigados: El trabajo en esta fase está diseñado para producir parámetros críticos de calidad y costos en varios aspectos, incluyendo:

- los costos progresivos de crear y mantener un subsistema de almacenamiento;
- los costos de hacer respaldos (*backup*) de los datos almacenados en un disco magnético y de renovación periódica de los datos almacenados en medios ópticos y magnéticos;

- los costos progresivos de crear y mantener el subsistema de acceso;
- los costos progresivos de crear y mantener el subsistema de impresión;
- los efectos de la carga de acceso en el subsistema de almacenamiento;
- los efectos de la carga de transmisión en la red; y
- el valor de los índices de los documentos para los lectores de la biblioteca.

Será importante recabar y monitorear estos diversos parámetros de manera continua a lo largo del proyecto.

Productos:

- El subsistema de almacenamiento de imágenes funcionando en niveles de producción.
- El subsistema de impresión funcionando en niveles de producción.
- Prototipo del administrador de presentación de imagen.
- El subsistema de acceso operando en niveles de producción en la biblioteca.
- Servidor para solicitudes de impresión.
- 1.300 volúmenes digitalizados y completamente indizados.
- Un informe que resuma los resultados del trabajo en esta fase, con inclusión de un análisis de costos.

Fase 5: Acceso remoto

Objetivo: El objetivo de la fase 5 es establecer el acceso a los documentos digitalizados desde el exterior de la biblioteca de Yale.

Lapso: Del mes vigésimo segundo al trigésimo tercero.

Actividades: Entre las principales tareas para el equipo del proyecto en esta fase, se contemplan:

- extender el funcionamiento del servidor para la presentación de imágenes, a fin de que esté en capacidad de interactuar con determinados tipos de estaciones de trabajo ubicadas fuera de la biblioteca;
- a través del servidor para presentación de imágenes, lograr que el subsistema de acceso a los documentos digitalizados esté disponible para determinados tipos de estaciones de trabajo que están conectadas a la red del campus;
- diseñar métodos para estimar los costos iniciales y continuos de introducir el acceso al sistema de imágenes digitalizadas desde fuera de la biblioteca;
- diseñar e implementar un estudio para medir el valor que tiene, para los usuarios fuera de la biblioteca, cada una de las herramientas de indización existentes para el documento digital; y
- digitalizar, mejorar, indizar y consignar en almacenamiento 7.800 volúmenes.

Requisitos: Para que esta fase pueda comenzar, la versión prototipo del servidor para presentación de imágenes debe estar concluida y operativa.

Puntos que se deben investigar: El trabajo en esta fase está diseñado para producir mediciones críticas del valor que los lectores externos de la biblioteca otorguen a la indización de los documentos. Será importante recabar y supervisar estas medidas de manera continua durante ésta y la última fase del proyecto.

Productos:

- Servidor para presentación de imágenes en nivel de producción.
- Funcionamiento del subsistema de acceso para estaciones de trabajo ubicadas fuera de la biblioteca.
- 7.800 volúmenes digitalizados y completamente indizados.
- Un informe que resuma los resultados del trabajo en esta fase, incluyendo un análisis de costos.

Fase 6: Finalización

Objetivo: El objetivo de la fase 6 es completar la conversión de los 10.000 volúmenes y realizar un último ajuste y una evaluación final de todos los componentes del sistema.

Lapso: Del trigésimo cuarto al trigésimo sexto mes.

Actividades: Las principales tareas para el equipo del proyecto en esta fase incluyen:

- revisar y evaluar todos los parámetros de calidad y rendimiento en cada uno de los subsistemas; y
- digitalizar, mejorar, indizar y consignar en almacenamiento 400 volúmenes.

Requisitos: Para que esta fase pueda comenzar, la versión de producción del servidor para presentación de imágenes debe estar concluida y en funcionamiento.

Productos:

- Un informe final.
- 400 volúmenes digitalizados y completamente indizados.

CONCLUSIÓN

La Yale University Library avizora un futuro en el cual la tecnología de la imagen digital constituya una herramienta crucial en el proceso de preservar el acceso a materiales en condición de deterioro en su rica y valiosa colección. Justo en un momento en el que el deterioro de documentos elaborados con papel ácido se está acelerando a una velocidad alarmante en Yale y en otras importantes bibliotecas de investigación en Estados Unidos, la arquitectura de los sistemas de digitalización de imágenes está ganando extensa comprensión y aceptación, mientras que los componentes del sistema están cada vez mejor integrados y se vuelven fácilmente asequibles. La Library of Congress, la National Medicine Library, la Cornell University y otras instituciones bibliotecarias y archivísticas han contribuido, de diversas maneras, al desarrollo de la tecnología de digitalización de imágenes y a su aplicación para preservar y mejorar el acceso general al acervo nacional e internacional de conocimiento registrado. El momento es oportuno para dar otro gran paso a fin de que la digitalización de imágenes se convierta en una herramienta bibliotecaria práctica.

El proyecto de Yale para preservar documentos en formato digital se basará en trabajos relacionados llevados a cabo en otros sitios, particularmente en la Cornell University, concentrándose en la conversión del microfilme a las imágenes digitales e intentando conformar una biblioteca digital de 10.000 volúmenes en el proceso. Buena parte del equipo y las aplicaciones necesarios para estructurar el sistema de digitalización en Yale ya está disponible o se encuentra en las etapas finales de desarrollo por parte de sus proveedores. Sin embargo, el sistema de preservación por digitalización de imágenes de Yale se caracterizará tanto por hacer un uso intensivo del capital como por hacer un uso intensivo del trabajo, particularmente en las fases de indización del proceso de conversión. Su éxito requerirá un gran apoyo por parte de las agencias de financiamiento y asociaciones significativas y productivas con

miembros de la comunidad de proveedores. El resultado para todos los participantes en el proyecto será un cuerpo sustancial de información acerca de los costos y beneficios de la digitalización de imágenes. En el futuro, las bibliotecas, y los proveedores que trabajan con ellas, dependerán de esa información para tomar decisiones prudentes y prácticas con respecto a la incorporación de la tecnología de digitalización de imágenes en la rutina diaria del trabajo de la biblioteca y del acceso de los investigadores a los materiales que ella contiene.

NOTAS

Agradecimientos: En la preparación de este informe sobre planificación, recibí ayuda muy valiosa de parte de numerosas instituciones privadas y universitarias. Son muchas para nombrarlas una por una; a todas les estoy agradecido. Sin embargo, quisiera dar gracias especialmente a mis colegas en la Oficina de Sistemas de la Yale University Library, Merri Beth Lavagnino y Greg Kaisen, por su valiosa colaboración. También agradezco a la Commission on Preservation and Access, por su apoyo al contratar a la Yale University Library para preparar este informe.

- 1 Nancy E. Gwinn, ed., *Preservation Microfilming : A Guide for Librarians and Archivists* (Chicago : American Library Association, 1987), p. xxvi. John C. Mallison, "On the Preservation of Human -and Machine-Readable Records," *Information Technology and Libraries*, 7:1 (March 1988): 22.
- 2 Michael Lesk, "Digital Imagery, Preservation and Access," *Information Technology and Libraries*, 9:4 (December 1990): 307.
- 3 *Ibid.*, 308.
- 4 Constance C. Gould, *Information Needs in the Humanities: An Assessment*, (Mountain View, Ca.: Research Libraries Group), p. 51. Constance C. Gould and Mark Handler, *Information Needs in the Social Sciences: An Assessment*, (Mountain View, Ca. : Research Libraries Group, 1989). Constance C. Gould, *Information Needs in the Sciences: An Assessment*, (Mountain View, Ca. : Research Libraries Group, 1991).
- 5 Eric Almquist, "An Examination of Work-Related Information Acquisition and Usage among Scientific, Technical and Medical Fields," resultados de un estudio comisionado por el Faxon Institute for Advanced Studies in Scholarly and Scientific Communications, presentado en la Conferencia Anual de 1991 del Faxon

Institute, *Creating User Pathways to Electronic Information*, April 1991, Reston, Virginia, Exhibit 2.

- 6 *Ibid.*, Exhibits 9, 22 y 25.
- 7 Ver por ejemplo, F.W. Lancaster, *The Measurement and Evaluation of Library Services* (Washington, D.C. : Information Resources Press, 1977), p. 313. Lancaster se refiere a una serie de estudios que respaldan el argumento de que "la selección de una fuente de información se basa casi exclusivamente en el acceso, por lo que la fuente (canal) más accesible se escoge primero; las consideraciones con respecto a la calidad y confiabilidad son secundarias, aunque estos factores son importantes en lo que concierne a influenciar el grado en que el usuario está dispuesto a aceptar información suministrada de una fuente particular".
- 8 Lawrence Dowler, "Conference on Research Trends and Library Resources," *Harvard Library Bulletin*, 1:2 (Summer, 1990): 8.
- 9 Lesk, *op. cit.*, 303.
- 10 El método de los indicadores visuales (*blipping*), es decir la colocación de marcas especiales en determinados fotogramas, puede facilitar el uso del microfilme, especialmente si las marcas están codificadas e indizadas en un sistema para la recuperación asistida por computadora. La técnica, sin embargo, ha sido aplicada principalmente en películas de 16 mm. UMI ha definido directrices de uso interno para la colocación de indicadores visuales en películas de 35 mm. Todavía está por verse si tales directrices serán ampliamente asequibles y aplicadas en forma generalizada a microfilmes de 35 mm para preservación en bibliotecas.
- 11 El reconocimiento de caracteres a partir de las imágenes escaneadas digitalmente genera un grupo de interesantes retos.

Primero, aunque los algoritmos de reconocimiento están mejorando constantemente, la precisión de la tecnología actual ha sido decepcionante. Segundo, todavía queda mucho trabajo por hacer para mejorar la habilidad de los algoritmos de reconocimiento de caracteres para tratar con la amplia variedad de tipografías e idiomas que uno encuentra en los materiales conservados en las bibliotecas. Tercero, dado el exitoso reconocimiento de los caracteres alfanuméricos que componen una imagen de texto, se necesita, para fines de almacenamiento y recuperación, relacionar los caracteres con la estructura y el formato del documento original, incluyendo cualquier fotografía, dibujo, diagrama u otras imágenes gráficas que éste pudiera contener.

La forma en que estos puntos sean tratados y resueltos afectará considerablemente el acceso a los documentos almacenados en medios electrónicos. Pretendemos tomar en consideración y estar informados acerca de las implicaciones de las decisiones que tomamos y las opciones que escogemos en este proyecto para la futura aplicación de la tecnología de reconocimiento de caracteres. De hecho, tenemos la intención de tomar todas las precauciones razonables para garantizar el éxito de dicha aplicación a las imágenes digitales que produzcamos. Sin embargo, en vista de la complejidad y la relativa inmadurez de la tecnología del reconocimiento de caracteres, consideramos que su aplicación está relacionada con el proyecto actual, pero que está decididamente fuera de su alcance.

Para una discusión útil de algunos de los aspectos relacionados con la aplicación del reconocimiento de caracteres a los materiales de bibliotecas, referirse a Stuart Weibel, John Handley y Charles Huff, "Automated Document Architecture Processing and Tagging", en: Donald L. Blamberg, Carol L. Dowling y Claudia V. Weston, eds., *Proceedings of the Conference on Application of Scanning Methodologies in Libraries*, (Beltsville, Maryland : National

Agricultural Library, 1989), pp. 3-14.

12. Ver, por ejemplo, Karen Markey, *Subject Searching in Library Catalogs: Before and after the Introduction of Online Catalogs* (Dublin, Ohio : Online Computer Library Center, 1984), pp. 75-117; Richard Van Orden, "Context-Enriched Access to Electronic Information : Summaries of Selected Research," *Library Hi Tech* 8:3 (1990): 27-32; y Marty Dillon y Patrick Wenzel, "Retrieval Effectiveness of Enhanced Bibliographic Records," *Library Hi Tech* 8:3 (1990): 43-46.

12 Sobre la continua utilidad del microfilme en la industria, ver Whitney S. Minkler, "Optical Disks vs. Micrographics. Is it an Adversarial Problem," *Micrographics and Optical Technology*, 7:4 (1989): 141-149; y John Blake, "War over Optical Disk or Microfilm Ends, Future Contains Multiple Media," *Micrographics and Optical Technology*, 8:3 (1990): 141-143.

14 Ver, por ejemplo, Clifford A. Lynch y Edwing B. Brownrigg, "Library Applications of Electronic Imaging Technology," *Information Technology and Libraries*, 5:2, (June 1986): 100-102.

15 Lesk, *op. cit.*, 307.

16 Felix P. Kraveski, "Transition of an Image System: From Paper to Microfiche to Optical Disk," Congressional Research Service, The Library of Congress, June 1990. Frank L. Walker and George R. Thoma, "Access Techniques for document Image Databases," *Library Trends*, 38:4 (Spring 1990): 751-786. Frank L. Walker, "Issues in Document Conversion," in Donald L. Blamberg, Carol L. Dowling and Claudia V. Weston, eds., *Proceedings of the Conference on Application of Scanning Methodologies in Libraries*, (Beltsville, Maryland : National Agricultural Library, 1989), pp.45-60.

17 Para una revisión útil de las normas pertinentes, ver Gerry Walter, "Standards

Help Advance Document Management System Progress," *Optical Memory News*, (May, 1990): 25-27.

18 La presentación de un índice por estructura del documento, en forma de patrón relacional, eventualmente ayudará a facilitar la creación de conexiones entre el documento en imágenes digitales y cualquier contenido informativo almacenado en otras fuentes de lectura automática, tales como la descripción bibliográfica en línea del documento.

19 Lesk, *op. cit.*, 306.

Impreso en julio de 1998
por Editorial **EX-LIBRIS**
Caracas-Venezuela