

Digitalización de colecciones
Texto e imagen. Volumen 1

Digitalización de colecciones
Texto e imagen. Volumen 1

Autores:

Gonzalo Lara Pacheco
Alberto Castro Thompson
Clara López Guzmán
Guillermo Chávez Sánchez
Dante Ortiz Ancona



Universidad Nacional Autónoma de México

México 2008

Coautores:

Gonzalo Lara Pacheco

Dirección General de Bibliotecas, Secretaría Técnica de Biblioteca Digital.

Alberto Castro Thompson

Dirección General de Bibliotecas, Secretaría Técnica de Biblioteca Digital.

Clara López Guzmán

Coordinación de Universidad Abierta y Educación a Distancia, Centro de Alta Tecnología en Educación a Distancia.

Guillermo Chávez Sánchez

Dirección General de Servicios de Cómputo Académico, Coordinador de Publicaciones Digitales

Dante Ortiz Ancona

Dirección General de Bibliotecas, Secretaría Técnica de Biblioteca Digital.

Presentación:

Silvia González Marín

Instituto de Investigaciones Bibliográficas

Corrección de estilo:

Adrián Estrada Corona

Dirección General de Servicios de Cómputo Académico, Coordinación de Publicaciones Digitales

Editán:

Dirección General de Bibliotecas, UNAM

Dirección General de Servicios de Cómputo Académico, UNAM

Coordinación de Universidad Abierta y Educación a Distancia, UNAM

Lara, G., Castro, A., López, C., Chávez, G., Ortiz, D., (2008). Digitalización de colecciones: Texto e imagen, Vol. 1. UNAM, Dirección General de Bibliotecas, Dirección General de Servicios de Cómputo Académico, Coordinación de Universidad Abierta y Educación a Distancia. México. [En línea] <<http://www.digitalizacion.unam.mx/digitalizacion.pdf>> [Consultado: dd/mm/aaaa].



Esta obra se encuentra bajo Licencia Creative Commons.

Eres libre de: copiar, distribuir y comunicar públicamente la obra, así como de hacer obras derivadas bajo las condiciones siguientes:

Atribución. Debes reconocer la autoría de la obra en los términos especificados por el propio autor o licenciante.

No comercial. No puedes utilizar esta obra para fines comerciales.

Licenciamiento Recíproco. Si alteras, transformas o creas una obra a partir de esta obra, solo podrás distribuir la obra resultante bajo una licencia igual a ésta.

Digitalización de colecciones [recurso electrónico] / autores, Lara, G., Castro, A., López, C., Chávez, G., Ortiz, D., -- México : UNAM, Dirección General de Bibliotecas : UNAM, Dirección General de Servicios de Cómputo Académico : UNAM, Coordinación de Universidad Abierta y Educación a Distancia, 2008.

v. : il.

Forma de acceso: World Wide Web

Disponible en: <http://www.digitalizacion.unam.mx>

Título tomado de la pantalla de presentación (vista el 28 de mayo, 2008)

Contenido: Texto e imagen. Volumen 1

Incluye referencias bibliográficas

ISBN: 978-970-32-5215-2 (obra general) – ISBN: 978-970-32-5216-9 (versión HTML) – ISBN: 978-970-32-5217-6 (versión PDF).

1. Materiales de biblioteca - Digitalización - Manuales, etc.
2. Preservación digital - Manuales, etc.
3. Procesamiento de imágenes - Técnicas digitales - Manuales, etc.
4. Procesamiento de textos (Computación) - Manuales, etc. I. Lara Pacheco, Gonzalo, colab.

CONTENIDO

Presentación..... V

Prólogo VII

1. Introducción..... 1

1.1. Objetivos 2

1.2. Alcances..... 2

1.3. Fuentes consultadas..... 4

2. Planeación (establecimiento del proyecto y plan de acción) 5

2.1. Definición de objetivos 5

2.2. Definición de metas 6

2.3. Definición de alcances 6

2.4. Definición de políticas 7

2.5. Formulación del proyecto 7

2.6. Selección de la colección..... 8

2.7. Recomendaciones 9

2.8. Fuentes consultadas..... 9

3. Formatos de archivos digitales 11

3.1. Recomendaciones: 14

3.2. Fuentes consultadas..... 14

4. Infraestructura tecnológica 15

4.1. Recomendaciones generales..... 15

4.2. Composición de una unidad de digitalización..... 16

4.3. Computadoras personales o estaciones de trabajo (CPU) 17

4.4. Monitores 20

4.5. Escáneres..... 22

4.5.1. Principales características técnicas de un escáner 24

4.5.2. Tipos de escáneres 26

4.6. Software 33

4.6.1. Gestores documentales..... 34

4.7. Servidores y unidades de almacenamiento..... 35

4.7.1. Servidores..... 35

4.7.2. Medios de almacenamiento..... 37

4.8. Recomendaciones 44

4.9. Fuentes consultadas..... 44

5. Preparación de los documentos 47

Digitalización de Colecciones. Texto e imagen. Volumen 1

5.1	Separación de documentos por su tipo	47
5.2	Orden de los documentos.....	48
5.3	Preparación física.....	48
5.4	Calidad de los documentos originales.....	50
5.5	Registro y control	50
5.6	Recomendaciones	51
5.7	Fuentes consultadas.....	51
6.	Proceso de digitalización.....	53
6.1.	La digitalización	53
6.2.	El archivo master	54
6.3.	Pasos del proceso de digitalización	55
6.4.	Digitalización de imágenes.....	56
6.4.1.	Profundidad de color.....	57
6.4.2.	Resolución	60
6.5.	Digitalización de textos	62
6.6.	Reconocimiento Óptico de Caracteres (OCR)	63
6.7.	Almacenamiento de los archivos	64
6.7.1.	Formato del archivo	65
6.7.2.	Nombres de archivos	66
6.8.	Recomendaciones	68
6.9.	Fuentes consultadas.....	68
7.	Respaldos y preservación digital.....	69
7.1.	Conceptos y definiciones	69
7.1.1.	Definición de respaldo.....	69
7.1.2.	Definición de preservación digital.....	69
7.1.3.	Diferencias entre preservación y respaldo	71
7.2.	Algunas problemáticas de la preservación.....	72
7.3.	Respaldo y recuperación.....	72
7.3.1.	Respaldo tradicional.....	73
7.3.2.	Respaldo con tecnología RAID.....	74
7.4.	Estrategias para la preservación digital	76
7.5.	Modelo de referencia OAIS	82
7.5.1.	Entidades de OAIS	82
7.6.	Esquemas y estándares de metadatos	87
7.7.	Arquitectura tecnológica	87
7.8.	Políticas y procedimientos	90
7.8.1.	Políticas y procedimientos de respaldo y preservación	90

7.8.2.	Políticas y procedimientos de respaldo.....	91
7.8.3.	Políticas y procedimientos de preservación.....	91
7.9.	Recomendaciones	92
7.10.	Para saber más... ..	92
7.11.	Fuentes consultadas.....	93
8.	Recomendaciones y puntos críticos.....	95
8.1.	Instalaciones.....	95
8.2.	Outsourcing.....	96
8.3.	Derechos de autor	97
8.4.	Colecciones en línea.....	98
9.	Glosario	101
10.	Fuentes consultadas.....	109

Presentación

En mayo de 2001 la Dirección General de Bibliotecas de la Universidad Nacional Autónoma de México comenzó la construcción de la Biblioteca Digital , BIDI-UNAM, misma que fue puesta a disposición del público en octubre de 2003. Tiene como propósito organizar, difundir y conservar la información científica en formato electrónico, en apoyo a la enseñanza y a la investigación que se realizan en esta Universidad.

El espectacular desarrollo que han experimentado en las últimas décadas las tecnologías de la información, ha hecho posible que la humanidad pueda transitar hacia la sociedad del conocimiento. En este paso, la información en formato digital ha jugado un importante papel al haber conquistado las preferencias de los académicos por las facilidades que ofrece de acceso a un vasto cúmulo de información científica y humanística, disponible para su consulta desde cualquier sitio y sin restricciones de horario.

Sin embargo, todavía falta un buen trecho de camino por recorrer para alcanzar la puerta de entrada a la sociedad del conocimiento, sobre todo en los países en vías de desarrollo que necesitan destinar recursos económicos crecientes para financiar la investigación científica generadora de nuevos conocimientos, decisivos para impulsar el avance tecnológico, el crecimiento económico, el desarrollo social y el mejoramiento de las condiciones de vida de sus pueblos.

También implica destinar recursos para la innovación científico-técnica y la conservación de la información. Este aspecto es de suma importancia, pues debe preservarse para las generaciones futuras todo ese conocimiento generado por la humanidad y que puede ponerse en peligro ante el rápido avance de las tecnologías y la continua obsolescencia de los soportes de almacenamiento.

La Biblioteca Digital de la Universidad tiene como función sustantiva brindar a los usuarios un servicio eficiente, de calidad, con rápido acceso a la información seleccionada y en el menor tiempo posible, lo cual significa contar con un grupo de académicos de alta profesionalización provenientes de diversas disciplinas del

Digitalización de Colecciones. Texto e imagen. Volumen 1

conocimiento dedicados a la investigación sistemática de las continuas innovaciones tecnológicas y de la actualización constante en los nuevos avances de las tecnologías de la información. A estos requerimientos se deben agregar los nuevos, producto del crecimiento acelerado de las colecciones digitales que ha exigido el replanteamiento de los modos de producir, organizar, almacenar, preservar, difundir y adquirir la información asentada en formatos digitales.

El creciente aumento que han alcanzado los proyectos de digitalización en diversas instituciones de educación superior, en las bibliotecas, en los centros de información y en las oficinas gubernamentales, plantea un problema serio para los informáticos y los bibliotecarios, sobre todo si se toma en cuenta que en pocos de esos lugares se cuenta con prácticas y procedimientos estandarizados a nivel internacional, que mejorarían en forma importante la calidad de las colecciones digitales.

Este problema al que se enfrenta también la Universidad, hizo que un grupo de académicos expertos en bibliotecas digitales, algunos de ellos con responsabilidades en el funcionamiento de la BIDI-UNAM, se dieran a la tarea de elaborar una obra que contribuyera a establecer las pautas y los procedimientos más recomendables para trasladar textos e imágenes del formato impreso al digital. El objetivo que persigue es contribuir a crear un consenso entre los especialistas en cuanto a los procedimientos de digitalización, con el propósito, afirman los autores “de facilitar la producción, la administración, el mantenimiento y la preservación de las colecciones digitales.”

La comunidad universitaria se congratula con los autores por la iniciativa que tuvieron al elaborar este libro que, seguramente, contribuirá a establecer los consensos necesarios para que la información en formato digital pueda contar con las pautas y procedimientos que ayuden a diversas instituciones a facilitar la producción, la administración y el mantenimiento de sus colecciones digitales, lo que repercutiría en un importante ahorro de tiempo y de recursos económicos.

Dra. Silvia González Marín

Prólogo

La conversión de imágenes y textos impresos a formatos digitales es una actividad en la que necesariamente participan técnicas, recomendaciones, estándares normalizados y criterios personales de quien digitaliza. Coordinarlos y aplicarlos equilibradamente a la digitalización de una colección de documentos no siempre es fácil.

Este manual ha sido creado por expertos de diversas disciplinas con experiencia en proyectos de digitalización. La temática abordada equilibra los aspectos técnicos con la organización y planeación que deben tenerse en cuenta en un proyecto de digitalización de colecciones. Así mismo se ha procurado claridad y sencillez en la terminología a fin de facilitar la lectura por parte de un público amplio, pues no es extraño que haya personal proveniente de diferentes áreas y profesiones dedicado a procesamientos digitales.

Confiamos en que este manual promoverá el incremento de iniciativas de digitalización de colecciones, y al mismo tiempo fortalecerá la participación de los diversos actores involucrados en la creación, difusión y administración de fuentes de información y conocimiento, en las actividades necesarias para construir colecciones en formatos digitales.

1. Introducción

Las necesidades de información que han creado los avances tecnológicos de los últimos años en diversas áreas del conocimiento han propiciado el desarrollo de colecciones digitales y el replanteamiento de los modos de producir, organizar, almacenar, preservar, difundir y adquirir la información asentada en formatos digitales.

Los proyectos de digitalización son cada vez más comunes en bibliotecas, instituciones, unidades de información, organizaciones y centros de todo tipo. Sin embargo, en muchos de ellos se pasan por alto prácticas y procedimientos estandarizados a nivel internacional, con los cuales se puede mejorar la calidad de las colecciones digitales¹.

Tomando como punto de partida esta situación y la variedad de procesos, parámetros y puntos de vista alrededor de la digitalización de una colección de documentos, en este manual se muestran algunas de las pautas y procedimientos más recomendables para trasladar textos e imágenes del formato impreso al digital. En la medida de lo posible se intenta establecer un consenso en las técnicas y pasos a seguir en la digitalización de documentos con el propósito de facilitar la producción, administración y mantenimiento de las colecciones digitales.

Generar colecciones digitales a partir de formatos impresos aplicando procedimientos y parámetros normalizados, significa contar con una amplia gama de ventajas para su producción, administración y difusión, principalmente cuando tales colecciones se destinan a la difusión remota por medios como Internet.

Algunas ventajas de usar estándares:

- Poner a disposición las colecciones sin necesidad de estar físicamente en un lugar determinado.
- Almacenar, preservar, difundir y actualizar las colecciones con mayor facilidad.

¹ Beagrie, N., Jones, M., Digital Preservation Coalition. *Preservation Management of Digital Materials: A Handbook*. En: <http://www.dpconline.org/graphics/handbook/> [Consultado en diciembre 2007]

Digitalización de Colecciones. Texto e imagen. Volumen 1

- Facilitar el intercambio y la compatibilidad de la información.
- Favorecer la migración de los materiales a formatos y tecnologías que sean desarrollados en el futuro.
- Prolongar la preservación y el uso de los materiales digitales.
- Contar con un producto confiable para generar diversas copias y versiones.
- Incrementar la fidelidad de las copias derivadas del original.
- Mantener un documento digital trasladable a diferentes formatos y soportes (cintas, discos, etcétera).
- Motivar el interés y el apoyo de las instituciones para el desarrollo de proyectos de digitalización de amplio alcance y explotación.²

1.1. Objetivos

- Ofrecer los principales pasos y procedimientos para involucrarse en un proyecto de digitalización de colecciones.
- Proporcionar una guía de referencia para planear, construir y administrar colecciones digitales de textos e imágenes.

1.2. Alcances

Con este manual se ofrece una guía para facilitar la digitalización de colecciones de textos e imágenes. Se ha pensado particularmente en instituciones académicas y organizaciones que administran información. Sin embargo, los procedimientos aquí expuestos se pueden aplicar también a colecciones de textos e imágenes diversas.

Los parámetros que aquí se presentan tienen la intención de ser punto de referencia, pues es difícil establecer parámetros fijos e invariables para todo tipo de documentos y para todos los proyectos de digitalización.³ De hacerlo así, se dejarían de lado las particularidades de cada colección, forzando o limitando las

² Greenstein, Daniel y Gerald, George (2008). *Digital reproduction quality: benchmark recommendations*. Disponible en: <http://www.rlg.org/preserv/diginews/diginews5-4.html#featured>. [Consultado en enero 2008].

³ Digital Library Federation. *Digital library standards and practices*. Disponible en: <http://www.diglib.org/standards.htm>. [Consultado en enero 2008].

capacidades de la infraestructura tecnológica y humana para desarrollar un proyecto de digitalización. En este sentido, se recomienda considerar esta guía en el marco de las necesidades, capacidades y posibilidades tecnológicas y humanas de cada institución y proyecto de digitalización en que se decida aplicar.

Los procedimientos, prácticas, parámetros, etcétera, que se presentan están respaldados por las recomendaciones de instituciones de reconocido prestigio pioneras en la digitalización de colecciones académicas, como la Biblioteca del Congreso de Estados Unidos (LC, por sus siglas en inglés), la Federación de Bibliotecas Digitales (DLF, por sus siglas en inglés), algunos participantes de la Iniciativa de Bibliotecas Digitales (DLI, por sus siglas en inglés), la Universidad Cornell de Michigan y las bibliotecas nacionales de Australia y Canadá, entre otras. Al mismo tiempo, hemos vertido recomendaciones producto de las experiencias que, con los mejores resultados, ha arrojado la práctica cotidiana.

Hay que tener en cuenta que, por sí misma, la tecnología digital propicia de manera constante la revisión y la actualización de las recomendaciones acerca de procedimientos, características de equipo, configuración, calidad, resolución y prácticamente todos los factores tecnológicos que intervienen.

Finalmente, es importante destacar que para la digitalización de una colección es necesario seguir un plan estratégico que incluya la planeación, operación y mantenimiento del proyecto de digitalización, como se muestra en la siguiente figura (figura 1).

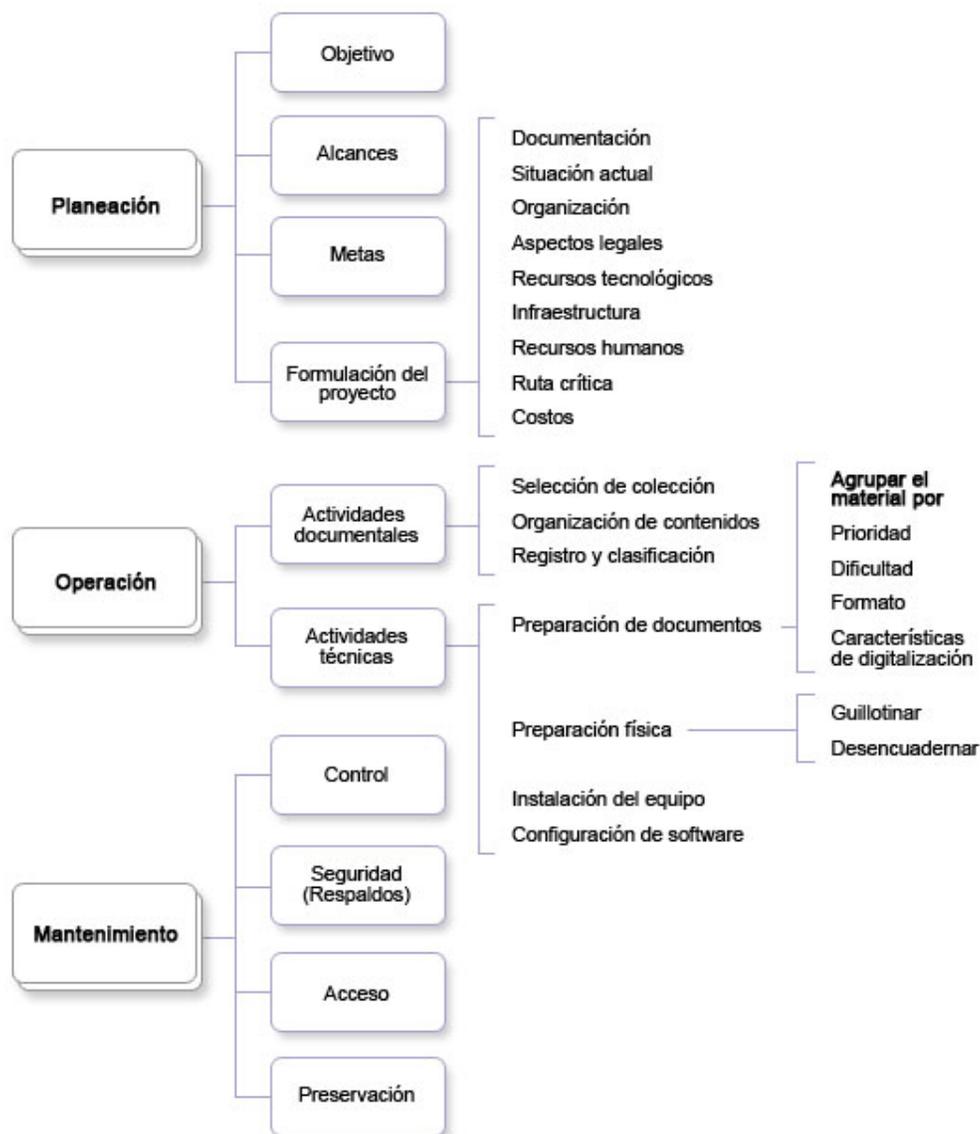


FIGURA 1. Mapa de un proyecto de digitalización de colecciones

1.3. Fuentes consultadas

- Beagrie, N., Jones, M., Digital Preservation Coalition. *Preservation Management of Digital Materials: A Handbook*. En: <http://www.dpconline.org/graphics/handbook/> [Consultado en diciembre 2007].
- Digital Library Federation. *Digital library standards and practices*. Disponible en: <http://www.diglib.org/standards.htm>. [Consultado en enero 2008].
- Greenstein, Daniel y Gerald, George (2008). *Digital reproduction quality: benchmark recommendations*. Disponible en: <http://www.rlg.org/preserv/diginews/diginews5-4.html#featured>. [Consultado en enero 2008].

2. Planeación (establecimiento del proyecto y plan de acción)

En la etapa de planeación es indispensable definir los pasos que se van a seguir a lo largo del proyecto de digitalización. El éxito o el fracaso de un proyecto dependen de una adecuada planeación, pues se trata de la guía que establece los aspectos técnicos y operativos que se llevarán a cabo. Debe estar previsto cada detalle, plenamente establecidas las responsabilidades y previstas posibles soluciones a contingencias.

2.1. Definición de objetivos

En los objetivos para digitalizar una colección debe plantearse precisamente qué se espera cubrir o satisfacer. Deben estar formalmente establecidos y tienen que ser dados a conocer entre los participantes del proyecto.

A continuación se muestran algunos de los objetivos más comunes para un proyecto de digitalización. No todos son aplicables al mismo proyecto.

- Brindar acceso a la colección en formato digital de manera local o remota.
- Trasladar documentos a formato digital con la intención de preservar los documentos físicos.
- Apoyar la educación, la investigación, las actividades empresariales e industriales, etcétera, de los usuarios, a partir de la consulta de una colección en formato digital.
- Modernizar el acceso a las fuentes de información.
- Compartir la colección, documentos, etc., en formato digital con las instituciones que han firmado convenios de colaboración, etcétera.

Debe plantearse si la digitalización se hará para:

- Preservación.
- Consulta.
- Difusión en línea o local.
- Difusión en soporte óptico.

Digitalización de Colecciones. Texto e imagen. Volumen 1

- Acceso a todo público.
- Acceso a una comunidad específica.
- Acceso total, parcial o restringido.
- Etc.

2.2. Definición de metas

Las metas orientan los objetivos acordados para digitalizar una colección. Si dentro de los objetivos se plantea poner a disposición de una comunidad de usuarios una colección en particular, la meta sería lograrlo en un tiempo y en unas condiciones determinados, de manera progresiva o total.

Conviene establecer metas para:

- Tiempos.
- Recursos económicos.
- Personal.
- Infraestructura tecnológica.
- Alcance (regional, temático, temporal, etcétera).

2.3. Definición de alcances

Otro paso importante en el desarrollo de un proyecto es definir hasta dónde se planea llegar, es decir:

- Qué volumen de documentos se digitalizará.
- Durante qué periodo se realizará.
- En cuánto tiempo se hará.
- Cuántos y qué tipo de usuarios atenderá.
- Quiénes tendrán acceso a la colección.
- Bajo qué condiciones, modalidades, etcétera, se dará el acceso.
- Etc.

Debe establecerse qué se espera de la colección en función de los objetivos y las metas trazados. El servicio y satisfacción de necesidades de información deberán ser prioritarios en la definición de los alcances.

2.4. Definición de políticas

Las políticas deben estar presentes en todo proyecto, fungiendo como las reglas del juego y enfocadas a determinar qué se puede y qué no se puede hacer, además de decidir de qué manera.

En la formulación de las políticas deberán destacarse los derechos de autor como un aspecto relevante en el mundo digital. Actualmente es uno de los puntos más críticos en la libre circulación de la información por los medios digitales. Si no se cuenta con los derechos para difundir, reproducir, etcétera, algún documento o colección, es mejor replantear la conveniencia o no de seguir adelante con el proyecto. Ante tal circunstancia hay que evitar una fuerte inversión de dinero y esfuerzo, pues es muy probable que las restricciones legales impidan la difusión-consulta de la colección digital.

Entre otros aspectos, las políticas deben regular procedimientos relacionados con:

- Las dinámicas de trabajo para el procesamiento digital íntegro, incluyendo mantenimiento, preservación, etc.
- Los costos, manejo y tratamiento del equipo, así como de documentos, personal y otros recursos que serán utilizados.
- El acceso a la colección, restricciones, condiciones de consulta, etc.
- La participación en cuanto a responsabilidades, compromisos y obligaciones de instituciones, autoridades o entidades involucradas.
- Cualidades y capacitación del personal que intervendrá en las distintas etapas del proyecto.

2.5. Formulación del proyecto

La formulación del proyecto debe incluir los puntos expresados anteriormente: podría decirse que son su carta de presentación. De la formulación depende que

Digitalización de Colecciones. Texto e imagen. Volumen 1

las autoridades y los responsables de instituciones u organizaciones acepten un proyecto de conversión digital de textos e imágenes.

Dentro de la formulación del proyecto deben integrarse objetivos, metas, alcances y políticas relacionados con tres aspectos principales:

- planeación
- operación y
- mantenimiento.

Será necesario subrayar:

- Costos aproximados para el proceso de digitalización y la etapa posterior de mantenimiento y preservación de la colección.
- Infraestructura tecnológica y de recursos humanos para antes, durante y después del proceso de digitalización.
- Beneficios directos e indirectos a corto, mediano y largo plazos, traducidos en ahorros de espacio, mantenimiento, personal, etc., en contraste con la colección física.
- Cualquier dato, circunstancia, etc., que se considere como determinante en la decisión para realizar el proyecto de la mejor manera, en sus etapas preparatorias, operativas y de mantenimiento (resguardo, preservación, etcétera).

2.6. Selección de la colección

Seleccionar una colección es un paso previo indispensable en el proceso de digitalización de la misma. Involucra estudios, sondeos y evaluaciones entre la comunidad de usuarios y los responsables del proyecto para definir específicamente qué colección y por qué será digitalizada.

Deben definirse claramente las necesidades de información que se busca satisfacer con la colección digital. Por ejemplo, en el caso de instituciones educativas deben priorizarse necesidades reales como documentos de alta demanda, pertenecientes a planes y programas de estudio, bibliografías básicas,

materiales producidos por los propios docentes, editados por la institución, etc. De manera complementaria, es recomendable digitalizar los documentos que presentan mayores facilidades para ser procesados digitalmente. En algunas colecciones se pueden considerar aquellos que:

- Sean los más demandados por parte de los usuarios.
- Estén libres de controversias de derechos de autor.
- Sean físicamente frágiles, estén dañados y cuya manipulación ponga en riesgo su integridad física.
- Se quiera conservar a largo plazo.
- Se desee difundir masivamente entre la comunidad de usuarios, por motivos académicos, estéticos, históricos, comerciales o de algún otro tipo.
- Sean únicos, de gran valor o similares.

2.7. Recomendaciones

- Fundamentar con argumentos sólidos y juicio crítico qué documentos, colección, etc., se planea digitalizar y para qué.
- Trabajar en coordinación y colaboración con las personas, autoridades y dependencias involucradas en el proyecto de digitalización.
- Visualizar claramente las necesidades, posibles contratiempos y beneficios que arrojará el proyecto.

Para abundar más conviene revisar *Planning digitization projects: a brief bibliography* de la Universidad de Harvard. Está disponible en: preserve.harvard.edu/bibliographies/digitalplanning.pdf.

2.8. Fuentes consultadas

- Hughes, Lorna M. (2004). *Digitizing collections: strategic issues for the information manager*. London: Facet.
- Hurst, Jill Ann (2002). *Planning for regional digitization projects : the SCRLC approach*. South Central Regional Library Council (september 2002). Disponible en: http://www.lakenet.org/scrlc/ad_comm/acits/2002digplan.pdf [Consultado en marzo 2007].
- Smith, Abby (1999). *Why Digitize?* Council on Library and Information Resources.

Digitalización de Colecciones. Texto e imagen. Volumen 1

Disponible en: <http://www.clir.org/pubs/reports/pub80-smith/pub80.html> [Consultado en mayo 2007].

- Smith, Abby (2001). *Strategies for Building Digitized Collections*. Washington, DC: Council on Library and Information Resources. Disponible en: <http://www.clir.org/pubs/reports/pub101/contents.html>. [Consultado en marzo 2007].

3. Formatos de archivos digitales

El o los formatos en que se digitaliza un documento o colección, determinan la manera como será administrada y consultada. La utilización de formatos regidos bajo normas y estándares ampliamente difundidos incrementa la certeza de que una colección estará disponible no sólo inmediatamente, sino también en el largo plazo, al margen de los cambios tecnológicos. Utilizar formatos estandarizados de amplio uso facilita la consulta de los documentos por parte de los usuarios de la colección. Para consultar un documento en un formato estándar ampliamente utilizado como .jpg o .gif, por ejemplo, no hay necesidad de instalar programas adicionales para la lectura / visualización. Si se presenta la necesidad de contar con un software para leer/ver una imagen, siempre debe procurarse que se trate de software de fácil acceso y preferentemente gratuito (por ejemplo, .pdf para documentos de texto e imagen).

El formato es la manera como se codifica la información para almacenarla en un archivo de computadora. Los formatos que se utilizarán deben ser elegidos en función de las características de los documentos por digitalizar y las necesidades de información que se espera cubrir con la colección.

En su elección habrá que considerar:

- Las características de los materiales.
- La infraestructura tecnológica disponible.
- La opción de contar con dos versiones distintas (una *master* y otra de consulta).
- Si además de digitalizar una colección se planea preservarla, deberán considerarse las características específicas del o los formatos que se utilizarán.
- Si sólo se planea la consulta local y no a través de Internet.

Dentro de las cualidades esenciales de un formato se encuentran la fidelidad que guarda con la fuente original, la manipulación y la duración. Debido a que hay formatos que son mejores para unas funciones que para otras (impresión,

Digitalización de Colecciones. Texto e imagen. Volumen 1

disponibilidad en Internet, respaldo, etc.), es necesario elegir el mejor para el tipo de documento que se vaya a digitalizar y el uso que se hará del mismo (véase cuadro 1).

Formato	Tipo de material	Características	Extensión
JPG (Joint Photographic Experts Group)	Imagen (Incluye texto-imagen)	<ul style="list-style-type: none">• Ampliamente utilizado para fotografías e imágenes distribuidas por Internet• Sin compresión produce imágenes de buena calidad y tamaño reducido• Amplia capacidad de reproducción y despliegue de colores.	.jpg
PDF (Portable Document Format: Adobe Acrobat File)	Texto e imagen	<ul style="list-style-type: none">• Reproducción casi exacta de la fuente original• Amplia difusión• Estándar en Internet para distribución de textos.	.pdf
TIFF	Texto e imagen	<ul style="list-style-type: none">• Recomendado para texto simple• Tamaño reducido• Muy utilizado para archivos <i>master</i>• Sin propietario	.tif
GIF (Graphics Interchange file Format)	Imagen	<ul style="list-style-type: none">• Ampliamente utilizado en Internet• Imágenes de mapa de bits• Tamaño reducido.	.gif
BMP (Bit Map)	Imagen	<ul style="list-style-type: none">• Imágenes de mapa de bits• Estándar para Windows y OS/2.	.bmp
TXT (Texto simple)	Texto	<ul style="list-style-type: none">• Amplio uso• Sin propietario• Estándar para múltiples sistemas operativos.	.txt
DOC (Documento de Windows Office)	Texto	<ul style="list-style-type: none">• Uno de los formatos más populares para la creación de textos electrónicos.	.doc

CUADRO 1. *Formatos estándar más utilizados en la digitalización de textos e imágenes.*

En la selección del o los formatos también habrá que tomar en cuenta:⁴

- Disposición de uso libre y abierto (sin propietario).
- Capacidad de alta resolución y compatibilidad.

⁴ Technical Advisory Service for images. *Choosing a File Format*. Disponible en: <http://www.tasi.ac.uk/advice/creating/format.html#fo5> [Consultado en marzo de 2007]; Technical Advisory Service for Images (TASI). *File formats and compression*. Disponible en: <http://www.tasi.ac.uk/advice/creating/fileformat.html> [Consultado en marzo 2007]; Technical Advisory Service for Images (2008) *New digital image file formats*. Disponible en: <http://www.tasi.ac.uk/advice/creating/newfile.html> [Consultado en marzo 2007].

- Nivel y tipo de compresión con que trabaja.
- Captura de metadatos como parte de la estructura del archivo
- Migración y compatibilidad.

Para seleccionar el formato más adecuado conviene plantearse algunas de las siguientes preguntas:⁵

- ¿Cuál es el uso que se dará al documento una vez puesto al alcance del público?
- ¿Qué tanta resolución requiere el documento para satisfacer las necesidades del usuario?
- ¿Se consultará en monitor o impreso?
- ¿Cuál es la capacidad del dispositivo de consulta y el rango de profundidad de bits que maneja?
- ¿Qué resolución permite?
- ¿Por qué ancho de banda viajará el documento?
- ¿Cuánta fidelidad se espera de la imagen con respecto al original?

En relación con el tipo de acceso, pueden ponerse a consideración las siguientes interrogantes:

- ¿Cómo se hará llegar al usuario? ¿Por medio de CD, cinta, OPAC, Intranet, Internet (tipo de conexión) o algún otro?
- ¿Es necesaria una marca de agua o tratar algún aspecto particular concerniente a los derechos de autor, distribución, copia, etcétera?
- ¿Los usuarios necesitan que el documento presente características particulares en cuanto a color, visualización, etc.?
- ¿Los archivos tendrán restricciones de acceso (inhibición de impresión, descarga, copia, modificación, etc.)?

⁵ Technical Advisory Service for images. *Choosing a File Format*. Disponible en: <http://www.tasi.ac.uk/advice/creating/format.html#fo5> [Consultado en marzo de 2007].

3.1. Recomendaciones:

Con la finalidad de hacer la elección más adecuada conviene seguir algunas recomendaciones:

- Utilizar formatos con normas y estándares internacionales
- Evitar formatos propietarios o condicionados que demanden constantes y costosas actualizaciones, que limiten las funciones del programa, etc.
- Definir formatos para diferentes funciones, como la consulta en línea y la preservación.
- Definir el o los formatos en función de la capacidad de acceso y recuperación que tenga el grueso de la comunidad de usuarios reales y potenciales.

3.2. Fuentes consultadas

- *Almost Every file format in the world!*. Disponible en: <http://www.ace.net.nz/tech/TechFileFormat.html> [Consultado en marzo 2007].
- File Formats Blog. Disponible en: <http://fileformats.blogspot.com/> [Consultado en marzo 2007].
- Fleischhauer, Carl. *Digital formats for content reproductions*. The National Digital Library Program. July 1998. Disponible en: <http://lcweb2.loc.gov/ammem/formats.html> [Consultado en marzo 2007].
- Perlman, Eric and Kallen Ian. *Common Internet File Formats*. Disponible en: <http://www.matisse.net/files/formats.html> [Consultado en enero 2007].
- Technical Advisory Service for images. *Choosing a File Format*. Disponible en: <http://www.tasi.ac.uk/advice/creating/format.html#fo5> [Consultado en marzo 2007].
- Technical Advisory Service for Images. *File formats and compression*. Disponible en: <http://www.tasi.ac.uk/advice/creating/fileformat.html> [Consultado en marzo 2007].
- Technical Advisory Service for Images. *New digital image file formats*. Disponible en: <http://www.tasi.ac.uk/advice/creating/newfile.html> [Consultado en marzo 2007].

4. Infraestructura tecnológica

Un factor de gran importancia que merece mucha atención en la digitalización de colecciones es la infraestructura tecnológica. En cualquier proyecto de digitalización se deben ponderar las posibilidades económicas para asumir los gastos de adquisición, mantenimiento y actualización del equipo requerido. Sin embargo, hay que tener presente que la infraestructura más costosa y novedosa no siempre garantiza el éxito de un proyecto. Lo que siempre facilita la generación de colecciones digitales útiles y funcionales es una adecuada planeación, pues ésta permite aprovechar al máximo los recursos tecnológicos, humanos y financieros.

4.1. Recomendaciones generales

Para arrancar el proyecto de digitalización de una colección es recomendable realizar siempre un análisis de los diferentes factores críticos que pueden influir, de manera general, en el éxito del mismo y, en particular, en la determinación de la infraestructura tecnológica requerida; por ejemplo:

- *La naturaleza de los materiales a digitalizar.* Identificar y cuantificar el material y los soportes físicos en los que se encuentra para determinar los periféricos y el software adecuado para su procesamiento.
- *El uso que se hará de la colección digital (modos de acceso).* Determinar el uso final que se les dará a las colecciones digitales, como la preservación y/o difusión. Con base en ello se podrá definir el equipo necesario para el almacenamiento y la publicación de los documentos, así como el ancho de banda requerido en la red de comunicación.
- *El tipo de usuarios que harán uso de la colección digital.* Identificar cuáles son los tipos de usuarios y el modelo de conexión con que accederán, pudiendo ser ésta interna o externa (Intranet o Internet), factor determinante para la adquisición de servidores y el ancho de banda de la red.
- *La infraestructura tecnológica con que se cuenta.* Valorar las características y la capacidad de los equipos disponibles, como computadoras, periféricos,

escáneres, cámaras y medios de almacenamiento y respaldo, a fin de determinar su grado de utilidad en las actividades del proyecto.

- *Las características y la velocidad promedio del equipo de consulta-acceso.* Determinar factores como capacidad, velocidad de procesamiento, memoria, concurrencias y tiempo de respuesta que son requeridos para el acceso y la consulta de las colecciones a través de una red.
- *El presupuesto.* Asegurar los recursos económicos disponibles para la actualización, adquisición y mantenimiento de la infraestructura tecnológica durante la vida de un proyecto.
- *Espacios mínimos adecuados.* Determinar con qué espacios físicos se cuenta y cuáles son los requerimientos para instalar la infraestructura tecnológica que será utilizada en el procesamiento y el almacenamiento de las imágenes.
- *Los recursos humanos con que se cuenta para operar el equipo.* Verificar si el personal cuenta con la capacidad suficiente para el manejo de equipo y software. De no ser así, debe presupuestarse su capacitación.

El resultado de este análisis, sin lugar a dudas, ofrecerá la pauta para determinar la inversión que se requerirá para actualizar y adquirir el equipo que se utilizará en el proyecto de digitalización. El análisis también permitirá definir los programas de cómputo acordes al proceso de digitalización, la edición de imágenes y la gestión de la información digital.

4.2. Composición de una unidad de digitalización

Para realizar los trabajos de digitalización es necesario contar con equipo de cómputo, equipo de captura digital y software. Esto permitirá trasladar las colecciones a formato digital sin importar el soporte físico en que se encuentren originalmente, ya sea papel, acetatos, diapositivas, negativos, etcétera.

La unidad de digitalización está conformada por el conjunto de elementos tecnológicos que se utilizan para generar colecciones en formato digital a partir de colecciones impresas.

Los elementos que conforman una unidad de digitalización no son definitivos y útiles en todos los casos, porque su composición depende principalmente de las características de la colección que se pretende digitalizar. No obstante, una unidad de digitalización debe constar en general de los siguientes elementos:

- Computadora de escritorio.
- Tarjetas, hardware y periféricos.
- Escáner o equipo para captura de texto / imagen digital.
- Software para la captura y edición de texto / imagen.

Para crear una colección digital a partir de fuentes impresas es necesario contar con estos elementos o similares, dependiendo de las necesidades de los documentos que conforman la colección. A continuación se presenta una breve descripción de los componentes de una unidad de digitalización.

4.3. Computadoras personales o estaciones de trabajo (CPU)

La computadora es uno de los principales elementos para la digitalización de cualquier tipo de documento. A través de ella el operador establece la conexión y la comunicación con los escáneres o las cámaras digitales. En primera instancia, las imágenes resultantes del proceso de digitalización deberán residir en los equipos asignados para esta actividad. En ocasiones será necesario hacer reprocesamientos digitales mínimos, como la nivelación de tonos, brillos y contrastes en la misma computadora. Una vez realizados los trabajos de digitalización y reprocesamiento, las imágenes deben ser transferidas para su preservación digital a las unidades de almacenamiento designadas *ex profeso*, como servidores, DVD, cintas, CDs, etcétera (véase figura 2).

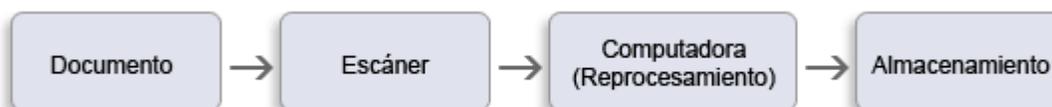


FIGURA 2. Proceso básico de digitalización

Digitalización de Colecciones. Texto e imagen. Volumen 1

Para el desarrollo de las actividades de digitalización es recomendable contar con equipos de alto desempeño y una amplia capacidad de almacenamiento, destinados específicamente para este tipo de tareas. En el mercado existe actualmente una gran variedad de computadoras tradicionales y estaciones de trabajo⁶ con características adecuadas para optimizar el procesamiento de imágenes. Así, cuando se tiene en puerta un proyecto de digitalización a gran escala, es conveniente considerar las siguientes recomendaciones técnicas para seleccionar el equipo:

- Elegir computadoras con una velocidad de procesamiento aceptable, que permitan digitalizar y reprocesar imágenes digitales con la mayor eficiencia posible.⁷
- Adquirir tanta memoria RAM como el presupuesto lo permita, ya que entre mayor sea ésta, mayor será la velocidad para el procesamiento de imágenes digitales.
- La capacidad del disco duro es también un factor importante para el almacenamiento temporal de las imágenes digitalizadas.
- También es recomendable que la computadora cuente con una unidad lectora de CD y DVD, así como la opción de un quemador de CD/DVD con la finalidad de poder leer y guardar imágenes en este tipo de medios. Otra opción es comprar un quemador de CD/DVD externo con el propósito de compartir el dispositivo con otros equipos para ahorrar presupuesto.
- Aunque debe considerarse sólo como una opción, si los recursos lo

⁶ Una estación de trabajo se define como "Micro o minicomputadora para un único usuario, de alto rendimiento, que ha sido especializada para gráficos, diseño asistido por computadora, ingeniería asistida por computadora o aplicaciones científicas". Disponible en: <http://www.monografias.com/trabajos41/estaciones-de-trabajo/estaciones-de-trabajo.shtml> [Consultado en septiembre 2007]. A diferencia de las computadoras tradicionales, debido a su composición arquitectónica, las estaciones de trabajo ofrecen una mayor capacidad de procesamiento, especialmente cuando se requiere mayor rendimiento para la realización de multitareas. Aun con grandes cargas de trabajo, las estaciones de trabajo pueden permanecer utilizables, mientras que las computadoras tradicionales pueden volverse incapaces.

⁷ Hoy en día la mayoría de las computadoras personales en el mercado están conformadas por procesadores cuya velocidad mínima es de 1.6 GHz, superando en algunos casos los 3.7 GHz. La tecnología que se puede encontrar en el mercado es muy variada, puede ir desde procesadores de un solo núcleo, como el Celeron, o de dos y cuatro núcleos como el Core 2 Duo o el Quad Core Intel Xeon, respectivamente. Sin embargo, la selección de la tecnología está directamente relacionada con el presupuesto que se tenga y las necesidades identificadas.

permiten, también es recomendable que la computadora cuente con una tarjeta aceleradora de video. Con esta tarjeta la recepción y la transferencia de información serán mucho más ágiles en el proceso de digitalización.

- De igual manera, el equipo de cómputo deberá tener capacidad para trabajar ágilmente con periféricos conectados en serie, del tipo USB 2.0, IEEE 1394 “Firewire” o mediante tarjeta SCSI (véase figura 3).

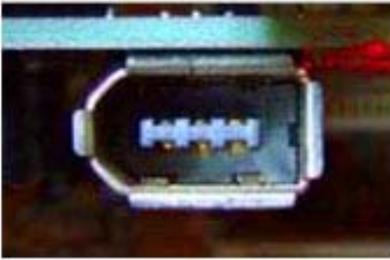
	<p>Conector tipo USB. El USB o Universal Serial Bus (bus universal en serie) es un tipo de conexión plug-and-play (enchufa y listo), que permite conectar dispositivos a la computadora para la transferencia de datos sin reiniciar el equipo. Actualmente, este sistema se ha convertido en el estándar de conexión para impresoras y escáneres. Su velocidad de transferencia es de aproximada 480 MB/seg</p>
	<p>Conector tipo IEEE 1394 “Firewire”. El IEEE 1394, también conocido como FireWire, es un estándar de conexión para la transferencia de datos a velocidades mayores a las manejadas por el USB 2.0. Este tipo de conexión es más utilizada en dispositivos como cámaras digitales o videocámaras.</p>
	<p>Conector tipo SCSI. La conexión SCSI (se pronuncia “escosi”) es un estándar para la transferencia de datos. Está diseñada para los escáneres que cuenten con este tipo de conexión y en los que se quiera agilizar la transferencia de la información. Sin embargo, se requiere de controladores para su uso. Una conexión SCSI puede alcanzar tasas de hasta 100 MB/seg.</p>

FIGURA 3. Tipos de conectores⁸

⁸ Imágenes tomadas de:
http://es.wikipedia.org/wiki/USB#Caracter.C3.ADsticas_de_Transimisi.C3.B3n
http://es.wikipedia.org/wiki/IEEE_1394
<http://www.servipc365.com/informatica/diccionario/s.htm>

La selección de la computadora depende directamente del tipo de proyecto de que se trate. Si la cantidad de documentos de la colección por digitalizar no es exorbitante ni lo son las dimensiones del material, bien podría emplearse un equipo usado con las mínimas características ya señaladas. No obstante, si las dimensiones de la colección son muy grandes, es recomendable la adquisición de un equipo con las mejores características que permita el presupuesto.⁹

Para tomar decisiones acertadas en la selección de un equipo es recomendable que participe el personal que se hará cargo del diseño y la difusión de los contenidos, así como los encargados de los sistemas y la infraestructura. Contar con opiniones diversas de los participantes directamente involucrados en el proyecto también resulta favorable.

Además, también es recomendable consultar en la Web la información que proporcionan los fabricantes sobre sus propios productos, hacer comparativos y, con base en estos datos adicionales decidir con qué equipo quedarán cubiertos los requerimientos planteados.

4.4. Monitores

Para evaluar la calidad de las imágenes digitalizadas y adecuarlas es conveniente un monitor que ofrezca una visualización de imágenes lo suficientemente amplia y clara. En este sentido, una imagen con unas características técnicas específicas puede visualizarse de diferente manera en monitores que utilicen distinta tecnología.¹⁰ Para prever un adecuado procesamiento de las imágenes, se recomienda considerar los siguientes puntos (véase cuadro 2):

- Muchas veces el monitor de una computadora común, de 15 pulgadas, no es suficiente para apreciar la calidad de las imágenes generadas. Lo recomendable es contar con una pantalla cuya resolución mínima sea de

⁹ Hay que tener bien claro que la determinación del tamaño de una colección es variable y relativa. Por lo tanto, se podría considerar como pequeña una colección de 10,000 imágenes, aproximadamente. Por el contrario, podría considerarse como una colección grande unas 100,000 imágenes.

¹⁰ Hay que tener en consideración que otro factor por el que se visualiza una misma imagen de forma diferente en distintos monitores, es la resolución del monitor. Si dos monitores están configurados con diferentes resoluciones, indudablemente esto influirá en la resolución y tamaño de la imagen en pantalla.

1024x768 pixeles y una visualización de al menos 17 pulgadas, pero se consideran más adecuadas las de 19 ó 21 pulgadas. De este modo será más sencillo hacer una evaluación integral de las imágenes y aplicar las técnicas de optimización, sin perder la perspectiva de los resultados.

- Básicamente, en el mercado existen tres tipos de tecnologías para monitores: Tubos de Rayos Catódicos o CRT (Cathode Ray Tube); pantalla de Cristal Líquido o LCD (Liquid Crystal Display) y pantalla de Transistor de Capa Fina o TFT (Thin Film Transistor). Si se cuenta con los recursos económicos suficientes, lo más recomendable será siempre un monitor LCD. Sin embargo, si el presupuesto es bajo o se quiere aprovechar el equipo con que ya se cuenta, los monitores CRT de 17 pulgadas son una solución aceptable.
- Un factor también importante en la selección de los monitores es su resolución. Entre mayor sea ésta, será también mayor su posibilidad de visualizar con un alto grado de fidelidad, lo que es de gran utilidad en el reprocesamiento de imágenes. Aunque los monitores de CRT manejan resoluciones mínimas de 800x600 y máximas de 1024x768 pixeles, las nuevas tecnologías ofrecen resoluciones que van desde los 800x600, superando los 1440x900 pixeles.
- Otro aspecto relevante en la elección de un monitor es el tipo de conexión. En monitores modernos, lo más recomendable es contar con la típica D-SUB de 15 pines (VGA) y la conexión digital DIV. Sin embargo, en monitores anteriores la conexión usual es la D-SUB.

Recomendaciones técnicas para la selección de monitor		
Características	Requerimientos mínimos	
	Equipo nuevo	Equipo usado
Tamaño	19"	17"
Resolución	1024x768 pixeles	1024x768 pixeles
Tecnología	LCD	CRT
Resolución	De 800x600 y superior a 1440x900	De 800x600 a 1024x768
Conexión	D-sub y DIV	D-sub

CUADRO 2. *Recomendaciones para la selección de monitores*

Aunque se adquiriera el mejor equipo, no hay que olvidar que esto no asegura la calidad óptima de las imágenes resultantes, pues la calidad de una imagen depende directamente de la capacidad y la sensibilidad del operador en la aplicación de los procedimientos que garanticen una información digital de alta calidad.

4.5. Escáneres

Además de la computadora, el escáner es otro elemento esencial en las actividades de digitalización. Es un periférico mediante el cual un documento impreso puede ser transformado en una imagen digital. La creciente oferta de este tipo de herramientas en el mercado y su costo relativamente accesible obliga a elegir con cautela el escáner que responda adecuadamente a las necesidades del proyecto por desarrollar. Actualmente existe una gran variedad de modelos con los cuales es posible digitalizar todo tipo de materiales como diapositivas, negativos, impresiones, mapas, planos, libros antiguos, periódicos, microfilms, etcétera. Por el papel que juega en la digitalización, es necesario conocer la base de su funcionamiento y los requerimientos mínimos de un escáner para su óptimo desempeño.

Para transformar imágenes análogas (impresas) en digitales, los escáneres trabajan con un elemento fotosensible llamado CCD (Charge Coupled Device), del que depende la resolución que puedan alcanzar. El funcionamiento de un escáner se inicia con la luminosidad de una lámpara, que envía un haz de luz hacia el objeto a digitalizar. El reflejo de esta luz es captado por el CCD, a través de espejos o lentes. El CCD convierte el reflejo de luz en pulsos electrónicos que pasan a través de un dispositivo llamado DAC (Digital Analog Converter), el cual los transforma en código binario, mismo que la computadora puede interpretar y presentar en una pantalla (véase figura 4).

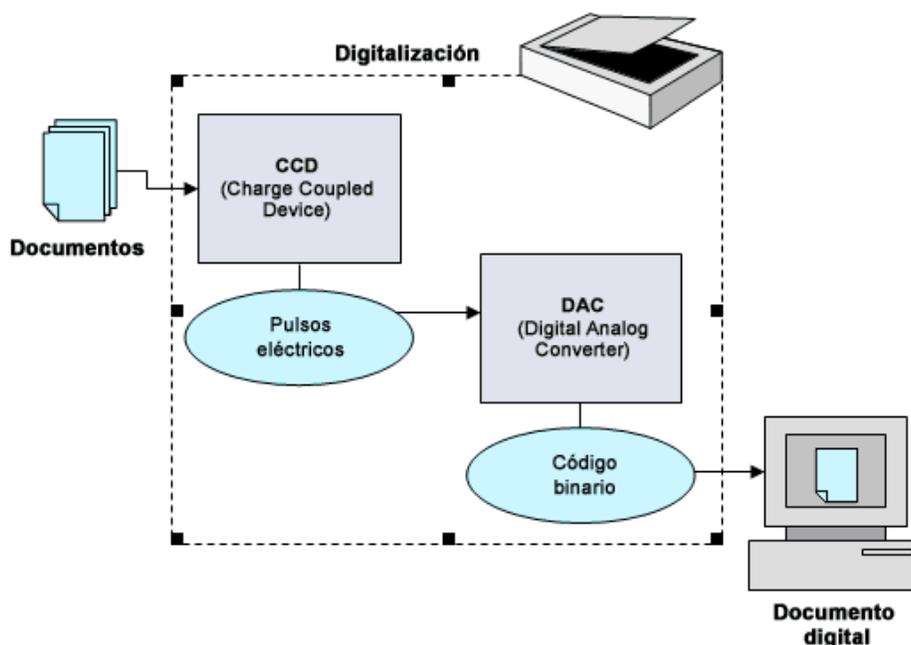


FIGURA 4. Funcionamiento del escáner

Cuando está por iniciarse un nuevo proyecto de digitalización es recomendable conocer los componentes del escáner que participan directamente en la digitalización, así como el software que se requiere para el manejo y el procesamiento de las imágenes resultantes (véase tabla 1).

Componentes	Software	Aplicación
Fuente lumínica Tamaño de la superficie de captura Velocidad de escaneo Ruta de elementos ópticos Mecanismo Sensor Requerimientos de energía Asistencia del distribuidor Rutas electrónicas Auto alimentación de documentos (ADF) Adaptador de transparencias	Número de bits por píxel Procesador de imágenes Productividad Calibración Selección de rangos Gamma Escalabilidad OCR Conversión <i>Raster</i> a vector Retención del formato de la imagen	Manejo de color Compresión Captura de textos Formato de archivos Ajuste de imagen Interacción con el texto Manejo de formatos traslúcidos

TABLA 1. Componentes, Software y aplicaciones de un escáner¹¹

¹¹ Adaptada de Williams, Don (2000). *Selecting a Scanner*. En: Guides to Quality in Visual Resource Imaging Digital Library Federation; Research Libraries Group. Disponible en: <http://www.rlg.org/visguides/visguide2.html> [Consultado en septiembre 2007]

4.5.1. Principales características técnicas de un escáner

Al seleccionar un escáner no basta con saber cuáles son los elementos que lo componen. Se requiere un conocimiento mínimo acerca de sus principales características técnicas, mismas que son determinantes en el grado de utilidad y la eficiencia del equipo al digitalizar una colección. Algunas características técnicas importantes:

- **Resolución óptica**

La resolución óptica de un escáner está determinada por el número de sensores que tiene para capturar una imagen. Regularmente el escáner expresa la resolución óptica en ppp (puntos por pulgada) o dpi (*dots per inch*, que equivale a lo mismo que las siglas anteriores pero en inglés). Es recomendable que la resolución óptica de un equipo rebase las necesidades mínimas del material que será escaneado. Por ejemplo, un escáner de cama plana cuya resolución es de 1200 dpi puede procesar una imagen de 20x25 cm a 600 dpi, pero no lo podrá hacer con una transparencia de 2x2 a 2000 dpi.

Con fines publicitarios, a muchos escáneres se les atribuye una resolución muy alta, cuando en realidad se hace referencia a su resolución *interpolada*. Esta resolución se caracteriza porque, al momento de escanear, "adivina" los colores y la luz que se encuentran en los espacios que no pueden ser captados por la lámpara de un escáner. Por lo tanto, al momento de seleccionar un escáner se sugiere tomar como punto de partida su resolución óptica y no la interpolada.

Es necesario poner especial atención en el grupo de datos que el escáner es capaz de manejar en el barrido, es decir, en el momento en el que las lámparas recorren la superficie del documento que está capturando. A menudo se señalan 1600x3200, correspondiendo el primer número a la resolución óptica, mientras que el segundo se refiere a la capacidad de obtención de datos en el área de captura del escáner al momento del barrido. En este sentido, en las especificaciones del equipo debe observarse que el segundo número siempre sea mayor que el primero, pues de lo contrario las imágenes tendrán una resolución interpolada o irreal.

- **Profundidad de bits (colores y tonos)**

La cantidad de información que el grupo de sensores de un escáner puede recoger en el barrido se representa por la profundidad de bits (*bit depth*). Entre mayor sea la profundidad de bits, mayor será la cantidad de información recogida y, por ende, será mayor la fidelidad de la imagen con respecto al original. Una profundidad de 8 bits puede representar 256 colores o grados de gris, mientras que una de 24 bits captura más de 16 millones de colores o matices de grises. Obviamente, una mayor cantidad de información impacta en el tamaño de la imagen; sin embargo, algunos escáneres muestrean o barren la imagen con una profundidad mayor que con la que generan una imagen, lo que permite tener una imagen con mayor fidelidad dimensional y tonal. Algunos equipos ofrecen profundidades de entre 36 y 48 bits, con una imagen de salida de 24 (véase cuadro 3).

Profundidad de color		
1 bit	2 colores	2^1
2 bits	4 colores	2^2
4 bits	16 colores	2^4
8 bits	256 colores	2^8
24 bits	16,777,216 colores	2^{24}
36 bits	68,719,476,736 colores	2^{36}
48 bits	281,474,976,710,656 colores	2^{48}

CUADRO 3. La profundidad del color de los escáneres

El número de colores que se pueden representar con respecto a la profundidad de color se obtiene al potenciar el número 2 al número de bits de profundidad:

- Una imagen capturada a un bit estará representada por $2^1=2$ colores.



- Una imagen capturada a 4 bits puede estar representada por $2^4=16$ colores.



- **Densidad óptica**

La densidad óptica es un parámetro que mide el máximo brillo y la máxima oscuridad de una imagen. Impacta en la calidad tonal, así como en la captura de luz y sombra de una imagen. Los equipos recientes incluyen este parámetro y lo denominan *máxima densidad* (dMax). Además de capturar gradaciones con gran detalle, ha mostrado buenos resultados en el procesamiento de transparencias, negativos y otros formatos similares.

- **Velocidad, conexión y funcionalidad con materiales diversos**

Es evidente que a una mayor velocidad se da un flujo de trabajo más ágil en sus diferentes etapas. Para asegurar un trabajo de alta velocidad es aconsejable que el equipo cuente con conectores como Universal Serial Bus (USB) 2.0, tarjetas y cables Small Computer Serial Interface (SCSI) o IEEE 1394 *FireWire*.¹² Siempre se deberá evitar el uso de equipos que utilicen los métodos de conexión más lentos, como puertos paralelos, seriales, USB 1.0, infrarrojo o bluetooth.

4.5.2. Tipos de escáneres

Debido a que existe una gran variedad de marcas y modelos de escáneres en el mercado actual (cama plana, ranura y tambor para formatos grandes, cámaras digitales, etcétera), es preciso conocer las características y posibilidades que presenta cada uno de ellos para seleccionar el más adecuado al tipo de proyecto por desarrollar (véase cuadro 4).

a) *Cama plana*

Los escáneres de cama plana regularmente son de bajo costo y fácil operación (figura 5). Son los más utilizados para cubrir necesidades personales y de pequeñas empresas, así como para procesar documentos, fotografías, libros y

¹² Norma de bus externo (external bus) de alta velocidad que soporta rangos de transferencia de datos superior a los 400Mbps (en 1394a) y 800Mbps (en 1394b). Los productos que se suscriben a la norma 1394 ostentan diferentes nombres, dependiendo de la empresa. Por ejemplo, Apple, quien originalmente desarrolló esta tecnología, usa la marca *FireWire*. Otras empresas adoptan nombres diferentes como i.link y Lynix para describir sus productos 1394.

otros impresos, cuando es pequeña la colección por digitalizar. Es mejor contar con un escáner especializado si la intención es digitalizar una colección de tamaño considerable. El escáner de cama plana resulta de utilidad en los siguientes casos:

- El ejemplar no se puede desencuadernar.
- Las hojas son demasiado gruesas, como diplomas, reconocimientos, etcétera.
- Las hojas sobrepasan las dimensiones de la bandeja de alimentación múltiple.

Los escáneres de cama plana pueden alcanzar resoluciones de 1600 dpi o superiores. Al seleccionar este tipo de equipos es importante observar las dimensiones del área de barrido. Los modelos poco sofisticados ofrecen dimensiones de alrededor de 8.5x11", pero los profesionales permiten escanear documentos de mayores dimensiones, como 12x18".



FIGURA 5. *Escáner de cama plana*¹

b) Escáner de alimentación múltiple (Auto-document Feed, ADF)

Los escáneres de alimentación múltiple permiten digitalizar grandes volúmenes de documentos (figura 6). Cuentan con una bandeja en la que se depositan las hojas sueltas y el aparato las ingresa una por una hasta terminar; los documentos pueden ser digitalizados por uno o ambos lados. Se recomiendan para la digitalización de documentos como facturas, notas o documentos similares conformados por hojas sueltas. Sin embargo, también son de gran ayuda para digitalizar documentos como libros, revistas y similares, siempre y cuando se puedan desencuadernar. Aunque el escáner de alimentación múltiple tiene un

costo relativamente bajo, su resolución es de apenas 600 dpi en color o escala de grises.



FIGURA 6. *Escáner de alimentación múltiple*

c) Escáner mixto.

El escáner mixto reúne las cualidades del de cama plana y las del de alimentación múltiple, aunque la velocidad en el alimentador es relativamente menor comparada con el de alimentación múltiple (figura 7).



FIGURA 7. *Escáner mixto*

d) Escáneres de ranura (para transparencias)

Los escáneres de ranura están diseñados para digitalizar transparencias como diapositivas, negativos y exposiciones de 35 mm, entre otros formatos (figura 8). Aunque es elevado el costo de este tipo de escáneres, esto se ve compensado

con la calidad del producto final, pues alcanzan resoluciones de hasta 4,000 dpi. Generalmente estos escáneres producen una imagen más refinada y detallada que los de cama plana. Esto se debe a que las transparencias reciben mayores cantidades de luz y tienen una mejor definición que los impresos.



FIGURA 8. Escáner de ranura¹³

e) Escáneres aéreos

Los escáneres aéreos son superiores en varios aspectos, empezando por el precio. Pero también pueden alcanzar resoluciones mayores a los 600 dpi. Este tipo de dispositivos está enfocado a la digitalización de libros y documentos antiguos o de gran tamaño (figura 9).



FIGURA 9. Escáner aéreo¹⁴

¹³ Imagen tomada de: <http://www.nikonusa.com/Find-Your-Nikon/Film-Scanners/index.page>

f) Escáner de tambor

En el escáner de tambor los materiales son fijados a una especie, precisamente, de tambor o rodillo, el cual gira para que se efectúe la digitalización (figura 10). Evidentemente, los materiales delicados están descartados para su captura en este tipo de escáneres.

Este escáner es usado frecuentemente en el trabajo de pre prensa y en el diseño gráfico con materiales de uso corriente. También se recomienda para capturar materiales como transparencias o negativos de gran tamaño, a partir de los cuales puede obtener una digitalización de gran precisión, alcanzando resoluciones superiores a los 3,000 dpi. Sin embargo, además de requerir mayores conocimientos para su operación y mantenimiento, su costo es alto.



FIGURA 10. Escáner de tambor¹⁵

g) Escáner para grandes dimensiones

Los escáneres para grandes dimensiones fueron diseñados para procesar planos y cartas geográficas, además de otros materiales de gran tamaño en cuya captura interviene un sistema de rodillos (figura 11). En ocasiones el material sufre a través de éstos algún desgarramiento o ruptura, razón por la cual no se recomienda este tipo de escáner para trabajar con material delicado e irremplazable. Estos escáneres alcanzan resoluciones de hasta 400 dpi y tienen

¹⁴ Imagen tomada de: <http://www.imageware.de/de/>

¹⁵ Imagen tomada de: http://www.aztek.com/Products/AZTEKPremier_lrg.gif

una capacidad para digitalizar documentos de hasta 54 pulgadas. Tiene un costo medio en comparación con los escáneres anteriores.



FIGURA 11. Escáner para grandes dimensiones¹⁶

h) Escáner de micro film

El escáner de microfilm fue concebido para la digitalización de rollos de película y microfichas (figura 12). La calidad o resolución que brinda no es totalmente buena aunque alcanza los 1200 dpi. La definición de las imágenes depende directamente del estado en que se encuentre la película. Su funcionamiento es complicado y costoso y actualmente son pocas las empresas que lo fabrican.



FIGURA 12. Escáner de microfilm¹⁷

¹⁶ Imagen tomada de: <http://h10010.www1.hp.com/wwpc/us/en/ga/WF30a/18972-18972-3328061-12600-3328081-1143093.html>

i) Cámaras digitales

En términos generales, las cámaras más comunes en el mercado no cuentan todavía con suficiente resolución para generar archivos *master* o garantizar la calidad suficiente que las convierta en el formato que preserve un original análogo (figura 13). No obstante, resultan prácticas para digitalizar algunas colecciones como carteles y mapas, así como obras de arte, material tridimensional o de gran tamaño, entre otras, que difícilmente tienen cabida en un escáner convencional. Al igual que los escáneres, entre mayor sea la resolución de la cámara, mejor será la calidad de la imagen.

Una forma de mantener niveles de fidelidad aceptables en la digitalización, es el empleo de una cámara identificada como *digital scan back*. Se trata de una cámara que, provista de un conjunto de sensores, digitaliza una imagen en el anverso y el reverso.

Sin embargo, para trabajar con una cámara *digital scan back* se requiere la asesoría de expertos en su manejo para determinar factores como iluminación, tipos de lentes a usar, etcétera. Por lo regular su costo rebasa el presupuesto de proyectos de mediana envergadura. Se recomienda, por lo tanto, considerar la viabilidad de trabajar con negativos o equipo menos costoso.



FIGURA 13. Cámara digital¹⁸

Tipo de escáner	Aplicaciones comunes	Resolución aproximada	Costo
-----------------	----------------------	-----------------------	-------

¹⁷ Imagen tomada de: <http://www.sunriseimaging.com/speedscan.htm>

¹⁸ Imagen tomada de: http://es.wikipedia.org/wiki/C%C3%A1mara_digital

Cama plana	Documentos como fotografías, libros y otros impresos	Superior a los 1,600 dpi	Bajo
Escáner de alimentación múltiple	Hojas sueltas, documentos guillotizados	600 dpi	Medio
Escáner mixto	Las dos anteriores	Superior a los 1,600 dpi	Medio
Escáneres de ranura	Diapositivas, negativos, exposiciones de 35 mm y similares.	4,000 dpi	Medio
Escáner aéreo	Libros y documentos antiguos o de gran formato y delicados	600 dpi	Alto
Escáner de tambor	Trabajo de pre prensa y diseño gráfico. Transparencias o negativos de gran formato	3,000 dpi	Alto
Escáner para grandes formatos	Planos, cartas geográficas y otro tipo de materiales de formato grande	400 dpi	Medio
Escáner de microfilm	Rollos de película y microfichas	Superior a los 1,200 dpi	Alto
Cámaras digitales	Carteles, mapas, material tridimensional o de gran tamaño		Bajo y alto en el caso de <i>digital scan back</i>

CUADRO 4. Resumen de tipos de escáneres

4.6. Software

Además del equipo de cómputo y los escáneres, hay que considerar el software utilizado en el procesamiento de las imágenes. La mayoría de las veces los escáneres vienen acompañados de programas que permiten realizar aplicaciones como la captura de la imagen, manejo de color, compresión, manipulación, captura de textos, etcétera.

En general, para la digitalización de documentos, cualquier escáner cuenta con la interfaz llamada “Tecnología Sin Un Nombre Interesante” (TWAIN, por sus siglas en inglés)¹⁹, un estándar para la captura de documentos a través de un escáner. Con esta interfaz, el operador tiene la posibilidad de determinar parámetros de captura, como el área de escaneo, la resolución, la profundidad de bits, el modo de color, formato, nivel de brillos y contrastes (véase figura 14).

¹⁹ TWAIN se utiliza como una interfaz de comunicación entre el operador y el escáner o la cámara digital. El estándar fue creado para funcionar con sistemas operativos Windows y Mac. Fue liberado en 1992 y actualmente es mantenido por el Grupo de Trabajo TWAIN. En: <http://es.wikipedia.org/wiki/TWAIN> [Consultado en septiembre 2007]

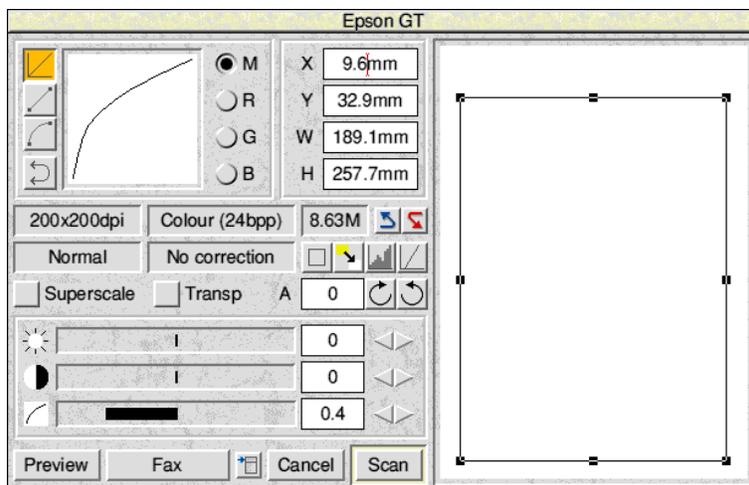


FIGURA 14. La interfaz TWAIN

Sin embargo, si se desea someter las imágenes a otro tipo de procesamiento, como OCR, edición y optimización, será necesario contar con el software especializado en este tipo de tareas. Actualmente, el software de los escáneres más recientes incluye este tipo de funciones.

4.6.1. Gestores documentales

Un apoyo en la creación, almacenamiento, consulta, revisión y distribución de los documentos es el uso de paquetes que permitan administrar cualquier clase de documento, como imágenes, reportes generados por computadora, archivos, formas en HTML, e-mails, fotos, video clips, etcétera. En ocasiones estos paquetes pueden ser creados por la entidad que está realizando el proyecto como traje a la medida, y en otras, pueden ser adquiridos de empresas en el mercado que ofrezcan este tipo de herramientas.

Algunos ejemplos de gestores documentales que actualmente pueden encontrarse en el mercado y que pueden apoyar en la gestión de la información digitalizada son:

On Base	http://www.onbase.com/english/index.aspx
Janium	http://www.janium.com/
DigiTool	http://www.greendata.es/index.php?option=content&task=view&id=35

Para tomar una decisión apropiada sobre la adquisición y el uso de un gestor documental se recomienda tomar en cuenta los parámetros de la norma ISO 15489:2001, en la cual se señala que el software debe permitir:

- a) Determinar qué documentos deberían crearse en cada proceso y qué información ha de contener cada uno.
- b) Decidir la forma y la estructura como deberían crearse los documentos, y las tecnologías que tienen que usarse.
- c) Determinar los metadatos que deberán crearse con los documentos y a lo largo de los procesos documentales.
- d) Determinar los requisitos para recuperar, usar y transmitir documentos entre los diferentes procesos del proyecto.
- e) Decidir cómo organizar los documentos de forma que se facilite su uso.
- f) Preservar los documentos y hacerlos accesibles a lo largo del tiempo.
- g) Cumplir con los requisitos legales y reglamentarios, las normas aplicables y la política de organización.
- h) Garantizar que los documentos se conserven en un entorno seguro.
- i) Garantizar la conservación de los documentos únicamente durante el periodo necesario o requerido.

4.7. Servidores y unidades de almacenamiento

Usar formatos digitales en lugar de analógicos o impresos es uno de los objetivos que se persiguen en un proyecto de digitalización. Con ello disminuye el uso de documentos originales y, por consiguiente, el riesgo de su deterioro físico. Conservar una copia digital de un documento destinada a la preservación hace posible la reproducción, visualización y consulta de la obra de manera eficiente a través de medios remotos como Internet u ópticos, como el CD o el DVD.

4.7.1. Servidores

Para el almacenamiento y la consulta remota de la colección es conveniente disponer de servidores con:

Digitalización de Colecciones. Texto e imagen. Volumen 1

- Capacidad de almacenamiento adecuada al tamaño de las colecciones.
- Memoria RAM y velocidad de procesamiento, tanta como el presupuesto lo permita.
- Unidad lectora y de quemado de CD/DVD y
- Tarjeta de Red con transferencia en Gigabits Ethernet.

Si la colección que se va a digitalizar es pequeña puede ser útil un servidor de mediana capacidad de almacenamiento para guardar los archivos *master* y consultar las imágenes. Si la colección es demasiado grande serán necesarias unidades de almacenamiento externas, capaces de soportar cantidades considerables de información (véase figura 15). En este caso, el servidor podrá funcionar como puente entre la unidad de almacenamiento y el usuario.

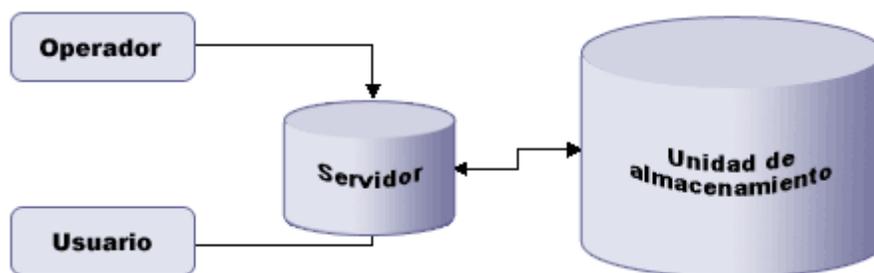


FIGURA 15. Conexión entre el servidor y la unidad de almacenamiento

De igual manera, en el servidor residen los metadatos o información asociada a las imágenes, con la cual el usuario puede consultar y recuperar documentos específicos. En este sentido, la adquisición de un servidor es de suma importancia para acceder a la información, consultarla y depositar las imágenes.

En el mercado existen distintas compañías enfocadas a la venta de servidores con diversas capacidades y cualidades. Partiendo de que las características técnicas de estos equipos evolucionan constantemente, es necesario apoyarse de especialistas y profesionales que asesoren sobre el tipo y necesidad de equipo que se requiera en el proyecto. Es importante destacar que cada empresa enfocada a la venta de servidores ofrece una amplia gama de equipos y regularmente los agrupa en diferentes escalas según capacidades y necesidades. Por ejemplo, para pequeñas, medianas y grandes cantidades de información,

velocidad, etc. Nuevamente, el equipo más adecuado deberá decidirse a partir de las necesidades de cada proyecto.

A continuación se enlistan algunas de las compañías más reconocidas que ofrecen servidores:

Sun Microsystems	http://mx.sun.com/
Apple	http://www.apple.com/es/server/
Hewlett-Packard	http://welcome.hp.com/country/mx/es/welcome.html
IBM	http://www.ibm.com/mx/
Dell	http://www.dell.com.mx

4.7.2. Medios de almacenamiento

Al planear trabajos de digitalización y resguardo es recomendable hacer un cálculo sobre la cantidad y uso de unidades de almacenamiento. Esto permitirá hacer un cronograma de los costos considerando la vida del proyecto. Debe tenerse en cuenta que los medios de almacenamiento mejoran constantemente y, por consiguiente, aumentan su capacidad. Para el almacenamiento de imágenes digitales *master*, debe considerarse la necesidad de grandes unidades de almacenamiento, capaces de soportar millones de Bytes (TeraBytes), además de ofrecer un acceso sencillo y rápido.

Con el propósito de ubicar el nivel de procesamiento del proyecto de digitalización se sugiere la siguiente escala de nivel de almacenamiento:

Baja escala = tamaño de la colección digital de 1 Mbytes a 500 Gbytes.

Media escala = tamaño de la colección digital de 500 Gbytes a 1 Tbytes.

Gran escala = tamaño de la colección digital mayor a 1 Terabytes.

De acuerdo con su tecnología, los medios de almacenamiento pueden clasificarse en ópticos, magnéticos y magnético-ópticos. A continuación se presentan algunos medios de almacenamiento para resguardar colecciones digitales, cuya elección dependerá del uso final que tendrá la colección.

a) Ópticos

- **CDs y DVDs**

Los CDs o DVDs son los medios de almacenamiento ópticos más conocidos. Tienen 12 cm de diámetro y físicamente son iguales. Para su lectura se emplea un rayo láser por medio de una lente que encamina el haz luminoso hacia donde se localizan los datos. El CD tiene una capacidad de almacenamiento de 750 MB en una sola capa. El DVD más comercial es de 4.6 GB, pero vienen en camino los HD DVDs, que prometen alcanzar hasta 45 GB en dos capas. Los DVDs pueden almacenar video con mejor calidad. Para guardar grandes cantidades de información los CDs o los DVDs de poca capacidad no son un medio óptimo, porque se emplearían muchos de ellos, además de que son frágiles y sensibles a los cambios bruscos de temperatura.

- **Blu-ray**

El *Blu-ray* es un sistema de almacenamiento muy similar al DVD, que fue desarrollado por un grupo de empresas: Apple, Dell, Hitachi, HP, JVC, LG, Mitsubishi, Panasonic, Pioneer, Philips, Samsung, Sharp, Sony, TDK y Thomson (figura 16). El desarrollo del formato tiene como objetivo permitir la escritura, la reescritura y la reproducción de video de alta densidad, así como el almacenamiento de información, que va desde los 25 hasta los 50 GB en discos de dos capas. Sin embargo, existen productos en desarrollo que pretenden alcanzar los 200 GB. El *Blu-ray* está basado precisamente en la tecnología del láser azul, con longitud de onda de 405 nm (nanómetros);²⁰ apertura numérica de 0.85, y grosor de capa de 0.1 mm, lo que permite el almacenamiento de grandes volúmenes de información con mayor velocidad y calidad.²¹

²⁰ Nanómetro: una millonésima de milímetro

²¹ Esto se traduce en pulsos más veloces y marcas más pequeñas, que se pueden colocar a una distancia menor, lo que produce una capacidad mayor y velocidades más altas.



FIGURA 16. Blue-ray

- **UDO**

El UDO (Ultra Density Optical) también está basado en la tecnología de láser azul (figura 17). Fue desarrollado por las empresas Hewlett-Packard, Sony y Plasmon. Su apertura numérica es de 0.7 y tiene una capacidad de 30 GB en su primera generación. Se pronostica que su capacidad aumentará hasta 120 GB por disco. El formato del UDO es de 5¼ y tendrá un diámetro de 130 mm. También existen productos en desarrollo que pretenden superar el TB.

Algunas de las bondades de estos medios, son, entre otros: una mayor capacidad de almacenamiento y transferencia de datos; menor costo a largo plazo, y facilidad para ser removibles. Sin embargo, carecen de estandarización, son más costosos y no existe la certidumbre de que sean aceptados fácilmente en el mercado.



FIGURA 17. Ultra Density Optical

b) Magnéticos

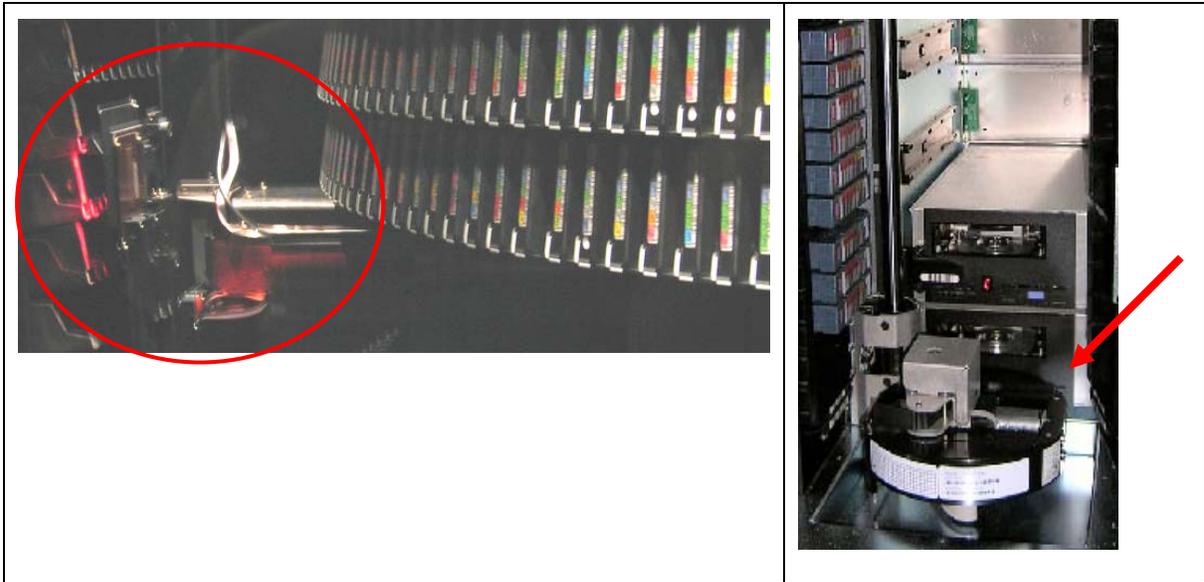
- **Disco duro**

El disco duro (en inglés *hard disk* y abreviado con frecuencia *HD*) es el medio por excelencia para almacenar información permanentemente en una computadora. Los discos duros utilizan generalmente un sistema de grabación magnética digital. Dentro de la carcasa de este medio se encuentra una serie de discos concéntricos de aluminio o cristal, que giran a la vez a gran velocidad. Sobre estos discos se sitúan los cabezales encargados de leer o escribir los impulsos magnéticos. Algunos estándares de comunicación entre un disco duro y la computadora son: Integrated Drive Electronics (IDE), SCSI, SATA (estandarizado en el año 2004), USB y Firewire. Su capacidad va desde 40 hasta 500 Gigabytes, pudiendo agruparse para formar discos duros de mayor capacidad, hasta llegar a los Terabytes. Con el abaratamiento de su costo, el disco duro es una nueva posibilidad para el almacenamiento secundario, en lugar de las cintas magnéticas, CDs, DVDs, etcétera.

- **Robot de cintas**

El Robot de cintas, también conocido como silos, librerías o bibliotecas de cintas, consta de un brazo robótico que mueve las cintas magnéticas en su interior, en una serie de estanterías (figuras 18 y 19). El brazo es capaz de identificar las cintas con la ayuda de un lector de código de barras. El almacenamiento de la información en las cintas es secuencial, por lo que la lectura se realiza en el orden que se guardaron los datos.

El robot más pequeño, que puede ser montado en un rack, maneja decenas de cintas. El más grande ocupa toda una habitación y maneja cientos de cintas. El promedio de capacidad de almacenamiento de los robots actuales de cintas va desde los 1.6 a los 800 TB, a velocidades de transferencia de hasta 14 TB/hora. Este medio es muy recomendado para hacer respaldos masivos de datos. Aunque puede ser escalable, es muy costoso y su mantenimiento requiere de personal calificado.



FIGURAS 18 y 19. Robot manipulando cintas

- **Xserve RAID**

El Xserve RAID es un sistema de almacenamiento diseñado por Apple, que combina tanto la tecnología de los discos Ultra ATA, como la interfaz Fibre Channel. Es compatible con Mac OS X Server y con servidores Windows y Linux (figura 20). Tiene la capacidad para albergar hasta 7 Terabytes distribuidos entre los 14 discos duros que lo componen. Si se conectan varias unidades, su capacidad puede ser prácticamente ilimitada, además de aumentar su rendimiento. Cada disco duro se conecta a un canal de unidad, dedicado a eliminar atascos y optimizar la conexión a 400 MB por segundo.

Es ideal para la gestión de archivos, la impresión, los servicios Web y la emisión multimedia, además de permitir el establecimiento de esquemas para proteger datos. Sin embargo, es costoso, su capacidad de crecimiento está limitada por las características del sistema de archivos del sistema operativo y la conexión a un servidor requiere de una interfaz apropiada (*Fibre Channel*).



FIGURA 20. *Xserve RAID*

- **Centera**

El Centera o Sistema de Almacenamiento de Información Accedido por Contenido de EMC, crea un identificador único, basado en los atributos del contenido, es decir, proporciona una única dirección de contenido a cada objeto almacenado, con el cual se puede llevar a cabo la recuperación de los datos.

Su máxima capacidad de almacenamiento es prácticamente ilimitada, pudiendo ser escalable a múltiples PetaBytes. Además, sus dimensiones son de 190 por 60 cm por cada 40 TB; puede manejar esquemas de respaldo que consumen sólo el 30 por ciento de espacio; cuenta con una resolución automática de problemas en caso de que un componente falle, y evita redundancia mediante la comparación binaria de los contenidos. Sin embargo, es más costoso y lento que el Xserve RAID.

c) Discos Magneto Ópticos (MO)

Estos discos se basan en un sistema de almacenamiento híbrido que funciona con un haz de láser en conjunto un campo magnético (figura 21). Físicamente el tamaño de los cartuchos MO es de 3.5 pulgadas, similares a los disquetes de 3¼. Su capacidad actual es de entre 2.3 y 2.6 GB, dependiendo de la marca. Son de múltiple escritura y no les afectan los campos magnéticos. Su velocidad de transferencia de datos es de 8.38 MB/s.

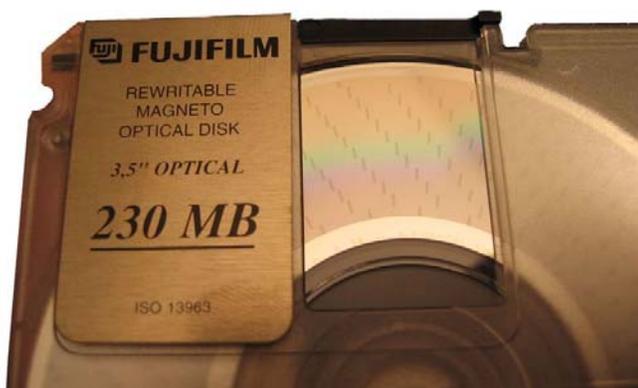


FIGURA 21. *Discos Magneto Ópticos*

En términos generales, para la adquisición y el mantenimiento de la mayoría de los medios de almacenamiento masivo se requiere una buena cantidad de dinero. Sin embargo, se ha podido ver en el mercado que los sistemas de almacenamiento tienen cada vez mayor capacidad, tamaño más reducido y menor costo, además de una mayor compatibilidad entre sí.

Sobre el almacenamiento digital aún existen varias preguntas y aspectos técnicos por resolver como la curación, preservación y arqueología digital, por mencionar algunos. Es importante estar pendientes de los adelantos que se presentan en estas áreas. Por mencionar un caso, Hitachi Data Systems ha elaborado una lista de las principales tendencias que marcarán el mercado del almacenamiento en el corto plazo²²:

1. Controlar las emisiones de carbono.
2. Incertidumbre económica.
3. Incrementar el uso de archivos.
4. Ser conscientes de las deficiencias del almacenamiento.
5. La movilidad de los datos será un requisito clave.
6. Unidad de Control de la Virtualización del Almacenamiento.
7. Almacenamiento orientado a servicios.

²² *Tendencias para el almacenamiento digital en 2008.* Disponible en: <http://www.faq-mac.com/noticias/node/27989> [consultado en abril 2008]

8. Convergencia de contenido, archivo y servicios de almacenamiento basado en bloque.
9. Thin provisioning.
10. De-duplicación.

4.8. Recomendaciones

Es claro que cada proyecto tiene objetivos y metas particulares, por lo cual la selección de la infraestructura tecnológica debe basarse en las necesidades y los requerimientos de cada caso. Pero también existen generalidades como las ya mencionadas en párrafos anteriores, que pueden ayudar en gran medida a tomar decisiones correctas. Aunque en general se piense que el éxito de un proyecto depende de los recursos financieros, debe recordarse que no siempre lo más avanzado en tecnología es garantía de resultados satisfactorios en el marco de las propias necesidades y expectativas.

Se recomiendan dos sitios web para la identificación de los tipos, características y precios aproximados de escáneres, computadoras personales, estaciones de trabajo, monitores y servidores:

- Para los escáneres <http://www.escaner.com.mx/>.
- Para las computadoras personales, estaciones de trabajo, monitores y servidores <http://www1.la.dell.com/content/default.aspx?c=mx&l=es&s=gen>.

4.9. Fuentes consultadas

- *Introducing the new Xserve* En: Apple Servers. Disponible en: <http://www.apple.com/xserve/> [Consultado en abril 2006].
- *Promise RAID and Xsan. Both built to build on* En: Apple Servers. Disponible en: <http://www.apple.com/xserve/raid/> [Consultado en abril 2006]
- Cornell University Library. *Moving theory into practice. Digital imaging tutorial*. Disponible en: <http://www.library.cornell.edu/preservation/tutorial/technical/technicalC-01.html> [Consultado en agosto 2007].
- *Familia Centera de EMC*. En: EMC²: where information lives. Disponible en: <http://spain.emc.com/products/systems/centera.jsp?openfolder=platform> [Consultado en abril 2006].
- Gómez, Luis Guillermo (1997). *Estaciones de trabajo*. En: Monografías.com. Disponible en: <http://www.monografias.com/trabajos41/estaciones-de->

[trabajo/estaciones-de-trabajo.shtml](#) [Consultado en septiembre 2007].

- Gómez, Félix y Werner Guillamón (2005). *La gestión documental y la norma ISO 15489:2001 Record Management*. En: Boletín de Asociación Andaluza de Bibliotecarios. Marzo 2005, No. 78. Disponible en: <http://www.aab.es/pdfs/baab78/78a2.pdf> [Consultado en septiembre 2007].
- Quiroga, Eduardo (2005). *El monitor: tipos y características*. En: Observatorio Tecnológico. Ministerio de Educación y Ciencia de España. Disponible en: <http://observatorio.cnice.mec.es/modules.php?op=modload&name=News&file=article&sid=268> [Consultado en septiembre 2007].
- *Tendencias para el almacenamiento digital en 2008*. Disponible en: <http://www.faq-mac.com/noticias/node/27989> [consultado en abril 2008]
- Wikipedia, la enciclopedia libre. *Copia de seguridad*. Disponible en: http://es.wikipedia.org/wiki/Copia_de_seguridad [Consultado en abril 2006].
- Wikipedia, la enciclopedia libre. *IEEE 1394* Disponible en: http://es.wikipedia.org/wiki/IEEE_1394 [Consultado en septiembre 2007].
- Wikipedia, la enciclopedia libre. *TWAIN*. Disponible en: <http://en.wikipedia.org/wiki/TWAIN> [Consultado en septiembre 2007].
- Williams, Don (2000). *Selecting a Scanner*. En: Guides to Quality in Visual Resource Imaging Digital Library Federation; Research Libraries Group. Disponible en: <http://www.rlg.org/visguides/visguide2.html> [Consultado en septiembre 2007]
- Las imágenes del apartado *Medios de almacenamiento*, han sido tomadas de www.wikipedia.org

5. Preparación de los documentos

Antes de digitalizar una colección es conveniente analizar las diferentes características físicas, gráficas y tipográficas de los documentos que la componen para evaluarlos y separarlos por cualidades similares con la intención de facilitar lo más posible el procesamiento digital, ya sea utilizando en escáner, cámara digital o algún otro dispositivo.

A continuación se enlistan algunos pasos básicos para organizar los documentos y agilizar el procesamiento digital.

5.1 Separación de documentos por su tipo

Los documentos deben separarse de acuerdo con sus características: imágenes en color, blanco y negro, tipografía de baja calidad, tamaño muy reducido, etc. Hacer esto evita reconfigurar frecuentemente el escáner o equipo de captura y ahorra tiempo en el procesamiento digital.

Recuérdese que todo arreglo o separación está en función de las características específicas de cada colección. Incluso es muy probable que ni siquiera haya que separar los materiales por alguna característica particular porque todos pueden ser muy similares; siendo así este paso podrá omitirse en el proceso de digitalización.

a) Un primer paso que se sugiere es separar los documentos en blanco y negro de los que cuentan con abundantes imágenes o textos a color.

b) Dentro de cada grupo, en color o blanco y negro, conviene hacer subgrupos para trabajarlos con mayor resolución o por separado (véase cuadro 5).

Subgrupo	Características
Textual	Predomina el texto con una ligera presencia de tablas, trazos sencillos, etc.
Texto con imagen	Cierta cantidad de ilustraciones, fotografías, diagramas, etc.
Texto de tipografía variada	Tablas con tipografía muy pequeña, diagramas con acotaciones, letra estilizada, manuscritos, etc.

CUADRO 5. Subgrupos para el manejo de documentos

c) Es muy frecuente encontrar documentos en los que es equilibrada la combinación texto-imagen en blanco y negro (grabados, fotografías, ilustraciones, etcétera) o color. Por ello resulta difícil determinar qué predomina, si el texto o la imagen. En tal caso, este tipo de documentos constituye una categoría aparte.

d) Lo ideal sería que todo el material quedara en un solo grupo y digitalizar el título completo de una sola vez, pues una menor separación implica una mínima reconfiguración del equipo, además de un ahorro de tiempo.

5.2 Orden de los documentos

Conviene digitalizar en primer lugar los documentos en los que predomine el texto simple; posteriormente los que muestren sólo imágenes, y después los que combinen ambos. En cuanto a la categorización de los documentos a digitalizar, la Federación de Bibliotecas Digitales (DLF por sus siglas en inglés) ofrece una serie de aspectos a considerar²³ (véase cuadro 6):

Características de contenido	Tipo de documento	Condiciones	Características gráficas
<ul style="list-style-type: none"> • B y n • Color • Monocromo – Texto – Manuscritos – Dibujos – Obras de arte – Miscelánea 	<ul style="list-style-type: none"> – Antiguo – De gran formato – Mapas y láminas – Fotografías – Encuadernado/ desencuadernado 	<ul style="list-style-type: none"> – Condiciones del objeto (rayado, frágil, desgarrado, doblado, difuso) 	<ul style="list-style-type: none"> – Reflejo de luz – Tipo de superficie (grosor, textura, tersura/arrugas) – Rangos de densidad – Detalles de espacios en el texto – Pigmentación y gama de colores – Medios tonos

CUADRO 6. Categorización de materiales de la Digital Libraries Federation

5.3 Preparación física

La encuadernación de libros y documentos actuales suele ser frágil. Las páginas de algunos están pegadas individualmente, en lugar de estar unidas en

²³ Adaptado de William, Don (2000) "Source Material Characterization. Selecting a Scanner" [en línea]. Digital Library Federation. Research Libraries Group. Guides to Quality in Visual Resource Imaging. July 2000. <<http://www.rlg.org/visguides/visguide2.html>> [Consulta: *falta fecha*].

Preparación de los documentos

cuadernillos o cosidas, lo que las hace muy desprendibles desde nuevas, incluso. Por otra parte, aunque hay encuadernaciones más consistentes, éstas se van deteriorando con el tiempo. Si los documentos que se van a procesar cuentan con un valor estético o estimativo, se debe tener en cuenta el riesgo de que sufran algún daño al ser manipulados, sobre todo en un escáner de cama plana: maltrato, desprendimiento de alguna página, quiebre de las mismas por acidez, pérdida de información, etcétera. Por ello se recomienda tratar en general los documentos con el mayor cuidado posible. Es recomendable que el personal encargado reciba capacitación mínima para manipular documentos delicados.

Si se decide digitalizar una colección con escáneres de alimentación múltiple debe considerarse que los documentos tienen que ser desencuadernados. Esto significa que es necesario dedicar más tiempo a la preparación de los documentos, o bien, acudir a una imprenta para que un experto se encargue de hacer la desencuadernación. Si se elige esta última opción debe supervisarse de cerca el proceso para cuidar que los documentos no resulten dañados y las hojas no conserven rebabas o excedentes que impidan su procesamiento. Considérese también que:

- Es riesgoso desencuadernar un documento que tiene pocas probabilidades de ser restituido a su estado original. Los documentos pertenecientes a fondos reservados u obras con un valor artístico están prácticamente fuera de esta opción, a menos que se cuente con la infraestructura para restablecer el documento con la mayor integridad posible. Este procedimiento es más aconsejable para documentos que se usan constantemente, como tesis u otro tipo de obras de las cuales es fácil obtener más ejemplares; o bien cuando la encuadernación es rústica o carece de valor estético.
- La desencuadernación puede hacerse con una guillotina industrial, procurando un corte mínimo y preciso en el lomo del documento, lo que deja una hoja muy limpia de rebabas y facilita su entrada al escáner de alimentación múltiple.
- Se deben conservar las cubiertas, solapas y lomos, con la finalidad de

digitalizar cualquier información relevante que contengan y restituir íntegramente el documento.

5.4 Calidad de los documentos originales

Como se ha señalado anteriormente, es necesaria la revisión visual de los documentos que serán procesados para separarlos por grupos de mayor a menor legibilidad y calidad tipográfica.

Los documentos con líneas tenues en texto y gráficos, suelen requerir mayor contraste o menor brillo, no necesariamente mayor resolución. Al determinar la legibilidad de un documento, es común aplicar criterios de apreciación a partir de la imagen que se presenta en la pantalla de la computadora. El procesamiento no se puede basar sólo en ese principio, pero es útil si se considera que los documentos originales presentan una amplísima gama de tonalidades, grosores de línea, tipos de letra, sombras, calidad de fotografías, grabados, dibujos, etcétera.

Algunos documentos impresos incluyen páginas desplegadas, como mapas, reproducciones, planos, etcétera, que regularmente son de mayores dimensiones. En estos casos deben preverse alternativas, como

- a) Escanear las páginas desplegadas por secciones y unirlos con un editor de imágenes
- b) Fragmentar las imágenes en archivos independientes.

La primera es la opción más recomendable, por ser la más práctica y la que ofrece un formato de lectura más amable para el usuario, sobre todo cuando no es necesario instalar ningún programa extra para visualizar las imágenes.

5.5 Registro y control

Una vez que han sido separados, ordenados y preparados los documentos que se van a procesar, es necesario llevar un registro detallado de ellos, que puede incluir los elementos que se describen en el cuadro 7.

Número identificador	Nombre o título	Fecha de procesamiento	Responsable	Status (listo, pendiente, etc.)	Observaciones

CUADRO 7. Elementos para el registro de documentos

El propósito es identificar todos los componentes de la colección y saber de un “vistazo” en qué situación se encuentra cada documento que la compone.

5.6 Recomendaciones

Dentro de las colecciones existen documentos delicados que deben ser manipulados con la precaución necesaria, a fin de evitar que sufran daños durante su digitalización.

Por otra parte, es necesario llevar un control detallado sobre el avance en cada uno de los procedimientos aplicados a cada documento, tanto como lo amerite el proyecto, pues en la digitalización de los documentos pueden participar varias personas o equipos y sin control de las actividades se corre el riesgo de duplicar o alterar lo ya hecho.

5.7 Fuentes consultadas

- Association of Research Libraries. *Managing digitization activities*. Disponible en: <http://www.arl.org/bm~doc/spec294web.pdf> [Consultado en noviembre 2006].
- *Preparing materials for digitization*. En: Capture Your Collections. A Guide for Managers Planning and Implementing Digitization Projects. Disponible en: http://www.chin.gc.ca/English/Digital_Content/Managers_Guide/pdf.html [Consultado en enero 2008]
- Wilson, Wayne. *Building and Managing a Digital Collection in a Small Library*. North Carolina Library Association. Fall 2003, number 3. Disponible en: http://www.nclaonline.org/NCL/ncl/NCL_61_3_Fall2003.pdf [Consultado en mayo 2007].
- William, Don. *Source Material Characterization. Selecting a Scanner*. En: Digital Library Federation. Research Libraries Group. Guides to Quality in Visual Resource Imaging. July 2000. Disponible en: <http://www.rlq.org/visguides/visguide2.html> [Consultado en mayo 2007].

6. Proceso de digitalización

En la digitalización se pone en funcionamiento la infraestructura que se eligió para el procesamiento de los documentos, los cuales han sido previamente seleccionados y preparados. En esta fase se ponen “manos a la obra” para comenzar el trabajo de producción o reproducción digital.

Debe quedar claro el resultado que se quiere obtener con la digitalización, así como las consideraciones técnicas y operativas que deben aplicarse. Para ello, en este capítulo se presentan algunos conceptos básicos sobre la digitalización y el tipo de archivos que son generados. Se muestran los pasos a seguir en la conversión de un documento a un archivo digital, además de las peculiaridades de la digitalización de textos e imágenes. Finalmente, se aborda el tema de cómo almacenar los archivos, tomando en consideración el formato y el nombre más apropiado.

Algo muy necesario antes de iniciar la digitalización, es contar con la autorización legal para la reproducción del material y su posterior publicación.

6.1. La digitalización

La digitalización de textos e imágenes es el proceso a través del cual se convierten los documentos, usualmente en papel, diapositivas y microfichas, entre otros, a un formato digital, haciendo uso de un dispositivo de captura.²⁴ Éste genera imágenes digitales en distintos tipos de formatos: .jpg, .tiff, .gif, .pdf, etcétera, legibles por medio de una computadora (obsérvese la figura 22).

Como resultado de la primera fase de digitalización se obtiene una imagen, aun cuando el original sea un texto. En este caso, lo que genera el dispositivo de captura es la imagen de un texto, aunque posteriormente, a través de otros métodos, se puede obtener un texto manipulable en un editor de computadora aplicándole tecnología OCR –que se describe más adelante..

²⁴ Como un escáner o una cámara digital, descritos en el capítulo 4.

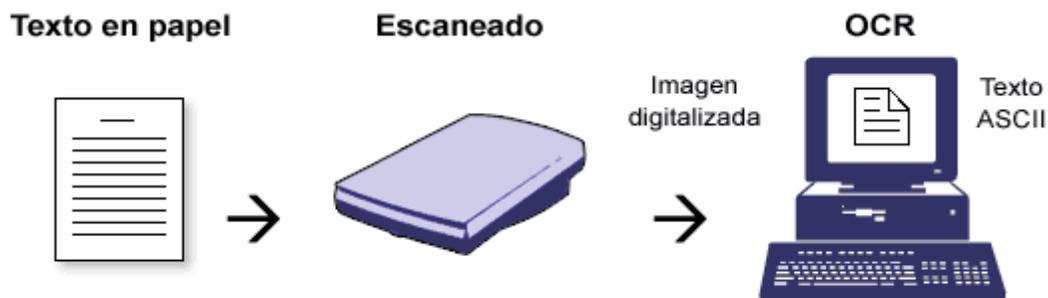


FIGURA 22. Proceso de digitalización

6.2. El archivo master

El archivo que resulta de una captura digital es conocido como archivo *master* porque es una copia maestra de la fuente original, como imagen digital. Es considerado como la base para otros productos o archivos derivados y, opcionalmente, para fines de preservación²⁵. Al tener un archivo *master* se evita volver a digitalizar el original y se cuenta ya con un *original* digital.

Un archivo *master* debe ser:

- De alta calidad o legibilidad;
- Persistente o resistir los cambios de tecnologías para su consulta;
- Confiable respecto a la fuente original: completitud, tonos, colores, resolución, etc.

Es recomendable que en todo proyecto de digitalización que lo requiera se tome en cuenta la creación de archivos *master* con las siguientes características:

- Con la mayor resolución posible.
- Sin compresión.
- Sin edición ni alteraciones.
- Al 100 por ciento de la imagen en captura original, siempre que sea posible y necesario.

²⁵ Véase capítulo 7

- En un formato estándar, fácil de manejar en diferentes editores de imágenes.

6.3. Pasos del proceso de digitalización

El proceso de digitalización consiste en una serie de actividades técnicas para convertir un documento original en un archivo digital.

Cada proyecto define sus especificaciones técnicas, que se derivan de sus particulares objetivos y recursos. Por lo tanto, a través de esta guía no se pretende hacer una revisión exhaustiva de todos los casos posibles, sino establecer un enfoque sobre los pasos generales recomendados para simplificar el proceso de digitalización y evitar posibles fallas:

Configuración de software. Además de instalar el software que se ocupará, se debe probar que su operación sea la adecuada, así como ajustar su configuración, según las especificaciones técnicas que den como resultado la imagen deseada (estas especificaciones se abordan más adelante haciendo una diferenciación entre imágenes y textos).

- *Cerciorarse de que el escáner sea el adecuado*, como puede consultarse en el apartado 4.5.2, y que su configuración sea la correcta. Para determinarla deben realizarse algunas pruebas. Si se presentan problemas debe recurrirse al proveedor o al área de soporte técnico que respalda al proyecto.
- *Tener acceso a los documentos por digitalizar.* Algunas veces los documentos no pueden salir de una zona de seguridad, por lo que la unidad de digitalización es la que debe desplazarse hasta ellos.
- *Verificar que el proceso de captura se realice sin contratiempos*, como errores de software, hardware, alimentación de los equipos, de recurso humanos, etc.
- *Verificar las imágenes resultantes*, para comprobar que el archivo digital se obtuvo como se había planeado originalmente. Cualquier error en este paso podrá repercutir en otros posteriores y se podrá corregir sólo reescaneando

la imagen original.

- *Almacenar las imágenes digitales.* El almacenamiento de grandes volúmenes de archivos requiere de una organización cuidadosa y bien estructurada. De otra manera, los archivos pueden confundirse o extraviarse y ello obligaría a trabajar doble.

6.4. Digitalización de imágenes

La digitalización de una imagen consiste en su captura por medio de un escáner u otro dispositivo digital, para generar así un archivo digital, es decir, un archivo *master* (figura 23).

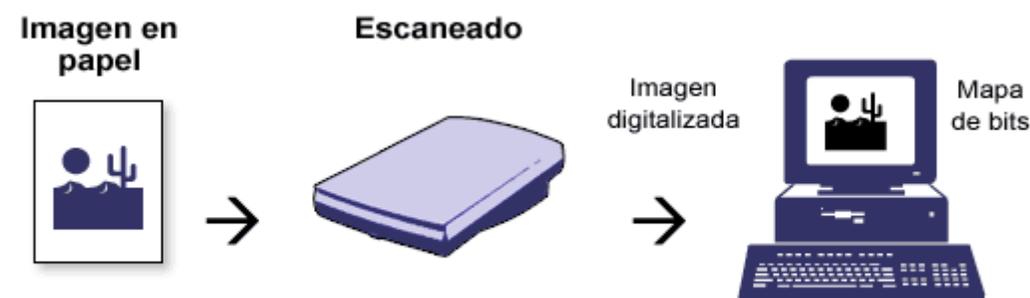


FIGURA 23. Obtención de un archivo *master*

La digitalización de una imagen depende del colorido o la profundidad de color que requiere la propia imagen digital, así como de la calidad que se obtiene en términos de resolución al ser escaneada, es decir, el número de puntos por pulgada lineal (dpi) que recorrerá el escáner y la cantidad de información que deberá contener cada punto. A mayor resolución y número de bits por píxel, se obtendrá un archivo de mayor tamaño.

En un proyecto de digitalización de imágenes consultables en la web es recomendable la generación de tres tipos de imágenes: *master* u original; *de acceso* –que consultan los usuarios en Internet, por ejemplo– y la *imagen miniatura*, también conocida como *thumbnail*, útil para tener una idea, en un solo vistazo, de lo que será descargado o consultado. Sin embargo, no debe perderse

de vista que estas imágenes tienen un impacto en el espacio del disco duro necesario para el almacenamiento de una colección (véase cuadro 8).

Imagen <i>master</i>	Imagen de acceso	Imagen miniatura
<ul style="list-style-type: none"> • Representa lo más fielmente posible la información contenida en la fuente original. • Sin compresión. • Sin edición. • Funge como fuente de larga duración para hacer copias de ella. • Funge como sustituto del original. • Cuenta con alta calidad. • Generalmente es de gran tamaño, pues conserva íntegra la información del escaneo. • Sirve para crear reproducciones impresas de alta calidad. • Comúnmente es formato TIFF (texto b&n), JPG (imágenes color) o PDF (texto o imagen/texto). 	<ul style="list-style-type: none"> • Se utiliza en lugar de la imagen <i>master</i> para acceso general vía Internet. • Generalmente se ajusta al área de visualización de un monitor estándar. • El archivo es de tamaño razonable para ser descargado en poco tiempo desde una conexión estándar. • Calidad aceptable para la visualización en pantalla (o incluso impresión). • Opcionalmente comprimida para agilizar el acceso. • Usualmente almacenada en formato estándar (pdf, jpg, gif, etc.). 	<ul style="list-style-type: none"> • Es una imagen muy pequeña. • Diseñada para desplegarse rápidamente en línea; permite al usuario decidir si desea o no desplegar o descargar la imagen (<i>imagen de acceso</i>) • Comúnmente se almacena en formato GIF o JPG. • No siempre es conveniente para las imágenes primordialmente textuales, partituras, etcétera. A una escala tan pequeña, el usuario no es capaz de distinguir el texto.

CUADRO 8. Adaptación de *Digital Imaging Best Practices*²⁶

El tamaño de un archivo repercute en la velocidad de recuperación, un factor crítico para la consulta en línea.

La calidad y el tamaño de un archivo *master*, dependen de dos factores principales: la profundidad de color y la resolución.

6.4.1. Profundidad de color

La profundidad de color es el número de bits utilizados para representar cada píxel. Es determinante en la calidad de una imagen y está dada por la gama de colores que pueden capturarse. Básicamente se manejan tres modalidades:²⁷

- *Bitonal o blanco y negro*. Este tipo de captura es recomendada para los documentos que requieren altos niveles de contraste, como los textos impresos (véase figura 24). Para su representación se utiliza sólo un bit,

²⁶ Western States Digital Standards Group. *Western States Digital Imaging Best Practices vers. 1.0*. University of Denver and the Colorado Digitization Program, 2003. disponible en: <http://www.bcr.org/cdp/best/digital-imaging-bp.pdf>. [Consultado en noviembre 2007]

²⁷ Western, *op.cit.*

Digitalización de Colecciones. Texto e imagen. Volumen 1

pues cada punto puede ser blanco o negro. En el caso de fotografías hechas con esta profundidad de color, la calidad visual será deficiente, la imagen austera y las líneas pueden aparecer irregulares.



FIGURA 24. Captura con alto nivel de contraste

- *Escala de grises.* Este tipo de captura se recomienda para documentos de tonos continuos, como fotografías en blanco y negro (véase la figura 25). Para su representación se utilizan 8 bits: cada punto puede estar dentro de 256 posibles tonos de gris, dando una imagen clara y buena. Es funcional para la mayoría de las imágenes que no tienen colores.

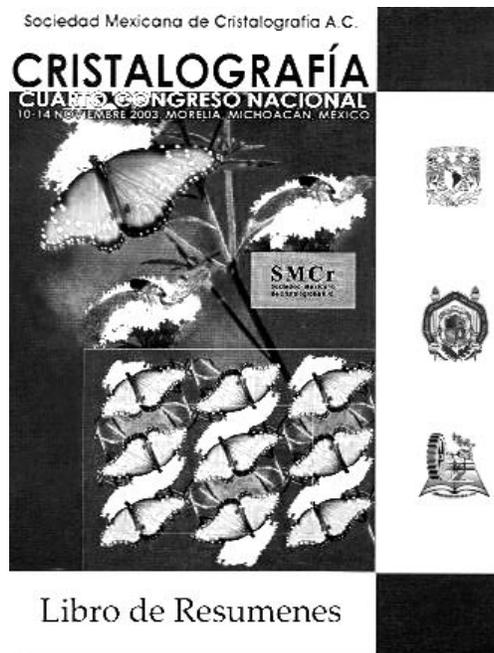


FIGURA 25. Imagen con tonos continuos

- **Color.** Es recomendable para documentos con tonos continuos y de variados colores, como fotografías a color, diapositivas y láminas (ver la figura 26). Se puede representar por 8 bits de color, lo que indica que cada punto puede equivaler a 256 colores, dando como resultado una imagen poco real en las fotografías, con una apariencia granulosa, aunque funciona bien con dibujos e imágenes, sobre todo para la web. Hablando de 24 bits, la gama se incrementa a 16.8 millones de colores, dando lugar a una digitalización de mayor calidad. El resultado es una imagen realista y perfecta para las fotografías, aunque el tamaño del archivo se vuelve grande.



FIGURA 26. Obtención de imágenes fotográficas

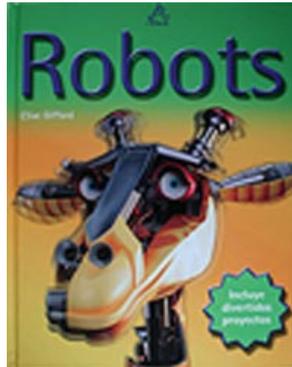
Para capturar un documento compuesto básicamente por texto es mejor utilizar la técnica del blanco y negro o bitonal. No hay necesidad de alcanzar una gran nitidez o conseguir detalles como en una imagen fina. En el caso de material manuscrito, antiguo o similar, es más conveniente utilizar la escala de grises, ante la probabilidad de que haya caracteres poco nítidos y detalles que valga la pena destacar. En el caso de textos en los que predomine la tinta negra, pero contengan anotaciones marginales a color que convenga mostrar, es mejor capturar la imagen a color. Sin embargo, no hay que olvidar que esto significa un aumento en el tamaño del archivo. Finalmente, esta decisión estará en función de las prioridades del proyecto.

6.4.2. Resolución

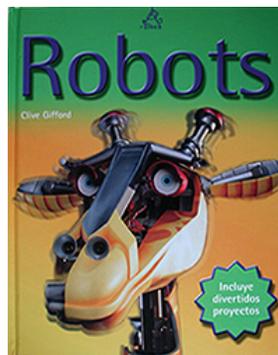
La apariencia y la calidad de una imagen digitalizada también dependen de la resolución, entendida ésta como el número de píxeles por unidad lineal, la cual es medida en puntos por pulgada (dpi). En la digitalización de imágenes, la decisión acerca de la cantidad de dpi a utilizar se rige regularmente por consideraciones

prácticas. Entre más grande sea el número de dpi, mayor será la información que contenga un archivo y más fino será el detalle de la imagen.

Para el despliegue de imágenes en pantalla, no se recomienda una resolución mayor a 100 dpi, pero si una imagen está pensada para impresiones en papel o ampliaciones en pantalla, sí es conveniente una resolución superior. Si se trata de documentos valiosos como manuscritos o ejemplares únicos, no debe escatimarse en su calidad a la hora de digitalizarlos para su preservación. En este caso puede archivar una copia en alta resolución y hacerse otra, a menor resolución, para ser difundida en línea o cualquier otro servicio que tenga como salida la pantalla de una computadora. Las imágenes de la figura 27 están digitalizadas a diferente resolución. Puede observarse una clara diferencia en la calidad de una y otra.



50 dpi. Tamaño: 8 kbytes



400 dpi. Tamaño: 40 Kbytes.

FIGURA 27. *Imágenes digitalizadas a diferente resolución*

Se realizó una prueba con una imagen escaneada a 24 bits, en cuatro resoluciones y dos tipos de almacenamiento, sin compresión y con compresión, TIFF y JPEG, respectivamente. El resultado se muestra en el cuadro 9. En él

Digitalización de Colecciones. Texto e imagen. Volumen 1

vemos la importancia de elegir la resolución adecuada, porque el tamaño varía notoriamente de una resolución a otra.

	DPI			
Formato	100	200	400	600
TIFF	1.93 Mbytes	7.81 Mbytes	31.3 Mbytes	70.3 Mbytes
JPEG	114 Kbytes	230 Kbytes	660 Kbytes	1.2 Mbytes

CUADRO 9. *Tamaño de imágenes a diferente resolución y diferente formato de archivo.*

Es importante destacar que no existe una resolución “única y perfecta” para la digitalización de *todas* las colecciones y *todos* los materiales. Para elegir la resolución apropiada, se debe considerar:

- El tamaño y la calidad de la fuente original.
- El tipo de material.
- Los usos que se les darán a los archivos digitales.
- Los tamaños mínimos y máximos en acordados para el archivo digital.

6.5. Digitalización de textos

Los textos digitales pueden producirse de tres maneras. La primera es la digital o digital de origen, lo que quiere decir que de los textos de este tipo no existe antecedente alguno en ningún medio impreso. Se generan directamente en un dispositivo, el cual permite su posterior almacenamiento en un medio digital, en cualquiera de sus formatos.

La segunda es obtener imágenes digitales a partir de la digitalización de textos impresos en papel, diapositivas o microfilms. El texto en imagen digital puede resultar útil, principalmente cuando se pretende mostrar íntegro un documento en forma y estilo, como un manuscrito.

La tercera es generando un texto digital manipulable, es decir, que pueda ser modificado, reorganizado o procesado, a partir de una imagen (de texto) digital; esto se logra aplicando tecnología OCR a la imagen de texto –que veremos más adelante.

6.6. Reconocimiento Óptico de Caracteres (OCR)

El Reconocimiento Óptico de Caracteres u OCR, por sus siglas en inglés (véase figura 28), es un proceso a través del cual textos en papel son convertidos en imágenes con el uso de un escáner y un software para el propósito. Los caracteres que contienen estas imágenes son interpretados y convertidos a texto digital, el cual puede ser almacenado en algunos de los formatos de texto más comunes: .DOC, .TXT, RTF o PDF. Este proceso es útil en la mayoría de las tipografías de los siglos XIX y XX. Sin embargo, para documentos con baja calidad de impresión o familias tipográficas complejas, el OCR no es una buena opción, pues cada carácter que no es reconocido correctamente y automáticamente por el software requiere de un trabajo extra de corrección. A veces ésta se convierte en una tarea más difícil y pausada que la captura manual del documento.

Hace algunos años el porcentaje de error del proceso de OCR se encontraba entre el 2 y el 5%, ahora el software es mucho más eficiente y se obtienen textos con un porcentaje de error menor al 1%. Claro que este porcentaje está directamente relacionado con la calidad del documento original, siendo recomendable todavía, en raras ocasiones, recurrir a un buen mecanógrafo para capturar el texto, evitando con ello el cotejo para encontrar y corregir los errores.

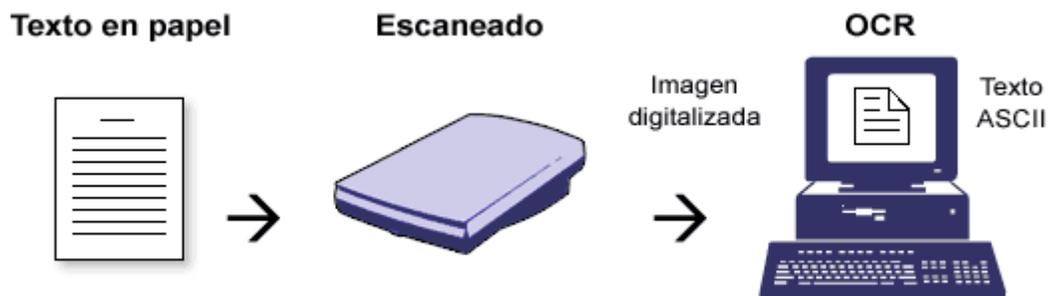


FIGURA 28. Reconocimiento Óptico de Caracteres (OCR)

Para evitar exhaustivas y grandes correcciones posteriores, es recomendable hacer pruebas antes de iniciar un largo proceso de OCR. Con frecuencia algunas

letras son intercambiadas por otras, sobre todo tipografías cursivas: una e puede ser tomada por una c, o una l por un 1.

Cuando se trata de manuscritos no es posible hacer muchas pruebas, porque usualmente no se pueden someter a un proceso de OCR. Si fuera el caso, el margen de error sería muy alto, por lo que la captura manual del texto es la mejor opción, o bien, la digitalización del documento como imagen digital.

El software para realizar el proceso de OCR ha evolucionado rápidamente y los productos cada vez son más amigables y eficientes, entre sus principales funciones está poder entregar diferentes formatos de los archivos de salida y el reconocimiento de diferentes idiomas. También cuentan con un proceso de “aprendizaje”, de tal manera que con el uso su capacidad de reconocimiento de cada carácter es mucho mejor y entregan textos prácticamente sin errores. En el mercado se encuentra software comercial como OmniPage, Abby FineReader, Docs2Manage y Top OCR, como los más populares; entre los productos de software libre están SimpleOCR y Softi FreeOCR.

6.7. Almacenamiento de los archivos

Una vez que un documento es capturado con un escáner, siguiendo las especificaciones técnicas requeridas, es el momento de almacenar la imagen digital obtenida. El almacenamiento de archivos digitales debe hacerse en una unidad de almacenamiento que permita el acceso inmediato y garantice la disponibilidad de los archivos digitales (Véase el subapartado 4.7 sobre la infraestructura de almacenamiento). Debe considerarse también una organización adecuada de los archivos generados para cada tipo de colección y los servicios que proporcionará.

Para producir una colección digital, cada imagen procesada debe almacenarse adecuadamente. Para ello debe considerarse el formato del archivo *master*, así como el nombre que le será asignado.

6.7.1. Formato del archivo

Los formatos con compresión y los formatos sin compresión tienen características diferentes y usos también específicos (véase apartado 6.4).

No debe olvidarse que el tamaño de un archivo repercute en la velocidad a la que es recuperado y en el espacio donde se encuentra almacenado, lo que no es crítico cuando se trata de una sola imagen, pero sí cuando se habla de cientos o miles de ellas. Así, es fundamental una valoración seria de la relación calidad-tamaño que debe tener un documento almacenado, según su uso: consulta, difusión o preservación.

En el cuadro 10 resulta interesante observar que el tamaño de un archivo disminuye drásticamente cuando una imagen es almacenada sin compresión.

Formato del archivo			
Profundidad de color y compresión	TIFF	JPEG	GIF
24-bit color	1.93 Mbytes	114 Kbytes	254 Kbytes
8-bit color	666 Kbytes	271 Kbytes	238 Kbytes
8-bit grises	666 Kbytes	81 Kbytes	431 Kbytes
1-bit Blanco/Negro	749 Kbytes	No aplica	106 Kbytes

CUADRO 10. *Tamaño de archivos según profundidad de color y compresión.*

Las pruebas realizadas y la naturaleza diferente de los documentos llevan a la conclusión de que no hay un método único para la digitalización de imágenes. La calidad es importante en algunos casos, debido a cuestiones de preservación de un documento. No debe escatimarse en el tamaño de un archivo para que una imagen cuente con los parámetros óptimos de digitalización. En otros casos, deberá encontrarse un buen balance entre el número de bits de digitalización, la resolución y el tipo de compresión, para obtener una imagen digital con una calidad aceptable sólo para fines de difusión.

6.7.2. Nombres de archivos

A los archivos que guardan las imágenes digitalizadas se les debe asignar un nombre único e inconfundible, pues la práctica demuestra que no es suficiente un número consecutivo²⁸. En algunos casos el nombre del archivo puede repetirse, pero no así el del documento.

Ejemplo:

1025487_A / 00001A, 00001B, 00002A, 00002B, etc.

1025488_A / 00001A, 00001B, 00002A, 00002B, etc.

Si los archivos individuales no se encuentran en el mismo subdirectorio o carpeta, no hay lugar a confusión.

Algunas instituciones reconocidas, como la Biblioteca del Congreso de los Estados Unidos²⁹ y la Federación de Bibliotecas Digitales, sugieren que los nombres de los archivos digitales deben obedecer a criterios establecidos por los propios organismos desarrolladores de proyectos de digitalización o compilación de recursos. En otras palabras, dan por válidas las decisiones locales, siempre que el nombre asignado sea funcional para un proyecto. Sin embargo, visualizando una normalización que favorezca el intercambio de información y de las colecciones, existe la opción de seguir algunos lineamientos básicos.

Tanto la Universidad de Michigan como la *Joint RLG/TASK Force on Digital Preservation*³⁰, recomiendan el seguimiento de las siguientes consideraciones para dar nombre a un archivo digital:

- Observar la norma de nombramiento ISO 9660, que define como útil un sistema de archivo en una variedad de sistemas operativos.
- Establecer un nombramiento convencional de archivos y las extensiones (variaciones, copias) que posteriormente se generarán a partir de éste.
- Basar los nombres en números de acceso, control, inventario, sistema u

²⁸ *Identifiers for digital resources*. Disponible en: <http://memory.loc.gov/ammem/award/docs/identifiers.html>

²⁹ *Identifiers...* *op cit.*

³⁰ Cory Snaveley. *Handle Service Conventions and Practices*. Disponible en: <http://www.umdl.umich.edu/um-rlg.html> [Consultado en diciembre 2007]

otro identificador único, establecido por la institución encargada del proyecto.

- Evitar la sensibilidad a las mayúsculas y las minúsculas, así como el uso de signos diacríticos.

Además, conviene observar los siguientes puntos:

- Utilizar una estructura de nombramiento de archivos compatible con cualquier sistema operativo y los medios de almacenamiento que se planee utilizar.
- Utilizar extensiones de archivo estándar para diferentes tipos de archivos.
- No sobrecargar directorios con demasiados archivos para no entorpecer la consulta.
- En el caso de colecciones grandes, dejar que sea el propio software de captura el que asigne nombres seriados.
- Nombrar calculando el crecimiento de la colección.

El nombre de un archivo puede contener caracteres alfabéticos, numéricos o una combinación de ambos. La cadena que lo componga puede tener un significado lógico: una fecha, siglas, claves de la institución, etcétera; guardar una relación con el contenido del archivo: iniciales, primeras letras, etcétera, o ser simplemente una secuencia progresiva de caracteres numéricos.

Hay dos tipos de nombres para los archivos: descriptivos y no descriptivos.

Nombres descriptivos. Denotan un significado. Por ejemplo, si una colección está separada por años, los cuatro dígitos que la identifican se pueden utilizar como nombre del directorio y del archivo o archivos. Esto es común en archivos históricos o fondos de imágenes.

Ejemplo:

a) 1950 / 19500001.tif

b) SSM / SSM00001.pdf

Nombres no descriptivos. Carecen de significado. Hay escáneres para trabajo masivo que asignan automáticamente un nombre secuencial a cada

Digitalización de Colecciones. Texto e imagen. Volumen 1

archivo que generan, aunque este nombre también puede estar preconfigurado.

Ejemplo:

a) 00001A, 00001B, 00002A b) A01.pdf, A02.pdf

6.8. Recomendaciones

- Cerciorarse de contar con nombres únicos para cada documento.
- Utilizar nombres significativos para los documentos siempre que sea posible y hacerlo en función de las capacidades del equipo y las necesidades del proyecto.
- Evitar nombres muy extensos que puedan dificultar su lectura en distintos sistemas operativos.
- Incluir la definición del nombramiento de archivos y documentos en las políticas que regirán al proyecto.

6.9. Fuentes consultadas

- Biblioteca de la Universidad de Cornell. *Tutorial de digitalización de imágenes*. Disponible en: <http://www.library.cornell.edu/preservation/tutorial-spanish/contents.html> [Consultado en enero 2008].
- Cory Snavely. *Handle Service Conventions and Practices*. Disponible en: <http://www.umdl.umich.edu/um-rlg.html> [Consultado en diciembre 2007].
- *Identifiers for digital resources*. Disponible en: <http://memory.loc.gov/ammem/award/docs/identifiers.html>. [Consultado en enero 2008].
- Tanner, Simon (2004). *Deciding whether Optical Character Recognition is feasible*. Disponible en: http://www.odl.ox.ac.uk/papers/OCRFfeasibility_final.pdf [Consultado en enero 2008].
- Western States Digital Standards Group. *Western States Digital Imaging Best Practices vers. 1.0*. University of Denver and the Colorado Digitization Program, 2003. Disponible en: <http://www.bcr.org/cdp/best/digital-imaging-bp.pdf>. [Consultado en noviembre 2007]

7. Respaldos y preservación digital

Las colecciones digitales crecen a un ritmo acelerado, como ha sucedido durante los últimos veinte años. Este crecimiento sostenido y, hasta cierto punto, incontrolado, plantea la necesidad de procedimientos que garanticen no sólo la permanencia de las colecciones, sino también que sean consultables y recuperables, independientemente de los cambios tecnológicos.

En este capítulo se presentan los fundamentos teóricos que marcan la diferencia que existe entre respaldar y preservar, conceptos que suelen emplearse simultáneamente porque tienen aspectos en común, como garantizar a futuro el acceso a los documentos, aunque también marcados contrastes.

7.1. *Conceptos y definiciones*

7.1.1. **Definición de respaldo**

El respaldo, también conocido como copia de seguridad, se refiere a la existencia de una réplica de los datos o la información de un sistema, para que éste pueda ser restaurado en caso de fallas o desastres. En este sentido un respaldo es utilizado como un plan de contingencia, para restaurar un equipo de cómputo a un estado operacional luego de un desastre, o bien, para recuperar datos o información que se hayan borrado o corrompido por cualquier causa.

7.1.2. **Definición de preservación digital**

El proceso de preservación digital amerita la siguiente reflexión:

- ¿Preservar es lo mismo que transferir documentos originales a un formato digital, o
- se refiere a preservar los documentos digitales ya creados?

Ambos enfoques son válidos. Dentro de este ámbito, además de que existen muchos métodos digitales para la preservación de originales, hay una cantidad importante de proyectos de digitalización de fondos históricos. La digitalización

Digitalización de Colecciones. Texto e imagen. Volumen 1

contribuye a mejorar el acceso y la consulta de los fondos históricos, además de preservar el original, pues al hacer uso de una copia digital se evitan los efectos nocivos que se derivan de la manipulación física de los documentos. En algunos medios analógicos, como las cintas magnéticas, la digitalización ayuda a proteger la calidad de la información ante la degradación natural que con el tiempo sufren los soportes.

La idea de preservar los documentos digitales surge por la fragilidad de los soportes de almacenamiento de la propia información digital, sin contar los rápidos avances de la tecnología y la continua obsolescencia de los soportes de almacenamiento, el hardware y el software.

La preservación digital consiste en una serie de actividades necesarias y muy bien administradas para asegurar el acceso continuo a las colecciones digitales durante el tiempo que sea necesario. Son acciones que permiten el acceso a las colecciones digitales, aun cuando se presenten fallas en los medios de almacenamiento o haya cambios de tipo tecnológico.

Por periodos, la preservación se clasifica en tres grupos:

- **Preservación de duración larga:** acceso constante a las colecciones digitales, prácticamente de manera indefinida.
- **Preservación de duración media:** acceso ininterrumpido a las colecciones digitales, aun después de cambios tecnológicos en un periodo determinado, sin garantizarse el acceso indefinido.
- **Preservación de duración corta:** acceso a las colecciones digitales durante un periodo definido o en un lapso menor a los cambios tecnológicos.

7.1.3. Diferencias entre preservación y respaldo

La preservación digital es diferente a un respaldo o una copia de seguridad. Lo que se guarda como copia de seguridad en un sistema de colecciones digitales, son básicamente dos cosas:

a) por una parte la información publicada en el servidor: documentos digitales más información de catálogo

b) y por otra, los documentos digitales en proceso de edición.

La preservación digital no se ocupa de respaldar los datos del servidor y el material de trabajo diario, sino salvaguardar los documentos digitales que se necesitarán en el futuro. Debido a que para muchos usuarios de colecciones digitales el ancho de banda es una limitante, la información publicada en el servidor comúnmente es comprimida o su calidad se sacrifica para reducirla de tamaño y descargarla con facilidad. La información digital seleccionada para ser preservada debe tener la mayor calidad posible para usos futuros.

Debe hacerse una distinción entre el material de archivo y el derivado para el acceso del público. Este modelo de almacenamiento digital incluye una versión *master* de la colección digital, con los documentos de alta calidad que se preservan y una colección de acceso público, con formatos generados automáticamente a partir del acervo preservado.

Las copias de seguridad y las de preservación se basan en la redundancia de la información mediante grabaciones periódicas, pero existe entre ellas una diferencia en cuanto a la forma de organizar las grabaciones y los tiempos. En las copias de seguridad se pueden seguir diversos métodos, como la copia integral, la copia incremental o las copias rotativas, con una periodicidad diaria y semanal generalmente alta. En el caso de las copias de preservación, por el contrario, el método suele ser la grabación integral del documento por una vez y el copiado del mismo una vez al año o al año y medio en un soporte nuevo como forma de rejuvenecimiento.

7.2. Algunas problemáticas de la preservación

Durante el proceso de digitalización, la preservación de colecciones digitales se enfrenta a aspectos críticos que deben tomarse en consideración. Algunos de ellos son:

- Enfoques emergentes y cambiantes para seleccionar y catalogar (en el caso de colecciones de bibliotecas).
- Multiplicidad de formatos.
- Cambios rápidos en la tecnología.
- Obsolescencia de hardware y software.
- Condiciones legales, sociales y económicas.

Por otra parte, hay documentos en los que deben hacerse algunas consideraciones especiales en el momento de hacer una preservación:

Libros electrónicos

- Diversidad de estándares;
- Bajo desarrollo de medidas de seguridad en el mercado.
- Dispositivos de hardware y software propietarios.

Revistas electrónicas

- Acceso remoto administrado por proveedores.
- Artículos con hipervínculos a documentos externos en línea.

Este último punto en particular lleva a la formulación de dos preguntas: ¿vale la pena preservar todos los enlaces vertidos en un artículo? y ¿tenemos el derecho y la capacidad de hacerlo?

7.3. Respaldo y recuperación

Los sistemas de cómputo que contienen las colecciones digitales están expuestos a riesgos latentes. Pueden interrumpir el servicio debido a alteraciones en la electricidad, el hardware, el software y la red, así como a fallas humanas, desastres naturales y ataques informáticos como virus y sabotaje, entre otras.

Por tal motivo es necesario contar con un adecuado plan de contingencia que permita respaldar la información para garantizar la recuperación de la misma. En un plan de contingencia es necesario:

- Identificar y priorizar los procesos y los recursos indispensables.
- Analizar el riesgo y el impacto por la pérdida de la información.
- Evaluar recomendaciones de protección.
- Contar con estrategias y alternativas de recuperación.
- Establecer los equipos de trabajo y las funciones de cada persona.
- Ejecutar simulacros del plan de contingencia.
- Elaborar un manual de contingencia.
- Retroalimentar el plan.

Para llevar a cabo un plan de contingencia es recomendable realizar algunas de las siguientes actividades:

- Seleccionar el medio de almacenamiento secundario.
- Determinar la frecuencia de realización de copias de seguridad
- Determinar el volumen de la información a respaldar.
- Determinar días y horario en que deben realizarse los respaldos.

Además de permitir la identificación de la mejor manera de recuperar la información en caso de desastre, una *estrategia de recuperación* es una guía para el desarrollo de los procedimientos mismos de recuperación.

7.3.1. Respaldo tradicional

El respaldo tradicional consiste en copiar los datos o la información de un sistema a un medio de almacenamiento secundario, como cinta, CD y DVD, entre otros, con el fin de que pueda ser restaurado en caso de fallas o desastres. Su periodicidad puede ser diaria, semanal o mensual y difícilmente menor a un día. Para realizar las copias los métodos a seguir pueden ser los siguientes:

- *Copiar sólo los datos.* No proporciona las facilidades para recuperar el entorno operacional que proporcionan los programas de aplicación para

acceder a los mismos.

- *Copia completa.* Incluye una copia de datos y programas que permite restaurar el sistema hasta el momento anterior a la copia.
- *Copia incremental.* Solamente se almacenan las modificaciones realizadas después de la última copia de seguridad. Debe mantenerse la copia original para restaurar posteriormente el resto de las copias.
- *Copia diferencial.* Es similar a la incremental, pero en lugar de copiar las modificaciones, son almacenados los archivos completos que han sido modificados. También se necesita la copia original.

7.3.2. Respaldo con tecnología RAID

En el mejor de los casos, el sistema de respaldo tradicional se aplica todos los días, comúnmente por la noche, cuando disminuye la carga de trabajo del servidor. Esto significa que si se presenta un incidente en el transcurso del día o, en las circunstancias más adversas, por la tarde, no sería posible recuperar el trabajo realizado. Para muchas empresas esto puede representar grandes pérdidas financieras. Para muchas empresas esto puede representar grandes pérdidas financieras. En el caso de los bancos, por ejemplo, no pueden perder las transacciones realizadas a lo largo del día. Para este tipo de contingencias la solución tecnológica es el uso del RAID (*Redundant Array of Inexpensive Disks* o Conjunto redundante de discos baratos y, actualmente, *Redundant Array of Independent Disks* o Conjunto redundante de discos independientes).

En informática, el acrónimo RAID se refiere a un sistema de almacenamiento en el que se usan múltiples discos duros, entre los que son distribuidos o replicados los datos. Dependiendo de su configuración, a la que suele denominarse “nivel”, los beneficios de un RAID con respecto a un único disco son:

- Mayor integridad.
- Tolerancia a fallos.
- Rendimiento y capacidad.

En sus orígenes, la principal ventaja de RAID radicaba en su capacidad de combinar varios dispositivos de bajo costo con una tecnología más antigua, para dar como resultado un conjunto que ofrecía mayor capacidad, fiabilidad, velocidad, o una combinación de éstas, que un solo dispositivo de última generación y costo mayor.

En el nivel más simple, RAID combina múltiples discos en una sola unidad lógica: en lugar de identificar diferentes discos, el sistema operativo sólo reconoce uno. Así, el RAID agrupa dos o más discos duros, ofreciendo una forma más avanzada de respaldo, puesto que:

- Es posible mantener copias en línea (redundancia).
- Agiliza las operaciones del sistema, sobre todo en bases de datos.
- El sistema es capaz de recuperar información, sin la intervención de un administrador.

Hablar del nivel o la configuración del RAID, es referirse a la arquitectura que determina la redundancia y cómo están distribuidos los datos a través de los discos duros del arreglo. Existen varias configuraciones del RAID, sin embargo, los cuatro tipos que prevalecen en muchas arquitecturas son RAID-0, RAID-1, RAID-3 y RAID-5.

Cabe aclarar que para la implantación de la tecnología del RAID se requiere un presupuesto mayor. Por este motivo es una tarea importante de las instituciones analizar y evaluar, en función de sus recursos financieros y necesidades, la tecnología a utilizar.

7.4. Estrategias para la preservación digital

Para garantizar la preservación digital en el corto, mediano y largo plazos, dependiendo del tiempo de vida de un documento, existen diferentes estrategias que pueden aplicarse. A continuación se describen algunas de ellas.

Preservación de la tecnología. Para visualizar y editar un contenido digital es recomendable preservar el ambiente tecnológico, incluidos el software y el hardware: sistemas operativos, programas de visualización y periféricos de lectura y escritura de medios de almacenamiento secundario, entre otros.

Migración. La migración es la transferencia o adaptación del contenido digital de una generación de hardware y software a otra, superando la obsolescencia tecnológica, aunque se tiene la desventaja de sufrir pérdidas en la información tras migraciones sucesivas (figura 29).



FIGURA 29. Migración de una generación a otra de hardware y software

Reformateo. El reformateo se refiere a cambiar el contenido digital de un formato a otro, como se ilustra en la figura 30.



FIGURA 30. Reformateo de un contenido digital

Refrescado o rejuvenecimiento. Refrescar un contenido digital significa copiarlo de un medio de almacenamiento a otro nuevo del mismo tipo o bien, escribir, cada determinado tiempo, un contenido digital en un medio nuevo, para evitar que el contenido se pierda por la degradación natural que conlleva el transcurso del tiempo (figura 31).



FIGURA 31. *Refrescado o rejuvenecimiento de un contenido digital*

Emulación. La emulación permite que el software original sea usado sin la necesidad de que siga existiendo el sistema original que lo ejecutaba, lo que obliga a preservar una cantidad importante de información. Una solución de emulación por hardware, por ejemplo, implica la preservación del emulador, el sistema operativo, la aplicación y los datos.

Un ejemplo de emulación es la ejecución de un sistema operativo dentro de otro (obsérvese la figura 32).



FIGURA 32. Emulación de software (ejecución de Linux dentro de Windows)

Replicación. La replicación es la generación y el mantenimiento de una o más copias de un mismo contenido digital, como puede apreciarse en las figuras 33 y 34.



FIGURA 33. Replicación de un contenido digital

Replicación geográfica

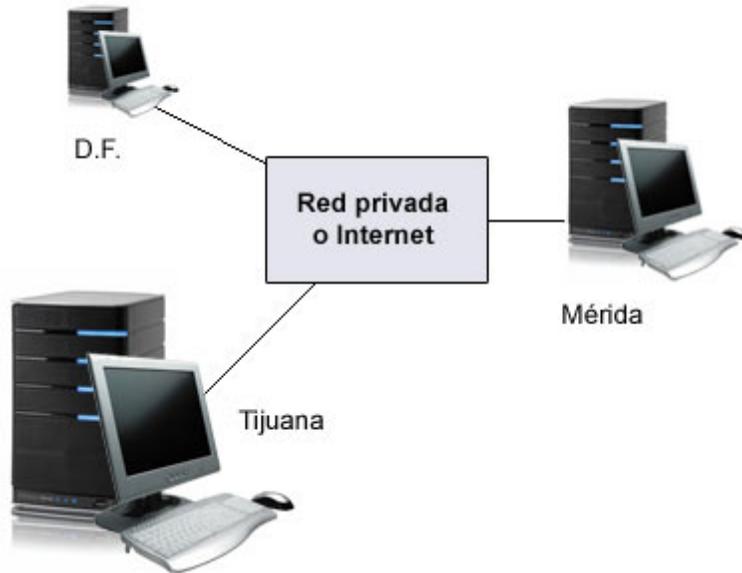


FIGURA 34. *Replicación de grandes contenidos digitales*

Estandarización. La estandarización se refiere a la utilización de un formato estándar para la representación de un documento digital, lo que garantiza un mejor soporte de herramientas para administrar la colección digital, una mayor duración del formato y una mejor migración ante los cambios tecnológicos (véase figura 35).



FIGURA 35. *Formatos estándares para representar documentos digitales*

Encapsulado. El encapsulado es el empaquetamiento de la información que se desea preservar, junto con un diccionario de datos o metadatos descriptivos, mantenidos en una única localización. Además, incorpora otros factores clave para la preservación de larga duración: a) auto documentación o la capacidad de entender y decodificar la información preservada sin hacer referencia a información externa; b) auto suficiencia o minimización de dependencias con respecto a sistemas, datos o información; c) documentación de contenido o

habilidad para que un futuro usuario encuentre o implante el software para visualizar la información preservada, y d) *preservación de organización* o habilidad para almacenar la información que permita a una organización el uso eficiente de la información preservada (véase figura 36).

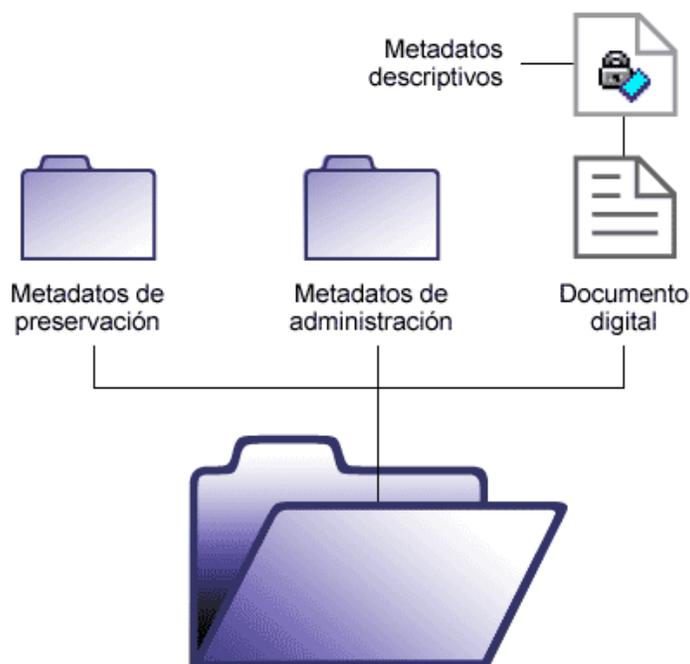


FIGURA 36. Paquete de información

Autenticidad. La autenticidad se refiere al aseguramiento de la integridad de una información digital. Existen muchas causas por las cuales se puede corromper: virus, negligencias, fallas de los medios de almacenamiento, ataques informáticos, etcétera. Para asegurar la autenticidad se propone utilizar huellas, firmas y certificados digitales sobre la información digitalizada.

Ejemplo: La función matemática **H** genera para el documento digital **D** una huella digital $h(d)=879d8a206e718d8e651a0df1e42ab7007f412a82$. La huella digital es única para cada documento, lo que quiere decir que si dos documentos tienen la misma huella digital, entonces se trata del mismo documento. El proceso de la firma digital para ofrecer autenticidad, es similar al de la firma autógrafa. Un certificado digital es un documento electrónico que demuestra identidad en transacciones electrónicas, validando que una firma digital pertenezca a una

entidad identificada. Una autoridad certificadora es el equivalente a un notario (véase figura 37).



FIGURA 37. Autenticidad de la información digital

Arqueología digital. La arqueología digital es un proceso para la recuperación de información, a partir de medios dañados o antiguos de almacenamiento digital (véase figura 38).

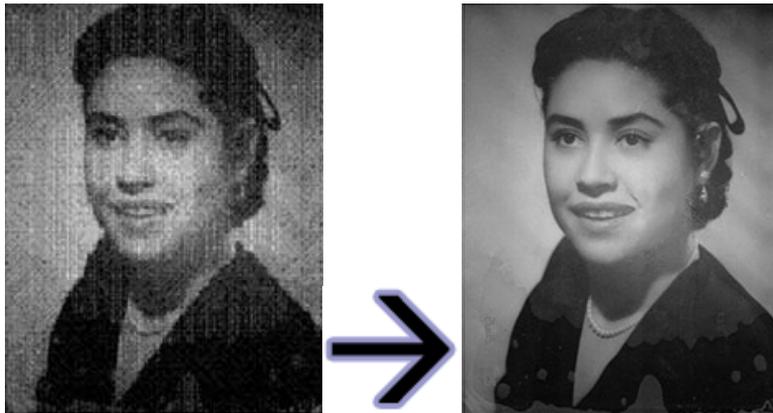


FIGURA 38. Restauración de imágenes

Cuidado duradero. El cuidado duradero debe ser visto como una estrategia continua para asegurar que los documentos digitales se encuentren en óptimas condiciones. En el cuidado de una colección los archivos deben almacenarse en medios y ubicaciones no sólo seguros, sino también confiables. Además, deben

manipularse con base en las pautas de aceptación internacional, orientadas a optimizar su expectativa y la calidad de duración.

7.5. Modelo de referencia OAIS

El modelo de referencia OAIS (Open Archival Information System) está enfocado a la preservación a largo plazo de la información en formato digital como una manera de garantizar el acceso a ella en el futuro. Consiste básicamente en un modelo lógico sobre la forma como los documentos digitales deben ser preparados, enviados a un archivo, almacenados durante periodos largos, conservados y recuperados.³¹

El modelo de referencia OAIS se ha convertido en el más reconocido para la preservación de información digital. Identifica las responsabilidades y los componentes de un sistema para archivar documentos digitales, incluyendo:

- Las funciones de las personas y las instituciones que interactúan con un documento digital: productor, administrador y consumidor.
- Los objetos digitales o documentos manejados por OAIS, denominados *paquetes de información*, y
- Seis funciones de alto nivel del modelo: Ingesta, Administración de datos, Almacén de archivos, Acceso, Planeación de la preservación y Administración, que representan treinta y tres funciones de nivel bajo.

7.5.1. Entidades de OAIS

El diagrama OAIS (véase figura 39) ilustra las relaciones entre las funciones. En él los rectángulos identifican grupos de funciones relacionadas. En la práctica no es necesario que las funciones estén en el mismo servidor o en la misma organización. Los grupos pueden estar separados y sus funciones distribuidas en muchas configuraciones.

³¹ Silió, Teresa (2005). *Fundamentos tecnológicos del acceso abierto: Open Archives Initiative y Open Archival Information System*. En: El profesional de la información. (septiembre-octubre 2005, Vol. 14, Núm. 5). Disponible en: <http://www.elprofesionalde lainformacion.com/contenidos/2005/septiembre/8.pdf> [Consultado en octubre 2007].

En el exterior de OAIS se encuentran los productores, los consumidores y los administradores:

- *Productor*. Es la persona que proporciona al sistema la información que va a ser preservada.
- *Administrador*. Es la persona que define las políticas de administración y el control de la administración de OAIS sólo una de sus responsabilidades. No está involucrado en las operaciones diarias del archivo, pues éstas son responsabilidad de la entidad funcional *Administración*.
- *Consumidor*. Interactúa con los servicios de OAIS para encontrar y obtener la información preservada de su interés.



FIGURA 39. Entidades funcionales de OAIS

El modelo de referencia OAIS está compuesto por seis entidades funcionales y sus interfaces relacionadas. En la figura 39, que muestra el modelo, sólo se presentan los flujos de información más importantes. Las líneas que conectan las entidades, identifican las rutas de información, sobre las cuales ésta fluye en ambas direcciones. Las líneas discontinuas se utilizan para evitar confusión.

Digitalización de Colecciones. Texto e imagen. Volumen 1

Así, el modelo de referencia OAIS, está conformado por:

Ingesta. Esta entidad proporciona los servicios y las funciones para aceptar los Paquetes de Información Sometida (PISs) de los productores o los elementos internos bajo el control de la Administración. Además, prepara el contenido para el manejo y almacenamiento en el archivo.

Las funciones de Ingesta incluyen:

- a) La recepción de PISs. aseguran su calidad y generan el Paquete de Información de Archivado (PIA); se encargan también de que cumpla con los estándares de documentación y el formateo de datos.
- b) Extracción de información descriptiva de los PIAs para su inclusión en la base de datos del archivo.
- c) Coordinación de actualizaciones en Almacén del archivo y Administración de datos.

Almacén del archivo. Proporciona los servicios y las funciones para el almacenamiento, mantenimiento y recuperación de PIAs. Sus funciones incluyen:

- a) La recepción de PIAs de Ingesta.
- b) Agregado de PIAs para el almacenamiento permanente.
- c) Administración de una jerarquía de almacenamiento.
- d) Actualización de los medios sobre los cuales los contenedores de los archivos son almacenados.
- e) Verificación de errores, brindando capacidades para la recuperación de desastres.
- f) Proporcionar PIAs para satisfacer las órdenes generadas por los consumidores.

Administración de datos. Brinda los servicios y las funciones para poblar, mantener y acceder a la información descriptiva, la cual identifica y documenta contenedores de archivos y datos administrativos para el manejo de un archivo.

En sus funciones se incluyen:

- a) Administración de la base de datos del archivo, con lo que mantiene las definiciones del esquema, así como vistas e integridad referencial.
- b) Ejecución de actualizaciones de la base de datos y carga de información descriptiva nueva o datos administrativos del archivo.
- c) Ejecución de consultas sobre datos para la administración de los mismos.
- d) Generación de conjuntos de resultados.
- e) Generación de reportes.

Administración. Esta entidad proporciona los servicios y las funciones para la operación global del sistema de archivo. Las funciones de administración incluyen:

- a) La solicitud y negociación de los acuerdos de sometimiento con los productores.
- b) La auditoría de los sometimientos, para asegurar que cumplan con los estándares de archivo.
- c) Mantenimiento de la administración de la configuración del software y el hardware del sistema.
- d) Proporcionar funciones de ingeniería del sistema para el monitoreo y el mejoramiento de las operaciones del archivo, inventario, reportes y migración/actualización del contenido de un archivo.

Finalmente, la Administración es la responsable de establecer y mantener las políticas y los estándares del archivo, brindando soporte a los usuarios y habilitando las solicitudes almacenadas.

Planeación de preservación. Además de proporcionar los servicios y las funciones para el monitoreo del ambiente de OAIS, esta entidad brinda recomendaciones para asegurar que la información almacenada en el sistema de archivado (es decir, el sistema que permite archivar los documentos digitales) permanezca disponible para la comunidad de

usuarios durante un tiempo muy prolongado, incluso si el ambiente original de computación se vuelve obsoleto.

Las funciones de esta entidad abarcan:

- a) Evaluación del contenido y recomendaciones periódicas de actualización de información de un archivo para migrar los contenedores actuales de los archivos.
- b) Emisión de recomendaciones sobre políticas y estándares de archivo.
- c) Monitoreo de cambios en el ambiente tecnológico y en los requerimientos de servicios de los usuarios
- d) Constitución de una base de conocimientos de la comunidad de usuarios.

En la planeación de la preservación también son diseñados modelos de paquetes de información que brindan asistencia y revisión del diseño para especializar estos modelos en PISs y PIAs y para sometimientos específicos. Por otro lado, se desarrollan planes de migración detallada, prototipos de software y planes de pruebas para liberar implantaciones de los objetivos de migración de Administración.

Acceso. Cuenta con los servicios y funciones de soporte a los consumidores en la obtención de la existencia, descripción, localización y disponibilidad de información almacenada en el sistema de archivo, permitiendo a los consumidores solicitar y recibir documentos. Las funciones de acceso incluyen:

- a) Comunicación con los consumidores para recibir solicitudes aplicando controles que limitan el acceso a la información protegida.
- b) Coordinación de la ejecución de solicitudes para que se completen satisfactoriamente.
- c) Generación de respuestas del estilo Paquetes de Información Diseminada (PIDs).
- d) Generación de resultados y reportes para los consumidores.

7.6. Esquemas y estándares de metadatos

Los metadatos de preservación son un conjunto de datos estructurados que permiten codificar, como parte del mismo documento digital, información relacionada con su preservación, es decir, en qué formato se generó, con qué compresión, calidad, etcétera. Asimismo, a través de ellos es posible identificar, describir, clasificar y localizar los documentos digitales que se preservarán.

A continuación se enlistan algunos proyectos e iniciativas en los que se han desarrollado estándares de metadatos para la preservación digital:

- RGL/ OCL Working Group on Preservation Metadata
<http://www.rlg.org./preserv/presmeta.html>
- NEDLIB (Networked European Deposit Library)
http://nedlib.kb.nl/results/D4.2/D4.2.htm#_Toc494249797
- Metadata for Digital Preservation : the Cedars Project Outline Specification
<http://www.leeds.ac.uk/cedars/colman/metadata/metadataspec.html>
- National Library of Australia, Preservation Metadata for Digital Collections
<http://www.nla.gov.au/preserve/pmeta.html>

7.7. Arquitectura tecnológica

La figura 40 muestra la arquitectura del servidor de publicación. Con el usuario general se representa la comunidad que puede consultar los documentos digitales que se ofrecen a través del servidor de publicación. El administrador de la colección es quien valida los documentos digitales que se someten a depósito en el servidor de publicación. Asimismo, aprueba los documentos digitales que serán depositados en el servidor de preservación. Otra de sus funciones es realizar el depósito de los documentos digitales de publicación y preservación en el servidor de preservación.



FIGURA 40. Arquitectura simple del servidor de publicación

Para el manejo de la preservación digital es necesario un requerimiento mínimo, el cual se muestra en la figura 41. En el resguardo de los medios de almacenamiento secundario, como es el caso de los documentos digitales preservados, es imprescindible contar con una bóveda. Si lo permiten los recursos, para ofrecer una mayor disponibilidad, además de mejores garantías de preservación ante fallas y desastres naturales, es ineludible tener una réplica de la bóveda localizada geográficamente.

Debe haber por lo menos un administrador local o curador, que mantenga organizados los medios de almacenamiento en la bóveda y tenga la autoridad para incorporarlos o retirarlos. Su obligación será cumplir las políticas y los procedimientos durante la administración de los documentos digitales publicados y preservados. Cuando le sea solicitado, el Administrador local podrá depositar uno o más documentos digitales publicados en el servidor de preservación. El acceso al servidor de preservación será posible únicamente a través del servidor de publicación y sólo cuando no se pueda recuperar por medio de su respaldo un documento digital publicado.



FIGURA 41. Arquitectura simple del servidor de preservación

Para garantizar un acceso continuo puede diseñarse una arquitectura más compleja (ver figura 42), pero esto dependerá de los recursos que se tengan destinados a esta actividad. Otra de las actividades del Administrador local será la generación de los medios de almacenamiento secundario de los documentos digitales publicados y preservados que hayan sido depositados en el servidor de preservación por el administrador de la colección, el servidor de validación o la unidad de digitalización.



FIGURA 42. Arquitectura de complejidad media de servidores

Para los servidores de publicación y preservación se requieren generalmente equipos de gran capacidad, alta disponibilidad, escalabilidad, adaptabilidad y bajo costo.

Es necesaria una gran capacidad para:

Digitalización de Colecciones. Texto e imagen. Volumen 1

- Brindar el servicio.
- Detectar y eliminar cuellos de botella.
- Tener velocidad de respuesta en los equipos y
- Favorecer la comunicación de la red.

Una alta disponibilidad es deseable para:

- Ofrecer un servicio continuo.
- Tener interrupciones breves.
- Operar con equipos o componentes redundantes (es decir, duplicados, replicados o repetidos).
- Tolerar fallas.
- Asegurar una recuperación automática frente a un problema.

Las características de los equipos, relacionadas con la escalabilidad y la adaptabilidad, les permitirá crecer conforme a la demanda.

7.8. Políticas y procedimientos

Dentro del contexto relacionado con el respaldo y la preservación, las políticas y los procedimientos son un conjunto de métodos que, aplicados sistemáticamente, sirven de apoyo en la realización del respaldo, el resguardo, la recuperación y la preservación de un contenido digital. Cada institución determina su propio conjunto de políticas y procedimientos aplicables sólo dentro de ella. Las siguientes son algunas políticas y procedimientos generales para el respaldo y la preservación de documentos digitales, aplicables en cualquier proyecto de digitalización.

7.8.1. Políticas y procedimientos de respaldo y preservación

1. Manejar con mucho cuidado los medios de almacenamiento.
2. Cumplir con las especificaciones del fabricante para el cuidado de los medios de almacenamiento, como las condiciones climáticas: humedad, calor, polvo, etcétera.

3. Ordenar los medios de almacenamiento en forma vertical.
4. No colocar objetos sobre los medios de almacenamiento.
5. Verificar la integridad del contenido almacenado en el dispositivo de almacenamiento secundario cada vez que se realice una copia de la información.
6. Verificar periódicamente el funcionamiento correcto del dispositivo periférico para la generación de copias de los datos.
7. Establecer reglas y procedimientos para la integración de metadatos.
8. Validar que los documentos digitales a ingresar se encuentren en un formato estándar.

7.8.2. Políticas y procedimientos de respaldo

1. Los respaldos deben hacerse en el horario de menor uso del servidor de publicación.
2. Se recomienda tener una copia del contenido digital cerca del servidor de publicación y otra lejos.
3. Retirar el medio de almacenamiento secundario de la unidad de lectura y grabación cuando haya concluido el proceso de respaldo.
4. Cumplir con los periodos de respaldo indicados en el plan de seguridad y contingencia.

7.8.3. Políticas y procedimientos de preservación

1. El acceso a la bóveda debe restringirse a un número limitado y bien definido de personas.
2. La consulta del servidor de preservación puede realizarse sólo a través del servidor de publicación de documentos digitales.
3. Refrescar los medios una vez al año.
4. Evitar el uso de los *masters* de preservación.
5. Por cada *master* de preservación generar un mínimo de dos copias.
6. Por cada *master* de publicación generar un mínimo de dos copias.
7. Asignar un límite de vida a cada documento electrónico.

8. Verificar semestralmente los cambios tecnológicos en los formatos de almacenamiento.
9. Verificar semestralmente los cambios tecnológicos de software y hardware, que impacten en la obsolescencia de los programas y los equipos en uso.
10. Verificar semestralmente la implantación de estándares nuevos en la representación y el intercambio de información digital, así como en los metadatos descriptivos y de preservación.
11. Cada vez que lo sugieran los cambios tecnológicos o los estándares nuevos, realizar la migración o el reformateo de los documentos digitales con el apoyo de programas computacionales, preferentemente.
12. Contar con un mínimo de dos bóvedas replicadas geográficamente para garantizar la preservación de los medios de almacenamiento ante desastres naturales.

7.9. Recomendaciones

Por sí solas las soluciones técnicas no son suficientes para asegurar la duración prolongada de los documentos digitales. Para lograr soluciones plenas y satisfactorias, se requiere la integración de aspectos técnicos y administrativos: recursos humanos, capacitación, requisitos financieros, criterios de selección, metadatos de preservación, etcétera.

Para una administración efectiva de las colecciones digitales, se debe desarrollar y seguir un plan de gestión en los proyectos de digitalización que permita evaluar los requisitos de preservación y el acceso a largo plazo. Simultáneamente, deben ser identificados los costos y los beneficios, además de calcularse los riesgos.

7.10. Para saber más...

- [NEDLIB](#) (Networked European Deposit Library)
- [Cedars](#) (curl exemplars in digital archives)
- [CAMILEON](#) (Creative Archiving at Michigan and Leeds Emulation the Old On the New)

- [DPC](#) (Digital Preservation Coalition)
- [PANDORA](#) (Preserving and Accessing Networked Documentary Resources of Australia)
- [NDIIPP](#) (The National Digital Information Infrastructure and Preservation Program) Library of Congress
- [PADI](#) (Preserving Access to Digital Information) National Library of Australia

7.11. Fuentes consultadas

- Beagrie, N., Jones, M., Digital Preservation Coalition. *Preservation Management of Digital Materials: A Handbook*. En: <http://www.dpconline.org/graphics/handbook/> [Consultado en diciembre 2007]
- Bia, Alejandro, Sánchez, Manuel (2002). *Desarrollo de una política de preservación digital: tecnología, planificación y perseverancia*. Disponible en: <http://www.cervantesvirtual.com/research/articles/JBIDI02a.pdf> [Consultado en octubre 2007].
- California Digital Library. *Digital preservation program*. Disponible en: http://www.cdlib.org/programs/digital_preservation.html [Consulta: octubre 2007].
- *Camileon*. Disponible en: <http://www.si.umich.edu/CAMILEON/> [Consulta: octubre 2007].
- *Curl exemplars in digital archives* (Cedars). Disponible en: <http://www.leeds.ac.uk/cedars/> [Consultado en octubre 2007].
- *Digital Curation Centre (DCC)*. (2007). Disponible en: <http://www.dcc.ac.uk/> [Consultado en octubre 2007].
- *Digital Preservation at EUL*. Disponible en: <http://www.lib.ed.ac.uk/sites/digpres/index.shtml> [Consultado en octubre 2007].
- *Digital Preservation Coalition (DPC)*. Disponible en: <http://www.dpconline.org/> [Consulta: octubre 2007].
- Hernández, Ignacio José (2005). *Métodos y políticas de respaldo (backup) en planes de contingencia*. Universidad politécnica de Madrid. Disponible en: www.criptored.upm.es/quiateoria/gt_m001l.htm. [Consultado en octubre 2007].
- *Interpares Project. International Research on Permanent Authentic Records in Electronic Systems*. Disponible en: <http://www.interpares.org/>. [Consultado en octubre 2007].
- Keefer, Alice y Gallart, Núria (2003). *La preservación digital y las universidades: el estado de la cuestión*. En: *Proceedings de las 8as. Jornadas españolas de documentación*. Barcelona. pp. 75-83.
- Lamarca, María Jesús (2008). *Metadatos*. En: *Hipertexto: el nuevo concepto de documento en la cultura de la imagen*. Disponible en: <http://www.hipertexto.info/documentos/metadatos.htm> [Consultado en octubre 2007].

Digitalización de Colecciones. Texto e imagen. Volumen 1

- *Llevando la teoría a la práctica: tutorial de digitalización de imágenes*. Disponible en: <http://www.library.cornell.edu/preservation/tutorial-spanish/toc.html> [Consultado en octubre 2007].
- McGray, A. T., Gallagher M. E. (2001). *Principles for Digital Libraries Development*. En: Communications of the ACM, 44, p. 49-54.
- *Networked European Deposit Library (NEDLIB)*. Disponible en: <http://nedlib.kb.nl/> [Consultado en octubre 2007].
- OCLC (2005). *PREMIS (PREservation Metadata: Implementation Strategies) Working Group*. Disponible en: <http://www.oclc.org/research/projects/pmwg/> [Consultado en octubre 2007].
- *Pandora: Australia's web archive*. Disponible en: <http://pandora.nla.gov.au/> [Consultado en octubre 2007].
- *Preserving access to digital information (PADI)*. Disponible en: <http://www.nla.gov.au/padi> [Consultado en octubre 2007].
- *Preserving our digital heritage. Plan for the National Digital Information Infrastructure and Preservation Program*. Disponible en: http://www.digitalpreservation.gov/library/pdf/ndiipp_plan.pdf [Consultado en octubre 2007].
- *Reference model for an open archival information system (OAIS)*. En: Consultative Committee for Space Data Systems, Recommendation for Space Data Systems Standards. Disponible en: <http://nost.gsfc.nasa.gov/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf> [Consultado en octubre 2006].
- Silió, Teresa (2005). *Fundamentos tecnológicos del acceso abierto: Open Archives Initiative y Open Archival Information System*. En: El profesional de la información (septiembre-octubre 2005, Vol. 14, Núm. 5). Disponible en: <http://www.elprofesionalde lainformacion.com/contenidos/2005/septiembre/8.pdf> [Consultado en octubre 2007].
- *The National Digital Information Infrastructure and Preservation Program (NDIIPP)*. Disponible en: <http://www.digitalpreservation.gov/> [Consultado en octubre 2007].
- Waugh Andrew, et al. (2000). *Preserving Digital Information Forever*. International Conference on Digital Libraries. Proceedings of the fifth ACM conference on Digital libraries. San Antonio, Texas. p. 175 - 184.

8. Recomendaciones y puntos críticos

En esta sección se enlistan algunos de los factores más importantes que deben considerarse al tomar decisiones sobre alguno de los cuatro rubros presentados. Se parte de la idea de que cada proyecto de digitalización define sus propias necesidades y demanda soluciones específicas ante situaciones concretas. Bajo este entendido, se exponen algunas de las circunstancias a las que se puede enfrentar usualmente un proyecto antes, durante o después de un proceso de digitalización.

8.1. *Instalaciones*

Lo concerniente a las instalaciones físicas para trabajar puede parecer obvio, sin embargo, a veces al comenzar a trabajar resulta evidente la necesidad de preparar los espacios en los que circularán los documentos y las personas que los procesarán.

Al decidir que una digitalización debe hacerse en un espacio físico determinado, se debe:

- a) Garantizar y controlar el suministro de energía eléctrica.
- b) Asegurar espacios físicos suficientes para el desplazamiento del personal y el manejo de los documentos.
- c) Acondicionar una ventilación e iluminación apropiadas a la cantidad y las dimensiones del equipo a utilizar, así como al número de personas que lo operarán.
- d) Mantener la integridad de la colección mediante condiciones adecuadas de higiene y seguridad.
- e) Proveer mobiliario adecuado para apilar, extender y separar documentos, considerando que los materiales se trabajarán en el mismo lugar donde estarán instalados escáneres, cámaras, computadoras u otros dispositivos.

El espacio físico debe contar con requerimientos mínimos para digitalizar una colección. Se recomienda un lugar definitivo para desarrollar el procesamiento, con el fin de evitar reconfiguraciones y pérdidas sorpresivas de información, así como maltratos y extravíos en una colección, entre otros factores.

8.2. Outsourcing

Se denomina de esta manera a la contratación que se hace de una compañía para que realice algún proceso, maquila, servicio, etcétera, en lugar de utilizar para ello los recursos propios. Ésta es una opción viable para la digitalización cuando se cuenta con el presupuesto suficiente y lo ameritan las condiciones y las necesidades de un proyecto. Antes de elegir esta alternativa para el traslado de una colección de formato análogo a digital, conviene poner cuidadosamente en la balanza las ventajas y las desventajas que tiene.

Vale la pena mencionar que esta modalidad de trabajo es cada vez más frecuente en ámbitos como la conversión tecnológica. Aunque la oferta es amplia, no siempre es sencillo saber si se está frente a la mejor opción del mercado, sino hasta que se comienza a trabajar con la compañía elegida. Los precios o la velocidad que ofrece una empresa no son garantía de un buen servicio, el cual, por otra parte, debe ser completo, garantizado, puntual, profesional, económicamente accesible y, en resumen, satisfactorio de acuerdo con lo encargado.

Si para contar con esta opción de servicio debe abrirse una licitación, tendrán que analizarse cuidadosamente todas las ofertas. Es muy recomendable que este examen lo haga un comité integrado por actores clave, relacionados con aspectos técnicos, organizativos y administrativos. La decisión debe tomarse siempre en función de las necesidades del proyecto y las capacidades reales del contratante.

Antes de optar por el *outsourcing*, se debe:

- a) Analizar todas las compañías que ofrecen estos servicios y definir claramente la conveniencia y la viabilidad de pagar por la realización de un trabajo.

Recomendaciones y puntos críticos

- b) Identificar a la empresa que parezca ser la idónea y conocer su experiencia, seriedad y confiabilidad.
- c) Hacer una comparación entre dos o más empresas para saber cuál ofrece las mejores tarifas, servicios y procedimientos.

Una vez que se ha elegido a una empresa para digitalizar bajo este esquema:

- a) Exponer a la compañía, de manera clara, concisa y precisa, lo que se espera del servicio, de lo general a lo particular, sin dejar nada sobreentendido u obviado.
- b) Acordar el flujo de trabajo que se seguirá.
- c) Hacerle saber todos los detalles de lo que se espera obtener, como niveles de resolución, tamaño de los archivos, calidad de las imágenes, OCR, indización, etcétera.
- d) Acordar las modalidades de entrega de los archivos: respaldos, formatos, etcétera.
- e) Saber qué tanto respaldará su trabajo la empresa, una vez entregado el producto final.
- f) Acordar cómo se solucionarán los errores detectados durante y después de la digitalización.
- g) Poner atención en lo que entrega la compañía y lo que debería entregar. Desde la óptica de ésta, atender también lo que recibe y lo que debería recibir a través del encargado o responsable del proyecto.

Finalmente, objetivos, visión y participación son elementos indispensables en la elección de la modalidad *outsourcing*.

8.3. Derechos de autor

Los derechos de autor son uno de los aspectos más delicados en la conversión de colecciones impresas o análogas a formatos digitales. Son determinantes para que una obra digital pueda o no ser reproducida o incluso dada a conocer al público en la modalidad de consulta en pantalla.

En materia de derechos de autor hay que considerar:

- a) La Ley Federal del Derecho de Autor (LFDA) que, aunque puede parecer complicada, es un excelente inicio para evitar futuras sorpresas.
- b) La existencia de los derechos para digitalizar documentos. Si hay duda sobre esto, es mejor iniciar su procesamiento hasta dejar resuelto el tema.
- c) No hay una “receta” sobre lo que se puede digitalizar y lo que no. Por ello, en el caso especial de las colecciones contemporáneas, es oportuno contar con la asesoría legal de los expertos en el tema de los derechos de autor, al menos en la fase de selección de los materiales que se pretende procesar.
- d) Si las obras no son propias, deben asegurarse por escrito los permisos de digitalización.
- e) No deben omitirse las autorías ni la referencia a los originales como fuente.
- f) Prever en los costos del proyecto un posible pago de regalías.
- g) Si los documentos digitalizados se alojarán en línea, deben mencionarse siempre las libertades o las restricciones de uso que tienen los usuarios que los consultan.
- h) Considerar los registros nacionales e internacionales necesarios para proteger los documentos producto de la digitalización.

8.4. Colecciones en línea

Una tendencia en la conversión de colecciones análogas a digitales apunta hacia el alojamiento en línea (es decir, en Internet) de las colecciones digitales que se están creando. Es una práctica común cuando no hay restricciones legales de difusión. En este caso la consulta o el acceso a una colección, puede ser para el público en general o una comunidad específica, identificada y controlada.

Cuando una colección digitalizada se pone a disposición de una comunidad de usuarios, vía acceso remoto en línea, es necesario observar:

- a) Las facilidades técnicas para el hospedaje y la consulta.
- b) El tamaño de los archivos para su consulta, descarga e impresión, desde un ancho de banda comercial promedio.

Recomendaciones y puntos críticos

- c) Los permisos o derechos de autor para difundir y permitir acciones como impresión, copia, modificación, descarga, etcétera, de archivos.
- d) El formato o formatos de consulta-despliegue.
- e) Las características de lectura y navegación en pantalla.
- f) Hospedaje, acceso, preservación, respaldos y todo lo relacionado con la administración-consulta de los archivos en línea.

Un factor muy importante, no incluido en la lista, es el ordenamiento de una colección. Éste es competencia de la organización bibliográfica de los acervos, que regularmente se aplica a las colecciones pertenecientes a bibliotecas, archivos, centros de documentación e información, etcétera. Sin embargo, la ventaja de tener colecciones catalogadas y clasificadas, es decir, organizadas sistemáticamente, radica en tener los datos de identificación de un documento. Éstos deben ser legibles a través de una computadora y por lo tanto, rescatables a través de sistemas de recuperación automatizados, locales o remotos, accesibles vía Internet.

9. Glosario

Acervo: conjunto de fuentes de información, por ejemplo, libros, revistas, fotografías, mapas, etc.

Acceso local: consulta de una fuente de información sin estar conectado a una red de cómputo

Acceso remoto: consulta de una fuente de información por medio de una conexión como Internet o similar.

Administración de información: los procedimientos involucrados en la identificación, selección, organización, transmisión, control, recuperación y preservación de la información.

Ancho de banda: cantidad de información o de datos que se puede enviar a través de una conexión de red en un período de tiempo dado. El ancho de banda se indica generalmente en bites por segundo (BPS), kilobites por segundo (kbps), o megabites por segundo (mps).

Arqueología digital: proceso de recuperación de información a partir de medios de almacenamiento digital dañados o antiguos.

Biblioteca: organismo o entidad que pone a disposición de determinado público fuentes de información en diferentes formatos (libros, revistas, fuentes digitales, etc.)

Certificado digital: documento electrónico con el que se demuestra la identidad de una persona o entidad en transacciones electrónicas. Con él se valida que una firma digital pertenece a una entidad identificada.

Cluster. Conjunto de computadoras o servidores compuestos, conectados y configurados para funcionar de manera integral y única, lo que agiliza la respuesta de procesamiento y aumenta la capacidad de almacenamiento de la información.

Colección de consulta: se compone de documentos de referencia o consulta, es decir: diccionarios, enciclopedias, atlas, anuarios, etc.

Digitalización de Colecciones. Texto e imagen. Volumen 1

Colección digital: conjunto de documentos digitales de información, como libros, revistas, tesis, etc.

Compatibilidad: en ambientes digitales, la capacidad que tiene una tecnología de interactuar con o adaptarse a otra. Por ejemplo, un archivo digital puede ejecutarse en uno o más sistemas operativos.

Compresión: Disminución del tamaño de un documento en formato digital con la intención de transferir o almacenar la misma información empleando la menor cantidad de espacio y disminuir el tiempo en la transferencia de datos. Una compresión puede ser con pérdida de información/calidad (generalmente para las imágenes y sonidos), o sin pérdida de información (para archivos o información que no debe ser degradada, como documentos de texto).

Comunidad de usuarios: el universo de personas a quienes está destinado un servicio.

Copia de seguridad: también conocida como respaldo, es la copia de los datos o información de un sistema para que lo puedan restaurar en caso fallas o desastres. Una copia de seguridad es utilizada para recuperar datos o información que se hayan borrado o corrompido por cualquier causa.

Copia integral: también conocida como copia completa. Consiste en copiar datos y programas, restaurando el sistema al momento anterior a la copia.

Copia incremental: almacena solamente las modificaciones realizadas después de la última copia de seguridad. Se requiere mantener la copia original para poder restaurar posteriormente el resto de las copias.

Copia rotativa: copia de seguridad o respaldo realizada con la reutilización periódica de los medios de almacenamiento.

Derechos de autor: conjunto de privilegios que tiene una persona sobre una obra de creación intelectual.

Digitalización: proceso de convertir un documento análogo a un formato digital legible por computadora.

Documento: texto, libro, revista, tesis o cualquier recurso que transmita información, datos o conocimiento. Un documento puede ser físico (análogo) o digital.

Documento de origen digital: documento que de origen fue creado en ambientes digitales, generalmente por medio de una computadora con ayuda de un procesador de textos o herramienta similar.

Documento digital: documento cuya información se encuentra codificada en formato digital.

Documento digitalizado: documento que ha sido creado a partir de un original analógico, generalmente impreso, por medio de alguna herramienta de captura digital como escáner, cámara, etc.

Emulación: estrategia de preservación que permite que el software original sea usado sin necesidad de que el sistema original que lo ejecutaba siga existiendo.

Encapsulado: estrategia de preservación digital que consiste en agrupar la información que se desea preservar junto con un diccionario de datos (metadatos descriptivos, administrativos y de preservación) y mantenerlos en una única localización.

E/S. También conocido como I/O (por su siglas en inglés Input/Output) es la abreviación de Entrada/Salida que hace referencia la recepción y envío de información por medio de dispositivos o puertos de comunicación de los equipos cómputo y sus periféricos.

Estandarización: normalización de procedimientos involucrados en una actividad.

Extensión: nombre del archivo electrónico que indica la forma en que está codificada la información que contiene. Suele escribirse abreviada con tres caracteres, y precedida de un punto. Por ejemplo: .doc, .pdf, .mp3, etc.

Fibre Channel. El Canal de Fibra es una tecnología utilizada principalmente en las redes de almacenamiento ofreciendo velocidad de transferencia de 1, 2, 4 y 8 Gb/s.

Digitalización de Colecciones. Texto e imagen. Volumen 1

Firma digital: análogo a la firma autógrafa para ofrecer autenticidad en medios digitales.

Formato digital: la forma en que está codificada la información contenida en un documento digital. Existen diferentes tipos de formatos para diferentes tipos de documentos: textuales, imágenes, audio, video, etc.

Formatos de uso frecuente en texto e imágenes: .doc; .pdf; .txt; gif; .jpg.

Huella digital: número hexadecimal generado al aplicar una función matemática sobre un documento digital con la propiedad de que si dos documentos digitales tienen la misma huella digital entonces tienen el mismo contenido.

Infraestructura tecnológica: equipos de cómputo, conexión, redes, cableado y servicios de telecomunicaciones con los que se cuenta para proporcionar servicios de información en ambientes digitales.

Internet: conjunto de redes de cómputo interconectadas mundialmente.

Intranet: es una red de computadoras privada o interna de tipo empresarial o educativa que utiliza las mismas tecnologías de Internet.

Marca de agua: sello digital, generalmente tenue, que se inserta a una imagen digital con la finalidad de mostrar la autoría de su creador, difusor, responsable, etc..

Master: versión íntegra de una obra. Generalmente conserva todas las características técnicas con las que fue creada. Suele emplearse como punto de partida para la creación de ulteriores copias.

Metadatos: Un metadato es un dato estructurado sobre la información, es decir, información sobre información, o de forma más simple, datos sobre datos. Son datos que se pueden guardar, intercambiar y procesar por medio de la computadora. Están estructurados de tal forma que permiten ayudar a la identificación, descripción clasificación y localización del contenido de un documento. También sirven para su recuperación.

Migración: transferencia o adaptación del contenido digital de una generación de hardware y software hacia otra generación.

OAIS (Open Archival Information System) Es un modelo de referencia que ha sido publicado como una recomendación del *CCSDS (Consultative Committee for Space Data Systems)* y como norma *ISO14721:2003*. Enfoca su actividad en la preservación a largo plazo de la información en formato digital, como garantía de que será accesible en el futuro. *OAIS* consiste en un modelo lógico que abarca todas las funciones de un repositorio digital, señalando la forma en que los objetos digitales deben ser preparados, enviados a un archivo, almacenados durante largos períodos, conservados y recuperados.

Parámetro normalizado: procedimiento o rango de acción normalizado para ejecutar una actividad. Por ejemplo, el tamaño mínimo que deben tener las imágenes para verse en un monitor de computadora.

Plan de contingencia: descripción de un conjunto de actividades para poder asegurar o restablecer la continuidad en el funcionamiento de un sistema informático ante las fallas de hardware, software o corrupción de información.

Planeación: conjunto de pasos razonados destinados a la ejecución de un proyecto.

Políticas: conjunto de acuerdos establecidos formalmente y dados a conocer entre los involucrados en la realización de un proyecto.

Preservación de la tecnología: Consiste en preservar el ambiente tecnológico para visualizar y editar el contenido digital, incluyendo software y hardware, como por ejemplo: sistemas operativos, programas de visualización, periféricos de lectura y escritura de medios de almacenamiento secundario.

Preservación digital: Serie de actividades necesarias y muy bien administradas para asegurar el acceso continuo a los materiales digitales por el periodo que sea necesario. Suele entenderse también como la digitalización de documentos originales con el fin de protegerlos de los efectos nocivos de la manipulación física.

RAID: son las siglas en inglés de *Redundant Array of Inexpensive Disks*, ‘conjunto redundante de discos baratos’. En la actualidad también de *Redundant Array of Independent Disks*, ‘conjunto redundante de discos independientes’

Recursos humanos: personal con el que se cuenta para desarrollar un proyecto.

Recursos tecnológicos: equipo técnico, aparatos, etc., con los que se cuenta para llevar a cabo un proyecto.

Reformateo: cambio del contenido digital de un formato a otro.

Refrescado (rejuvenecimiento): copia del contenido a un medio nuevo.

Replicación: se refiere al hecho de mantener una o mas copias de un mismo contenido digital.

Reproducción: generación de una copia de un documento en el mismo o en otro formato.

Resolución de una imagen: nivel de detalle que muestra una imagen digital. A mayor resolución, mayores detalles muestra la imagen.

Respaldo: También conocido como copia de seguridad, copia de los datos o información de un sistema para que lo puedan restaurar en caso de fallas o desastres. Un respaldo es utilizado para recuperar datos o información que se haya borrado o corrompido por cualquier causa.

Ruta crítica: pasos que se siguen en un procedimiento de inicio a fin. Suelen señalarse algunos aspectos relevantes a tener en cuenta al momento de llevarlo a cabo.

Seguridad: medidas precautorias que se toman con la intención de garantizar la integridad física, técnica e intelectual de los documentos digitales.

Tecnología digital: tecnología basada en ambientes de cómputo, dentro de los cuales se crean y difunden productos y creaciones intelectuales.

Unidad de información: espacio o entidad en la que se administra y difunde información, por ejemplo, una biblioteca, un centro de información, una hemeroteca, etc.

10. Fuentes consultadas

- Association of Research Libraries. *Managing digitization activities*. Disponible en: <http://www.arl.org/bm~doc/spec294web.pdf> [Consultado en noviembre 2006].
- Beagrie, N., Jones, M., Digital Preservation Coalition. *Preservation Management of Digital Materials: A Handbook*. En: <http://www.dpconline.org/graphics/handbook/> [Consultado en diciembre 2007]
- Bia, Alejandro, Sánchez, Manuel (2002). *Desarrollo de una política de preservación digital: tecnología, planificación y perseverancia*. Disponible en: <http://www.cervantesvirtual.com/research/articles/JBIDI02a.pdf> [Consultado en octubre 2007].
- Biblioteca de la Universidad de Cornell. *Tutorial de digitalización de imágenes*. Disponible en: <http://www.library.cornell.edu/preservation/tutorial-spanish/contents.html> [Consultado en enero 2008].
- California Digital Library. *Digital preservation program*. Disponible en: http://www.cdlib.org/programs/digital_preservation.html [Consultado en octubre 2007].
- *Camileon*. Disponible en: <http://www.si.umich.edu/CAMILEON/> [Consultado en octubre 2007].
- *Copia de seguridad*. En: Wikipedia, la enciclopedia libre. Disponible en: http://es.wikipedia.org/wiki/Copia_de_seguridad [Consultado en abril 2006].
- Cornell University Library. *Moving theory into practice. Digital imaging tutorial*. Disponible en: <http://www.library.cornell.edu/preservation/tutorial/technical/technicalC-01.html> [Consultado en agosto 2007].
- Cory Snaveley. *Handle Service Conventions and Practices*. Disponible en: <http://www.umdl.umich.edu/um-rlg.html> [Consultado en diciembre 2007]
- *Curl exemplars in digital archives* (Cedars). Disponible en: <http://www.leeds.ac.uk/cedars/> [Consultado en octubre 2007].
- *Digital Consultancy Services*. Disponible en: <http://www.digitalconsultancy.net/> [Consultado en octubre 2007].
- *Digital Curation Centre (DCC)*. (2007). Disponible en: <http://www.dcc.ac.uk/> [Consultado en octubre 2007].
- Digital Library Federation. *Digital library standards and practices*. Disponible en: <http://www.diglib.org/standards.htm>. [Consultado en enero 2008].
- *Digital Preservation at EUL*. Disponible en: <http://www.lib.ed.ac.uk/sites/digpres/index.shtml> [Consultado en octubre 2007].
- *Digital Preservation Coalition (DPC)*. Disponible en: <http://www.dpconline.org/> [Consultado en octubre 2007].
- *Familia Centera de EMC*. En: EMC²: where information lives. Disponible en: <http://spain.emc.com/products/systems/centera.jsp?openfolder=platform> [Consultado en abril 2006]
- *File Formats Blog*. Disponible en: <http://fileformats.blogspot.com/> [Consultado en

Digitalización de Colecciones. Texto e imagen. Volumen 1

marzo 2007].

- Fleischhauer, Carl. *Digital formats for content reproductions*. The National Digital Library Program. July 1998. Disponible en: <http://lcweb2.loc.gov/ammem/formats.html> [Consultado en marzo 2007].
- Gómez, Luis Guillermo (1997). *Estaciones de trabajo*. En: Monografías.com. Disponible en: <http://www.monografias.com/trabajos41/estaciones-de-trabajo/estaciones-de-trabajo.shtml> [Consultado en septiembre 2007].
- Gómez, Félix y Werner Guillamón (2005). *La gestión documental y la norma ISO 15489:2001 Record Management*. En: Boletín de Asociación Andaluza de Bibliotecarios. Marzo 2005, No. 78. Disponible en: <http://www.aab.es/pdfs/baab78/78a2.pdf> [Consultado en septiembre 2007].
- Greenstein, Daniel y Gerald, George (2008). *Digital reproduction quality: benchmark recommendations*. Disponible en: <http://www.rlg.org/preserv/diginews/diginews5-4.html#featured>. [Consultado en enero 2008].
- Hernández, Ignacio José (2005). *Métodos y políticas de respaldo (backup) en planes de contingencia*. Universidad politécnica de Madrid. Disponible en: www.criptored.upm.es/guiateoria/gt_m001l.htm. [Consultado en octubre 2007].
- Hughes, Lorna M. (2004). *Digitizing collections: strategic issues for the information manager*. London: Facet.
- Hurst, Jill Ann (2002). *Planning for regional digitization projects : the SCRLC approach*. South Central Regional Library Council (September 2002). Disponible en: http://www.lakenet.org/scrlc/ad_comm/acits/2002digplan.pdf [Consultado en marzo 2007].
- *IEEE 1394*. En: Wikipedia, la enciclopedia libre. Disponible en: http://es.wikipedia.org/wiki/IEEE_1394 [Consultado en septiembre 2007].
- *Interpares Project. International Research on Permanent Authentic Records in Electronic Systems*. Disponible en: <http://www.interpares.org/>. [Consultado en octubre 2007].
- *Introducing the new Xserve* En: Apple Servers. Disponible en: <http://www.apple.com/xserve/> [Consultado en abril 2006].
- Keefer, Alice y Gallart, Núria (2003). *La preservación digital y las universidades: el estado de la cuestión*. En: Proceedings de las 8as. Jornadas españolas de documentación. Barcelona. pp. 75-83.
- Lamarca, María Jesús (2008). *Metadatos*. En: Hipertexto: el nuevo concepto de documento en la cultura de la imagen. Disponible en: <http://www.hipertexto.info/documentos/metadatos.htm> [Consultado en octubre 2007].
- McGray, A. T., Gallagher M. E. (2001). *Principles for Digital Libraries Development*. En: Communications of the ACM, 44, p. 49-54.
- Moraski, Brittney L. (2007). *Libraries go digital, and books go on*. Disponible en: <http://www.thecrimson.com/article.aspx?ref=519188>. [Consultado en diciembre 2007].
- *Networked European Deposit Library (NEDLIB)*. Disponible en: <http://nedlib.kb.nl/> [Consultado en octubre 2007].

Fuentes consultadas

- OCLC (2005). *PREMIS (PREservation Metadata: Implementation Strategies) Working Group*. Disponible en: <http://www.oclc.org/research/projects/pmwg/> [Consultado en octubre 2007].
- *Pandora: Australia's web archive*. Disponible en: <http://pandora.nla.gov.au/> [Consultado en octubre 2007].
- *Planning digitization projects: a brief bibliography*. Harvard University. Disponible en: <http://preserve.harvard.edu/bibliographies/digitalplanning.pdf>. [Consultado en octubre 2007]
- Perlman, Eric and Kallen Ian. *Common Internet File Formats*. Disponible en: <http://www.matisse.net/files/formats.html> [Consultado en enero 2007].
- *Preparing materials for digitization*. En: Capture your collections. A guide for managers planning and Implementing Digitization Projects. Disponible en: http://www.chin.gc.ca/English/Digital_Content/Managers_Guide/pdf.html [Consultado en enero 2008]
- *Preserving Access to Digital Information (PADI)*. Disponible en: <http://www.nla.gov.au/padi> [Consultado en octubre 2007].
- *Preserving our digital heritage. Plan for the National Digital Information Infrastructure and Preservation Program*. Disponible en: http://www.digitalpreservation.gov/library/pdf/ndiipp_plan.pdf [Consultado en octubre 2007].
- *Promise RAID and Xsan. Both built to build on* En: Apple Servers. Disponible en: <http://www.apple.com/xserve/raid/> [Consultado en abril 2006]
- Quiroga, Eduardo (2005). *El monitor: tipos y características*. En: Observatorio Tecnológico. Ministerio de Educación y Ciencia de España. Disponible en: <http://observatorio.cnice.mec.es/modules.php?op=modload&name=News&file=article&sid=268> [Consultado en septiembre 2007].
- *Reference model for an open archival information system (OAIS)*. En: Consultative Committee for Space Data Systems, Recommendation for Space Data Systems Standards. Disponible en: <http://nost.gsfc.nasa.gov/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf> [Consultado en octubre 2006].
- Silió, Teresa (2005). *Fundamentos tecnológicos del acceso abierto: Open Archives Initiative y Open Archival Information System*. En: El profesional de la información (septiembre-octubre 2005, Vol. 14, Núm. 5). Disponible en: <http://www.elprofesionalde lainformacion.com/contenidos/2005/septiembre/8.pdf> [Consultado en octubre 2007].
- Smith, Abby (1999). *Why Digitize?* Council on Library and Information Resources. Disponible en: <http://www.clir.org/pubs/reports/pub80-smith/pub80.html> [Consultado en mayo 2007].
- Smith, Abby (2001). *Strategies for Building Digitized Collections*. Washington, DC: Council on Library and Information Resources, September. Disponible en: <http://www.clir.org/pubs/reports/pub101/contents.html>. [Consultado en marzo 2007].
- Tanner, Simon (2004). *Deciding whether Optical Character Recognition is feasible*. Disponible en: http://www.odl.ox.ac.uk/papers/OCRFfeasibility_final.pdf [Consultado

Digitalización de Colecciones. Texto e imagen. Volumen 1

en enero 2008].

- Technical Advisory Service for images. *Choosing a File Format*. Disponible en: <http://www.tasi.ac.uk/advice/creating/format.html#fo5> [Consultado en marzo 2007].
- Technical Advisory Service for Images. *New digital image file formats*. Disponible en: <http://www.tasi.ac.uk/advice/creating/newfile.html> [Consultado en marzo 2007].
- Technical Advisory Service for Images (TASI). *File formats and compression*. Disponible en: <http://www.tasi.ac.uk/advice/creating/fileformat.html> [Consultado en marzo 2007].
- *Tendencias para el almacenamiento digital en 2008*. Disponible en: <http://www.faq-mac.com/noticias/node/27989> [Consultado en abril 2008]
- *The National Digital Information Infrastructure and Preservation Program (NDIIPP)*. Disponible en: <http://www.digitalpreservation.gov/> [Consultado en octubre 2007].
- *TWAIN*. En: Wikipedia, la enciclopedia libre. Disponible en: <http://en.wikipedia.org/wiki/TWAIN> [Consultado en septiembre 2007].
- Waugh Andrew, et al. (2000). *Preserving Digital Information Forever*. International Conference on Digital Libraries. Proceedings of the fifth ACM conference on Digital libraries. San Antonio, Texas. p. 175 - 184.
- Western States Digital Standards Group. *Western States Digital Imaging Best Practices vers. 1.0*. University of Denver and the Colorado Digitization Program, 2003. disponible en: <http://www.bcr.org/cdp/best/digital-imaging-bp.pdf>. [Consultado en noviembre 2007].
- Williams, Don (2000). *Selecting a Scanner*. En: Guides to Quality in Visual Resource Imaging Digital Library Federation; Research Libraries Group. Disponible en: <http://www.rlg.org/visguides/visguide2.html> [Consultado en septiembre 2007].
- Wilson, Wayne. *Building and Managing a Digital Collection in a Small Library*. North Carolina Library Association. Fall 2003, number 3. Disponible en: http://www.nclaonline.org/NCL/ncl/NCL_61_3_Fall2003.pdf [Consultado en mayo 2007].