

# Guía de la información electrónica.

Cómo tratar los datos legibles por máquina y la  
documentación electrónica.

GUÍA DE LA INFORMACIÓN ELECTRÓNICA.  
Cómo tratar los datos legibles por máquina y la  
documentación electrónica.  
Edición revisada y ampliada.  
DLM-FORUM. Electronic Records.

Traducción a partir de las ediciones originales  
en lenguas francesa e inglesa.  
*José María Fernández Hevia, 2000.*  
<http://www.jazzfree.com/jazz8/jmfhevia>  
[jmfhevia@airtel.net](mailto:jmfhevia@airtel.net).

Esta guía multidisciplinar fue diseñada originariamente por el consultor Jean-Michel Cornu en estrecha cooperación con el Archivo Histórico de la Comisión Europea y con expertos de los Estados miembros. En una versión provisional, fue presentada al forum DLM sobre archivos electrónicos (Bruselas, 18 al 20 de diciembre 1996) ISBN 92-827-9130-0 (EN, FR, DE). Los comentarios y anotaciones recibidos durante el forum, transmitidos poco después al Archivo Histórico de la Comisión Europea, fueron incluidos por Sylis SA (Lille, Francia).

Esta versión de la guía, puesta al día y ampliada, es uno de los resultados concretos del forum DLM; es asimismo reflejo de los esfuerzos continuos de cooperación hechos por diferentes disciplinas (administración pública, archivos, industria, comunidad de investigadores), afectadas por el problema de los datos electrónicos y de la gestión de la información dentro de una sociedad de la información en rápida mutación.

Otras informaciones sobre la Unión Europea se encuentran disponibles en Internet vía servidor EUROPA (<http://europa.eu.int>).

Al final de la obra se encuentra una ficha bibliográfica.

Luxemburgo: Oficina de las publicaciones oficiales de las Comunidades Europeas, 1997.

ISBN 92-828-2286-9

© Comunidades Europeas, 1998.

Reproducción autorizada, a condición de citar la fuente.

Impreso en Italia.

<b>PRESENTACIÓN.</b>	<b>5</b>
<b>1. INTRODUCCIÓN.</b>	<b>7</b>
1.1. Presentación de la guía multidisciplinar.	7
1.2. ¿Cómo leer esta guía multidisciplinar?.	9
1.3. Creación individual / Uso general.	10
1.4. El ciclo vital de la información: sus tres fases.	10
<b>2.- DEL DATO INDIVIDUAL A LA INFORMACIÓN ELECTRÓNICA ESTRUCTURADA.</b>	<b>11</b>
2.1. ¿Qué es información?.	11
2.2. ¿Qué es un dato?.	11
2.3. ¿Por qué es diferente la información electrónica?.	11
2.4. Creación de datos electrónicos.	12
2.5. Dos modos de estructurar los datos.	13
2.5.1. Documentos.	14
2.5.2. Bases de datos.	15
2.5.3. Adquisición de equipos y aplicaciones para la información electrónica.	15
<b>3.- CICLO DE VIDA Y REPARTO DE RESPONSABILIDADES.</b>	<b>17</b>
3.1. Vista de conjunto del ciclo de vida.	17
3.2. Definición de responsabilidades para cada fase.	18
<b>4. DISEÑO, CREACIÓN Y MANTENIMIENTO DE LA INFORMACIÓN ELECTRÓNICA.</b>	<b>20</b>
4.1. La fase de diseño.	20
4.2. Creación de la información electrónica.	20
4.3. Integración, migración y valoración de la información electrónica	21
4.3.1. Integración de la información.	21
4.3.2. Migración de información.	22
4.3.3. Del papel o el microfilm a la imagen escaneada.	24
4.3.4. De la imagen escaneada a los formatos codificados.	24
4.3.5. De un formato electrónico a otro.	26
4.3.6. Valoración de la información electrónica.	27
4.4. Gestión y clasificación de la información electrónica.	30
	<b>3</b>

4.5. Transferencias.	32
<b>5.- CONSERVACIÓN DE LA INFORMACIÓN ELECTRÓNICA A LARGO Y A CORTO PLAZO.</b>	<b>34</b>
5.1. Soportes para el almacenamiento de datos.	34
5.2. Formatos de archivos.	37
5.2.1. Imágenes de mapa de bits.	38
5.2.2. Archivos de gráficos vectoriales.	39
5.2.3. Archivos de texto.	39
5.2.4. Datos y programas.	41
<b>6. ACCESO Y DIFUSIÓN DE LA INFORMACIÓN.</b>	<b>43</b>
6.1. Hacia la sociedad de la información.	43
6.2. Normas y estándares para el acceso a los datos.	44
6.3. Seguridad.	45
6.3.1. Derechos de acceso.	45
6.3.2. Encriptación y autenticación.	46
6.4. Acceso a los datos.	47
<b>7. CONCLUSIONES.</b>	<b>48</b>
<b>8. ANEXOS.</b>	<b>49</b>
8.1. Terminología.	49
8.2. Preguntas en suspenso.	50
8.3. Para entender la normalización.	50
8.4. Lista de tests para los DLM.	53
8.5. Modelo ideal: ¿qué metadatos es necesario crear?.	55
8.6. Modelo ideal: ¿Cómo elegir los estándares correctos?.	57
8.7. Índice de términos.	58
8.8. Índice de ilustraciones.	60
8.9. Bibliografía.	60

## **PRESENTACIÓN.**

El forum DLM 96, organizado conjuntamente por los Estados miembros de la Unión Europea y la Comisión Europea en diciembre de 1996 en Bruselas, reunió expertos de la industria, de la investigación, de la administración y de los archivos alrededor de un tema central, denominado cada vez más frecuentemente la "memoria de la sociedad de la información". Hace solamente algunos años, nadie habría imaginado que las innovaciones tecnológicas afectarían al sector de los archivos hasta el punto que la autenticidad y la conservación duradera de informaciones indispensables se encontrasen gravemente amenazadas a corto plazo en caso de no tomar medidas adecuadas a tiempo.

Los servicios de archivo son parte integrante de la sociedad de la información. Desde la publicación del informe titulado "Europa y la sociedad de la información" (informe Bangemann) en 1994, los programas de acción y apoyo de la Comisión Europea les otorgan una posición más destacada. Ellos jugarán también un papel cada vez más importante dentro de la gestión moderna de la información. El forum DLM 96 ha marcado un hito importante en este contexto al abrir la vía a nuevos cambios.

Esta Guía de la Información Electrónica es uno de los resultados más señalados del forum DLM. Fue difundida entre los participantes como anteproyecto y sirvió de base de trabajo y de discusión. Esta nueva versión, considerablemente revisada y ampliada, fue elaborada gracias a las casi 300 proposiciones de modificación y mejora que, en cuestiones fundamentales, los expertos nacionales presentaron a los servicios de la Comisión durante y después del forum DLM.

Estoy convencido que esta guía redactada en común y de manera interdisciplinaria contribuirá a resolver los urgentes problemas que plantean la gestión y el archivo de la información almacenada en soporte electrónico, al proponer estrategias a corto y medio plazo. Corresponde ahora a las personas competentes en la administración y la industria el tomar las medidas necesarias, en acuerdo con los especialistas de la gestión de archivos electrónicos, y el diseñar soluciones prácticas en el marco de una cooperación europea. De esta manera, saldría reforzada la confianza de las empresas y de las personas físicas en la seguridad de las nuevas tecnologías.

Gracias a su estructura, esta guía constituye una fuente de información de gran riqueza, tanto para expertos como para aficionados. Representa un paso importante en la preservación duradera de la "memoria de la sociedad de la información" y en la promoción activa de la transparencia para todos los ciudadanos, tanto de la actividad como de los procesos de decisión de las administraciones nacionales e instituciones de la Unión Europea.

Dr. Martin Bangemann

Miembro de la Comisión Europea, encargado de los asuntos industriales, las tecnologías de la información y las telecomunicaciones.



## 1. Introducción.

### 1.1. Presentación de la guía multidisciplinar.

**Documentos y bases de datos** son cada vez más utilizados bajo forma de soporte electrónico.

El formato de la gran mayoría de los documentos sigue siendo en la actualidad el papel. Y sin embargo, en muchas ocasiones, éstos son producidos originariamente por equipos informáticos. Es el caso, por ejemplo, de los mensajes electrónicos, de las notas internas o incluso de las minutas y actas de reuniones de trabajo que jalonan toda vida profesional. Con el desarrollo del correo electrónico, el número de documentos electrónicos ha aumentado rápidamente en administraciones y empresas, de tal manera que éstos han empezado a reemplazar al papel como soporte. Es necesario, por tanto, empezar a analizar el impacto de este fenómeno en la práctica diaria así como en las reglas de uso y archivo de la documentación.

Por otro lado, muchos organismos han desarrollado sus propias bases de datos, cuya información debe generalmente conservarse cierto tiempo después de su uso inmediato -por razones legales o para permitir ulteriores investigaciones-. Es necesario así tener en cuenta aspectos como la conservación a largo plazo y la accesibilidad a los datos, y la capacidad de éstos de producir nueva información.

**Sugerencias** para ayudar a definir estrategias en el ámbito de la información electrónica.

Esta guía multidisciplinar no ofrece respuestas definitivas en todas estas cuestiones, que aún necesitan una mayor discusión entre las diferentes disciplinas afectadas -administraciones públicas, archivos, industria y comunidad de investigadores-. La guía ofrece sugerencias y consejos para ayudar a cada organismo a definir su propia estrategia en el ámbito de la información electrónica. No se intenta ofrecer un modelo europeo único para tratar estas cuestiones, sino más bien reunir la experiencia adquirida en organismos nacionales, regionales y europeos en beneficio de todos.

Una versión previa de esta guía multidisciplinar, fue entregada en su día por la Comisión Europea a los participantes al forum DLM que sobre archivos electrónicos se celebró en Bruselas del 18 al 20 de diciembre de 1996. El forum dio lugar al envío de comentarios y anotaciones escritas por expertos nacionales, y a discusiones posteriores destinadas a la preparación de esta versión puesta al día y ampliada.

Ésta guía puede también ser utilizada como complemento a la del Consejo Internacional de Archivos (CIA) sobre documentación electrónica, que se limita sin embargo a abordar el tema desde el punto de vista de una única disciplina.

**Cadena DLM.** Comprende varios tipos de acción: indicadores específicos, estudios, ayuda a los usuarios.

Las acciones de seguimiento del forum DLM, precisadas en lo que se denomina "diez puntos"<sup>1</sup> (ver el siguiente gráfico), están ligadas a toda la cadena de producción y mantenimiento de los archivos electrónicos. Controladas por el comité de vigilancia DLM, creado como una de las medidas de seguimiento, comprenden:

- Ayuda a los usuarios (formación y elaboración de esta guía multidisciplinar).
- Estudio detallado de las relaciones existentes entre administración pública y archivos.
- Indicaciones específicas de los requisitos funcionales de las especificaciones DLM, de las normas TI para fabricantes de aplicaciones y equipos informáticos y los organismos de normalización, sobre las implicaciones jurídicas para los gestores DLM así como sobre el acceso a la información para el ciudadano y los investigadores.

<sup>1</sup> Ver: Actas del Forum DLM sobre documentación legible por máquina, Bruselas, 18 a 20 de diciembre de 1996, INSAR - Correo europeo de archivos, suplemento II, EUR-OP, Luxemburgo, 1997, p.353, y en Internet: <http://www.echo.lu/dim/en/home.html>

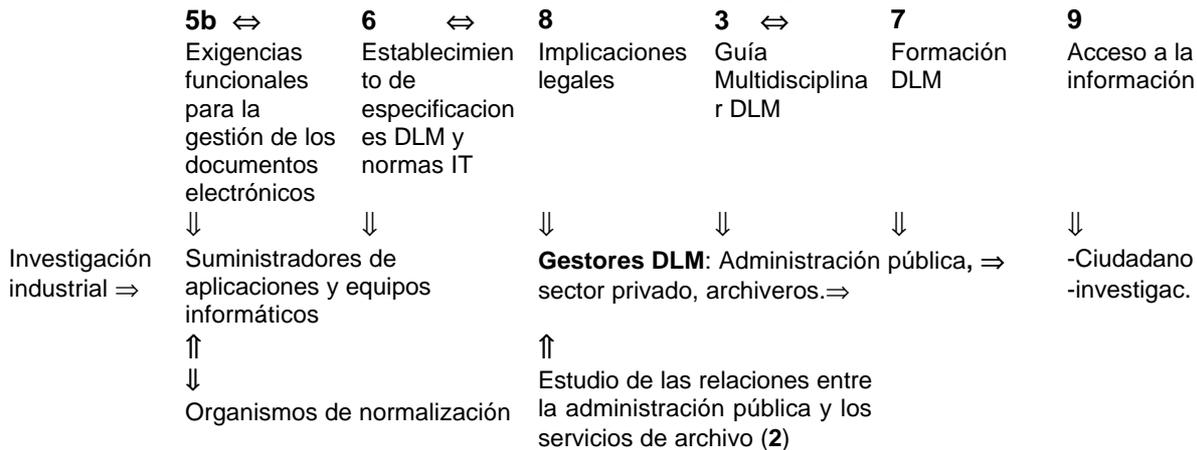
**10 DLM MONITORING COMMITTEE**

Grupos de trabajo multidisciplinares (ver cadena DLM)

5A, Puntos focales nacionales / regionales

Difusión:  
1, Actas del Forum DLM (INSAR Suppl. II)  
4, DLM Website  
(<http://www.echo.lu/dlm/en/home.html>)

**La cadena DLM y las medidas de seguimiento**



**Acercamiento multidisciplinar** para tener en cuenta las necesidades de los diversos implicados.

Esta guía multidisciplinar se dirige tanto a personas que no poseen más que un conocimiento básico de la gestión de documentación electrónica, como a aquellas que tienen un conocimiento más avanzado, en particular:

- Las administraciones públicas.
- Los servicios de archivo.
- La industria.
- La comunidad investigadora

en los países miembros y las instituciones de la Unión Europea.

## 1.2. ¿Cómo leer esta guía multidisciplinar?

El cuerpo del texto está dirigido a todo tipo de público.

Para facilitar que los lectores localicen la información adecuada a sus necesidades y a su nivel de conocimientos, ésta se ha incluido en cuatro tipos diferentes de recuadros.

### **Nociones de base.**

Estos recuadros revisan conceptos básicos utilizados en cada profesión.



### **El CD Rom**

(recuadros para nociones básicas)

### **Nociones avanzadas.**

Estos recuadros presentan nociones más técnicas para quienes deseen entrar en detalles.

### **Ficheros gráficos**

(recuadros para nociones avanzadas)

### **Ejemplos.**

Diversos ejemplos, que permiten hacerse una idea de las diferentes soluciones adoptadas.

### **Alemania.**

(recuadros que presentan ejemplos).

### **Posibilidades.**

Estos recuadros presentan consejos o posibilidades para llevar a cabo una elección correcta.

### **Escanear un documento.**

(recuadros que proponen ayudas a la hora de tomar decisiones)

**Estándares**  
El consenso de cada norma o estándar no es siempre el mismo.

No todos los estándares y normas indicados en la guía disfrutaban del mismo grado de consenso. Determinados símbolos permiten advertir rápidamente el grado de confianza que puede otorgarse a cada especificación.

Especificaciones estables y reconocidas.

Especificaciones en curso o con pocos suministradores.

Especificaciones "propietarias" sin garantía de perdurabilidad.

### **1.3. Creación individual / Uso general.**

**Un documento o una base de datos**  
creados por una persona en el marco de sus funciones es un elemento del sistema de información del organismo

¿Por qué es tan importante tener una estrategia a largo plazo cuando se crea, pone al día o suministra información?

Incluso en el caso de que los documentos o las bases de datos fuesen creados sólo por una o varias personas, tienen a menudo un interés mucho más general del que podría suponerse:

- Pueden ser utilizados o puestos al día por personas que no participaron en su creación.
- Son utilizados a menudo junto a muchos otros documentos, algunos de los cuales ni siquiera son conocidos por el creador.
- La información contenida puede ser reutilizada mucho tiempo después del final de su vida ordinaria (por ejemplo, por razones legales o históricas).

Esto quiere decir que documentos y bases de datos deben ser diseñados como componentes de un sistema global de información. Es fundamental tener en cuenta, desde las primeras etapas, los aspectos de conservación, acceso a la información y protección de la vida privada.

### **1.4. El ciclo vital de la información: sus tres fases.**

**3 fases:**  
Diseño, Creación,  
Mantenimiento.

Existen tres fases principales en la vida de la información electrónica:

- **Fase de diseño**, durante la que se define una estrategia global.
- **Fase de creación** efectiva. Implica generalmente un número limitado de personas.
- **Fase de mantenimiento**, que incluye la utilización y conservación de los datos.

Estas tres fases no pueden abordarse del mismo modo. Existen diferencias en diversos conceptos:

- Puesta al día de la información.
- Frecuencia de acceso a la información.
- Responsabilidad.
- Etc.

## 2.- Del dato individual a la información electrónica estructurada.

### 2.1. ¿Qué es información?.

**Información.** Una información es una indicación o un evento llevado al conocimiento de una persona o de un grupo. Es posible crearla, mantenerla, conservarla o un evento y transmitirla.  
Es una indicación o un evento llevado al conocimiento de una persona o de un grupo. La información está en la base de la organización de todo proceso de trabajo. Este concepto se ha vuelto tan importante, que la sociedad de la información es considerada en el día de hoy como la etapa siguiente a la de la “sociedad industrial”.

### 2.2. ¿Qué es un dato?.

**Datos.** Un dato es una unidad elemental de información.  
Pueden reagruparse bajo forma de documento o de catálogo. En un documento, por ejemplo, se agrupan numerosos datos para presentar una argumentación o rendir cuentas de una acción. Hasta hace muy poco, la mayor parte de los datos eran conservados y transmitidos mediante documentos en soporte papel (o a veces, para los más antiguos, en otros soportes, como la piedra tallada).  
A veces encontramos datos en forma de listas en soporte papel, como ocurre con las guías de teléfonos. Ya no se trata de presentar una argumentación, sino de suministrar materia prima para una acción futura (localizar el número de teléfono del señor García, por ejemplo). En estos casos, es crucial la clasificación de los datos para una fácil localización de la información requerida (por ejemplo, en una guía de teléfono los apellidos se clasifican por núcleos de población y después por orden alfabético).  
Es posible conservar datos en soportes diferentes al papel: por ejemplo, el soporte electrónico, que permite que la información pueda tratarse más fácilmente.

**Gestión de los datos.** La rápida evolución de las tecnologías ha vuelto cada vez más complejo el garantizar soluciones duraderas a la hora de conservar los datos. Los formatos de ficheros y los soportes electrónicos evolucionan rápidamente, mientras que su esperanza de vida sigue siendo aún más escasa que la del papel.  
Hoy es cada vez más complejo administrar los datos. Además, la producción de información ha sufrido un fuerte aumento, sobre todo en las administraciones públicas. Como consecuencia, las tareas de clasificación y estructuración de los datos para poder acceder a ellos tiempo después de su producción, se vuelven cada vez más complejas.

### 2.3. ¿Por qué es diferente la información electrónica?.

**El soporte no es el mensaje.** Cuando los datos se conservan en un soporte electrónico, ya no es posible leerlos sin la ayuda de una herramienta específica, una “máquina” (generalmente, un ordenador).  
Es necesario un instrumento para leer la información electrónica. En efecto, a diferencia de lo que ocurría con los documentos en papel, “el soporte no es el mensaje”. En esta guía, se entiende por “información electrónica” todo dato conservado en un formato que permite su tratamiento automático, tratándose por regla general de soportes electrónicos.

### **La información electrónica**

aporta ventajas e inconvenientes a la gestión de los datos.

Son varias las ventajas de almacenar datos en un soporte legible por máquina:

- el tratamiento de los datos es mucho más fácil. Ya no es necesario rehacerlos en su integridad en el caso de tener que hacer algunas modificaciones parciales.
- los soportes electrónicos permiten en general almacenar más datos en un volumen menor.
- es más fácil copiar un documento completo.
- es más fácil y rápido transferir información de un sitio a otro.
- un soporte como éste permite una utilización más elaborada al recurrir a una estructura de tratamiento electrónica.

Sin embargo, la utilización de estos soportes añade nuevos inconvenientes:

- es necesario utilizar un instrumento para que un operador pueda leer los datos.
- los soportes digitales tienen, generalmente, una vida más corta que el papel o los microfilms.
- es más fácil duplicar o modificar un original (esto plantea problemas de estabilidad y de perdurabilidad en los formatos).

### **2.4. Creación de datos electrónicos.**

Los datos son conservados en un soporte. En el caso de la información electrónica, aquellos pueden ser tratados, comunicados e interpretados por ordenador. Un conjunto de datos puede constituir en ocasiones un documento.

#### **¿Qué es un documento?.**

Definición del Consejo Internacional de Archivos (record).

El Comité CIA sobre archivos electrónicos define al documento (*record*) como un fragmento particular de información registrada generada, acumulada o recibida en el inicio, desarrollo o finalización de una actividad institucional o personal, y que comprende un contenido, un contexto y una estructura suficientes para constituir una prueba o evidencia de esta actividad

En una administración pública, un documento tiene un valor legal, de información o de prueba. El término "documento" se toma en esta guía en su sentido administrativo y archivístico, y no en el utilizado habitualmente en informática.

El hecho que un documento electrónico sea distinto de su soporte, tiene numerosas consecuencias. Por ejemplo, es fácil copiar un documento electrónico de un soporte a otro. Esto quiere decir que es más fácil copiar y difundir la información, pero también se hace más difícil la crucial delimitación del propio concepto de "original" (véase a este respecto el anexo 8.2 sobre preguntas en suspenso).

Los datos almacenados en un documento electrónico deben constituir un conjunto coherente y consistente de información. Diseñar un sistema de clasificación es una de las tareas más importantes a la hora de delimitar documentos. Suele haber varias posibilidades de agrupar datos en un documento, según el nivel de detalle deseado: por ejemplo, en el caso de una base de datos, la totalidad de la base o un subconjunto coherente pueden formar un único documento.

## El documento electrónico.

Está compuesto: de contenido, estructura, contexto y presentación. Como ésta última depende fuertemente del soporte de visualización utilizado, puede que su conservación sea imposible

## Documentación electrónica.

### Nociones avanzadas.

Un documento electrónico está constituido por cuatro elementos principales. Los tres primeros deben ser conservados:

1) El **contenido** del documento, que puede incluir varios tipos de datos:

- Textuales (páginas, párrafos, palabras);
- Numéricos (enteros, flotantes...);
- Gráficos, de imágenes, de vídeo y sonido;
- De enlaces hipertexto

2) La **estructura lógica** del documento (que puede estar incluida en el documento o base de datos, o también ser añadida como entidad separada, en cuyo caso una misma estructura puede ser utilizada por varios documentos). La estructura lógica puede ser muy diferente de la estructura física del documento.

3) El **contexto** descrito en un documento asociado. Puede incluir informaciones como:

- Los metadatos técnicos (aplicaciones y equipos informáticos necesarios, números de versión, estructura del fichero, descripción de los datos, un extracto histórico de las relaciones con otros documentos);
- La descripción del contexto administrativo.

El contexto descrito en la documentación puede volverse complejo cuando el documento se encuentra integrado en una arquitectura de red.

4) La **presentación**, se contempla cada vez más separadamente del propio documento, lo que permite abstraer la información en sí misma del modo en que ésta será presentada. La difusión de la información en soportes variados (CD-ROM, acceso en línea, papel...), se denomina "gestión intersoporte". El instrumento que permita, dentro de varios años, visualizar los datos documentados hoy en día, aún no se ha inventado.

## 2.5. Dos modos de estructurar los datos.

### Bases de datos y documentos.

Dos tipos de estructuras de datos frecuentemente combinados para formar un conjunto compuesto de información.

Para tener la posibilidad de recuperar una información específica, es necesario estructurarla. Según la finalidad de la información, existen principalmente dos medios de estructurar los datos:

- **la base de datos:** en esta estructura, los datos son almacenados en un "pool" o reserva de información, de donde pueden extraerse o ponerse al día.
- **el documento:** estructura utilizada cuando los datos han sido agrupados para dar cuenta de un razonamiento o una actividad. El documento puede a menudo servir de prueba de una actividad particular (acto jurídico...), debiendo registrarse en este caso.



Ilustración 1 ¿Documento o base de datos?.

Por supuesto, existen distintas formas de estructurar los datos. La creciente importancia de la información en los organismos, da lugar a estructuras cada vez más complejas, tanto de documentos como de bases de datos (por ejemplo, una página en un CD-ROM o en un servicio en línea configurada automáticamente con datos extraídos de una base de datos).

Además, a menudo existe una estrecha relación entre una base de datos y varios documentos. Por ejemplo, una actualización de la base de datos puede dar lugar a la producción de un documento definido por procedimiento administrativo. Este tipo de solución es cada vez más frecuente con los programas de Workflow (flujo de tareas), que regulan el desarrollo de los procesos en la empresa.

Los siguientes puntos enumeran algunas diferencias entre “documento” y “base de datos”:

- su conservación plantea diferentes requerimientos.
- un documento debe ser consistente e inalterable (su puesta al día genera un nuevo documento).
- una base de datos puede ser regularmente puesta al día.
- existe también información compuesta (base de datos/documentos), que plantea nuevos problemas.

### 2.5.1. Documentos.

#### Documentos en soporte papel y electrónico.

Coexisten y es necesario tener en cuenta las particularidades de cada soporte.

Entre la oficina basada en documentos en soporte papel y la basada en documentos en soporte electrónico, existen varios estadios:

- 1) oficina tradicional basada en documentos en papel;
- 2) oficina mixta, con ambos tipos de documentos (papel y electrónicos);
- 3) conversión de documentos en papel (mediante escaneado);
- 4) oficina completamente digitalizada (todos los documentos producidos, recibidos o difundidos son electrónicos).

Esta guía se basa sobre todo en las etapas 2 y 3, que son las que corresponden a la situación actual, informando sobre las relaciones existentes entre los documentos electrónicos y los tradicionales (principalmente en papel y microfichas). La etapa 4 es la tendencia futura más que un estadio sensu stricto. Los documentos en papel y electrónicos continuarán coexistiendo probablemente durante muchos años.

Existen numerosas formas de documentos que necesitan tratamientos específicos: cartas, notas, formularios, informes...

Los documentos pueden agruparse en expedientes para formar un conjunto coherente de información. La clasificación de los documentos es un importante factor para conseguir una cómoda recuperación de la información.

**Plan de clasificación y palabras claves.**  
Dos instrumentos complementarios para facilitar la búsqueda de información.

### Clasificación de los documentos.



Nociones básicas.

Existen varias formas de clasificar los documentos. Las dos principales son:

- codificación cronológica (se da un número a cada documento en el momento en el que se registra).
- codificación por tema (por ejemplo, cada documento se adscribe a un número de acuerdo a un plan de clasificación).

La segunda solución es a menudo la más eficaz para recuperar fácilmente un documento, aunque es necesario el diseño previo de un plan de clasificación eficaz.

Otra solución consiste en recuperar la información gracias a palabras claves identificadas o a una búsqueda dentro del mismo texto.

La relación entre el plan de clasificación y el sistema de palabras claves es similar a la existente entre la tabla de materias y el índice de un libro. Los dos son eficaces instrumentos de búsqueda, y uno no debe reemplazar al otro.

Las modernas tecnologías de la información pueden ofrecer nuevos modos de acceso y clasificación.

#### 2.5.2. Bases de datos.

**Conservar una base de datos.**  
Un problema complejo, a causa de la falta de formatos estándares que puedan asegurar la perdurabilidad.

Las bases de datos plantean problemas a la hora de acceder a los datos (p.e., por razones legales o por investigaciones), después de utilizadas aquellas. En la actualidad existen muy pocos formatos estandarizados de bases de datos, por lo que a menudo no existen más que dos opciones:

- copiar la base de datos en un formato de más bajo nivel (en texto plano o en formato ISAM -método de acceso secuencial indizado-).
- conservar la aplicación que generó la base de datos, incluida su documentación (sistema de gestión de la base de datos, aplicación contable, etc.).

La primera solución puede significar perder los elementos estructurales de la base. La segunda, implica a menudo la necesidad de conservar, además de la aplicación y la plataforma informática capaz de hacerla funcionar, los conocimientos necesarios para administrar aquellos (lo que no es fácil después de un tiempo).

En muchos casos, la base de datos está totalmente integrada en una aplicación propietaria, como ocurre a menudo en los programas de gestión.

#### 2.5.3. Adquisición de equipos y aplicaciones para la información electrónica.

La adquisición de equipos, aplicaciones y servicios, es un aspecto importante del uso y archivo de datos electrónicos. Hoy en día, no sólo existen muchos tipos de equipos, aplicaciones y servicios basados en diferentes tecnologías, sino también numerosos estándares, normas y especificaciones técnicas. Sin embargo, la rápida evolución de las tecnologías de la información y de las comunicaciones, hace temer que la supervivencia a largo plazo de los datos electrónicos dependa indebidamente de su formato de almacenamiento, que bien podría no ser aceptado por futuras tecnologías.

## **EPHOS.**

Un programa de la Unión Europea dirigido a ayudar a compradores en el sector público.

Las adquisiciones dentro del sector público de la Unión europea están regidas por la decisión del Consejo de Ministros 87/95 CEE, que obliga a todos los compradores públicos a recurrir a normas oficiales (es decir, normas aprobadas por instancias oficiales de normalización) en sus procedimientos de subasta. EPHOS (el manual europeo de compra de sistemas abiertos) es un programa europeo dirigido a ayudar al sector público a aplicar en sus compras la decisión del Consejo 87/95 CEE, suministrando para ello una serie de manuales que contienen consejos estratégicos sobre tecnología de sistemas abiertos. EPHOS hace referencia normalmente a perfiles internacionales estandarizados de normas internacionales cuando éstos son aplicables. Los módulos EPHOS actuales tratan exclusivamente sobre normas oficiales, aunque algunos en curso de preparación abordan especificaciones disponibles para el gran público.

## **Compras con normas oficiales.**

Plantean problemas prácticos debidos a la indisponibilidad o al coste de los productos y a los estándares de hecho que concurren en competencia.

Sin embargo, la experiencia ha demostrado que en ciertos ámbitos de la información y de las comunicaciones, en particular dentro del sector de las tecnologías de la información, las disposiciones de la decisión del Consejo 87/95 CEE plantean problemas prácticos. Esto es debido:

- a la indisponibilidad de productos que participen de las normas oficiales.
- al coste elevado de los productos que participan de las normas oficiales.
- al peligro de que estas normas no se vuelvan obsoletas en el futuro, de tal manera que no sean contempladas en futuros productos.
- a los estándares *de facto* con los que entran en competencia (que pueden ser especificaciones disponibles para el gran público), profusamente aplicados en productos también muy extendidos entre los usuarios (por ejemplo, las especificaciones de Internet).

De cualquier manera, es necesario recalcar que el problema del control futuro debe aplicarse tanto -y quizás más particularmente- al desarrollo de productos basados en normas *de facto*, como a los basados *de iure* en normas oficiales.

## **Compra de productos.**

La compra basada sobre unas normas estables y abiertas es uno de los principales criterios de evaluación de los productos.

Consecuentemente, el análisis planteado sugiere que las compras públicas estén basadas más en productos que en normas. Los usuarios que produzcan o mantengan datos legibles por máquina, deben asegurarse de disponer de los equipos, aplicaciones y documentación necesarios para recuperar a largo plazo los datos y documentos generados por sus aplicaciones. Sin embargo, debe admitirse que no es siempre práctico o realista el conservar la totalidad de equipos, aplicaciones y documentación durante un largo período. Es evidente que la compra de productos basados en normas estables y abiertas es uno de los principales criterios de evaluación de los productos.

Más concretamente, los usuarios deben tener una idea clara de las normas aceptadas por los productos de las tecnologías de la información, incluyendo una evaluación de la estabilidad y apertura de estas normas. Esto supone la necesidad de una política de compras a largo plazo. Quizás sea necesario elaborar unas líneas directrices comunes de compras para la comunidad de intereses vinculada a la conservación de los datos legibles por máquina.

## **Documentación sobre formatos.**

Es conveniente obligar a los suministradores a que pongan los medios adecuados para recuperar los datos creados por sus aplicaciones, para facilitar su conservación y uso a largo plazo.

### **Cláusula de compra.**

#### Posibilidad

Mientras tanto, se sugiere incluir la siguiente cláusula estándar en las subastas.

#### **Cláusula tipo para insertar en los pliegos de condiciones.**

“A fin de garantizar la conservación y el acceso a largo plazo de los datos del o de los organismos adjudicadores, el suministrador se compromete a suministrar todo el equipo, programas informáticos y documentación que permita la recuperación de los datos generados por sus aplicaciones y su exportación hacia otros entornos y otros formatos”

### 3.- Ciclo de vida y reparto de responsabilidades.

#### 3.1. Vista de conjunto del ciclo de vida.

**El ciclo de vida.** La información electrónica pasa por varias fases durante su ciclo de vida. Un conjunto de fases que es necesario coordinar de la mejor manera posible. Durante estas diversas fases, pueden definirse diferentes tareas. Es importante definir las reglas y los estándares que permitan asegurar la sincronización entre todas las personas implicadas en estas tareas.

*Fase de diseño*

Diseño de una estrategia

*Fase de creación*

Creación de la información

*Fase de mantenimiento*

Migración y valoración



Ilustración 2- Ciclo de vida de la información electrónica.

Pueden existir varias fases de mantenimiento allí donde la responsabilidad sobre la información electrónica sea transferida de un organismo a otro.

El capítulo 4 presenta las tareas referidas a la creación y gestión de la documentación electrónica, con sugerencias para ayudar a definir una estrategia coherente.

El capítulo 5 presenta varias sugerencias para facilitar la conservación de la documentación electrónica (en particular sobre los diferentes tipos de soportes y de formatos de ficheros).

El capítulo 6 presenta varias formas de organizar el acceso y la utilización de la información electrónica, teniendo en cuenta tanto aspectos de organización y seguridad, como los estándares de intercambio de datos.

### 3.2. Definición de responsabilidades para cada fase.

**Definir las responsabilidades** Documentos y bases de datos son de una importancia vital para una administración pública, así como para todo organismo. La información es la materia prima para el funcionamiento de los organismos, y la definición de una estrategia a corto, medio y largo plazo en el tratamiento, conservación y acceso a la información, constituye uno de los aspectos de mayor importancia que condicionan el éxito de un organismo en sus funciones.

Sean cuales sean las posibilidades de cada organismo, es necesario definir correctamente las responsabilidades para cada fase. El reparto de responsabilidades varía según países y organismos. De cualquier manera, sean cuales sean las opciones elegidas, es crucial que la definición clara de responsabilidades sea resultado de una estrategia global y no del azar o consecuencia de relaciones tradicionales entre departamentos.

#### Reparto de responsabilidades.

Ejemplo.

Las administraciones nacionales y comunitarias han optado por diferentes soluciones en lo referente al reparto de responsabilidades.

- la Unesco recomienda “servirse de la profesionalidad de los archiveros para valorar la información, pero conservar la fuente de la información como custodia física del documento” (estudio Unesco/RAMP).
- en los Estados Unidos, los documentos que aún no poseen valor histórico son conservados por la administración de los archivos nacionales (NARA) en más de veinte servicios repartidos a través de los Estados Unidos, así como por los “archivos afiliados” autorizados por el NARA. Estos documentos dependen del NARA, mientras que los archivos afiliados se responsabilizan de su conservación, gestión y acceso.
- en la Comisión Europea, algunos servicios/oficinas de archivos son responsables del registro y distribución del correo en algunas direcciones generales, mientras que otros se organizan de un modo diferente.

Un correcto reparto de responsabilidades debe tener en cuenta cada tarea, pero también la cultura y conocimientos técnicos del organismo. Debe implicarse desde el inicio a los diferentes servicios afectados y a los archiveros, debiéndose tener también en cuenta las informaciones provenientes del mercado y del mundo de la investigación.

Cuando la información electrónica es producida y mantenida por varios organismos, la identificación de las responsabilidades debe hacerse mediante el diálogo entre los diferentes organismos y servicios de archivos.

#### ¿Qué papel juegan los archiveros?.

Tienen valiosos conocimientos en valoración documental y pueden ser útiles desde el inicio del ciclo de vida de la documentación.

El archivero es el responsable de la conservación de los documentos. Puede también ofrecer una valiosa experiencia en lo referente a la evaluación del valor de un documento. Su papel se ha ampliado, pasando de la recepción pasiva de los documentos al final de su vida a una participación activa en las primeras etapas de su ciclo.

Es necesaria una cooperación más estrecha entre los archiveros y los encargados de la información electrónica, tanto en la administración pública como en el sector privado.

**Los diferentes tipos de información requieren de diferentes expertos:**

-La información en directo está generalmente gestionada por el autor del documento o el administrador de la base de datos.

-Los archiveros son expertos en la gestión de los documentos.

-Los bibliotecarios son expertos en la gestión de las publicaciones.

**Tres tipos de información.**

Nociones avanzadas.

Existen tres tipos de información.

- Información en directo. Producida por un organismo como información no registrada, puede ser efímera o acabar convertida en documento.
- Documentos. Los archiveros son los expertos que ayudan a administrar los documentos durante todo su ciclo de vida. La información puede presentarse directamente como documento, o devenir como tal a raíz de la captura de información en directo -to capture a record-. En este caso, la información está “congelada” y su contexto cobra importancia primordial para facilitar el acceso y la utilización de los documentos.

Mientras que esta guía presenta la información y la documentación electrónica desde un punto de vista multidisciplinar, la guía del CIA sobre documentación electrónica (ver anexo 8.9) da más detalles sobre la conservación de los documentos desde el punto de vista de los archiveros.

## 4. Diseño, creación y mantenimiento de la información electrónica.

Cada una de las tareas presentes en el ciclo de vida tiene sus propias particularidades. Este capítulo describe las referentes al diseño, creación y mantenimiento de la información electrónica.

### 4.1. La fase de diseño.

**Equipo estratégico.**  
La puesta en marcha desde el comienzo de un equipo multidisciplinar para definir y seguir la estrategia de la información electrónica permite optimizar los resultados.

La mejor manera de optimizar la gestión de la información electrónica es definir una estrategia global y coherente desde el comienzo, que debe ser desarrollada mediante la implicación de todos los interesados. Una de las soluciones consiste en el establecimiento de un equipo multidisciplinar que defina y mantenga esta estrategia; aunque no sea una práctica muy extendida, puede considerarse que esta forma de abordar el problema es la más deseable.

#### Definir una estrategia

##### Posibilidad

El equipo multidisciplinar responsable de la estrategia de la información electrónica puede encargarse de los puntos descritos a continuación:

- toma en consideración de las necesidades de los usuarios (incluyendo también aspectos legales).
- identificación de los documentos importantes.
- establecimiento de reglas para el diseño de un plan de clasificación eficaz.
- definición de normas y especificaciones que aseguren la independencia de los datos en relación a los soportes y garanticen su pervivencia.
- establecimiento de un plan de valoración de la información electrónica.
- identificación de los responsables de cada tarea del ciclo de vida.
- definición de una política de formación e información en de la/s organización/es implicada/s<sup>2</sup>.
- seguimiento de la puesta en funcionamiento de los nuevos sistemas.

Véase igualmente la lista de comprobación de la estrategia para la información electrónica en el anexo 8.4.

**Aspectos legales:** Los aspectos legales son uno de los más importantes en el contexto de una administración pública. Sin embargo, en lo referente a la información electrónica existen aún algunas cuestiones poco estudiadas, por lo que es necesario obrar con sumo cuidado (véase la lista de las principales cuestiones en suspenso en el anexo 8.2.).

### 4.2. Creación de la información electrónica.

Es importante que cada una de las tareas descritas en el ciclo de vida se tengan en consideración lo más tempranamente posible. Por ello, es necesario utilizar reglas comunes durante el proceso de creación de la información electrónica. Esto facilitará además la comunicación entre los diversos responsables encargados más adelante del tratamiento de aquellas.

---

<sup>2</sup> Sería deseable que la información y el tratamiento de la documentación electrónica pasase a ser un elemento en la formación de los archiveros y empleados de las administraciones públicas.

### Reglas de creación.

Cuando se crea la información electrónica, es necesario respetar ciertas reglas que facilitarán las etapas posteriores del ciclo de vida.

### Creación de la información electrónica

#### Posibilidad

A la hora de crear información electrónica, deben aplicarse las siguientes reglas:

- cada documento o base de datos debe estar claramente identificado por el organismo responsable de la gestión de la información electrónica
- toda información electrónica debe incluir una documentación contextual, que debe ser mantenida por el responsable de la gestión de la información electrónica en cada etapa de su ciclo de vida.
- debe articularse un procedimiento para tratar la información electrónica de procedencia dudosa.
- No debe efectuarse ninguna destrucción o modificación de datos sin un procedimiento previo de aprobación. Esto permite garantizar la conservación de las modificaciones en los documentos (la destrucción o la puesta al día de los datos hace imposible combinar, reconocer, recuperar o identificarlos)

La información puede generarse inicialmente en soporte papel (y después digitalizarse), o bien directamente en un formato electrónico (tratamiento de texto, base de datos, mensajería electrónica). En todos los casos, debe ser generada en un formato estándar o convertida a él (a menos que su autenticidad no sea puesta en peligro por ello). El capítulo 5.2 ofrece detalles sobre los formatos más apropiados.

### ¿Información electrónica o documentación en soporte papel?.

#### Posibilidad

- Generalmente, la información en papel se conserva en este mismo formato. Los instrumentos de búsqueda electrónica pueden ayudar a su gestión y recuperación.
- Los documentos electrónicos deben ser conservados en soportes electrónicos.

Los documentos en papel -o una parte de su información contextual-, pueden ser escaneados para producir documentos electrónicos con la finalidad de facilitar su consulta y utilización.

## 4.3. Integración, migración y valoración de la información electrónica

### 4.3.1. Integración de la información.

#### Agrupamiento de datos.

Puede ser útil para consultas posteriores, pero también delicado.

La reorganización, selección y reunión de datos para crear un conjunto más compacto es una tarea delicada, puesto que es necesario respetar el derecho a la vida privada (ver sección 6.3. sobre el anonimato).

Sin embargo, a veces puede ser necesario agrupar datos provenientes de diferentes organismos: los futuros usuarios de datos públicos querrán poder consultar tanto por "tema" como por el organismo productor.

### ¿Qué debe incluirse?

La documentación, su estructura y un conjunto coherente de textos y datos.

Siempre existe, en efecto, más de una manera de reunir información. ¿Es necesario vincular un informe con sus correcciones posteriores?. ¿Con qué documentos debe conservarse una base de datos?. Sea cual sea la solución elegida, además del documento o de la base de datos, el archivo debe contener su propia documentación (incluyendo sus “metadatos”, es decir, su información contextual) y su estructura (en forma de archivo separado o incluido en el principal).

#### Documentos compuestos creados a partir de una base de datos.

##### Ejemplo

Cuando se crea dinámicamente un documento a partir de una base de datos, existen varias soluciones para su conservación a largo plazo:

- “congelar” la base de datos con una consulta específica para producir un documento tradicional, en cuyo caso se pierden posibles combinaciones.
- en la medida que se incluya un análisis retrospectivo total de la aplicación, exportar la base de datos completa cuando el sistema ya no sea operativo.
- bien conservar la base de datos, pero igualmente la aplicación que genera el documento dinámicamente.

Aún no están completamente claras las soluciones que deben adoptarse con los documentos compuestos, más complejos que los simples.

El tamaño de la unidad de información puede variar en función de las opciones tomadas: por un lado, la multiplicidad de referencias y enlaces hipertexto entre documentos existentes en la actualidad, podría dar lugar a conservar el conjunto de datos del planeta como un sólo documento; sin embargo, las unidades de información deben tener unas dimensiones aceptables para poder ser utilizadas con provecho.

### 4.3.2. Migración de información.

#### Migración del soporte papel al electrónico:

existen dos soluciones.  
\*Escanear solamente.  
\*Escanear y codificar.

Cuando se desea hacer “migrar” un documento en soporte papel a un formato electrónico, existen dos soluciones principales:

- escanear el documento para obtener de él una imagen.
- escanear el documento y después codificarlo en un formato electrónico (por ejemplo, con el reconocimiento óptico de caracteres o la vectorización de gráficos, tal y como se describe en la sección 4.3.4.).

Documentos iniciales    escaneado

OCR o vectorización  
de gráficos (opción)

Documento  
electrónico

Ilustración 3: Conversión de un documento en soporte papel a formato electrónico.

## ¿Codificar o no codificar?.

### Posibilidad

Para texto, el escaneado simple es más fácil y no requiere un tratamiento añadido, como ocurre con el reconocimiento óptico de caracteres.

Sin embargo, las dimensiones del fichero que debe conservarse son mucho mayores (más de 50 kb. por página, frente a sólo unos pocos kb. en el segundo caso). Además, es bastante más fácil editar un texto formado por caracteres que una imagen que representa este texto. La indización para establecer referencias cruzadas es también bastante más simple.

Otra posibilidad consiste en combinar las dos soluciones. Así, es posible la conservación de las imágenes en su formato inicial mientras que las partes textuales se pasan a un sistema de reconocimiento óptico de caracteres. En los capítulos 5 (Conservación de la información electrónica) y 6 (Acceso y difusión de la información), se reseñan algunas indicaciones detalladas sobre los formatos indicados para la conservación a corto y a largo plazo y el acceso a los datos.

### **Conversión de formato electrónico.**

Los cambios de formato electrónico afectan a todos los tipos de documentos.

Existe un tercer tipo de conversión que transforma un formato electrónico en otro diferente. Debería llevarse a cabo únicamente cuando se trata de transformar el formato de un documento a otro más estandarizado y de más larga perdurabilidad, o de transformarlo en un estándar que ofrezca mayores posibilidades (por ejemplo, añadir estructura a un fichero plano). Puede realizarse también como una tercera etapa que permita obtener a partir de un documento en soporte papel, otro en formato electrónico más estructurado (por ejemplo, un documento con una estructura explícita o una base de datos). En ocasiones, esta tercera etapa permite dar una estructura de base de datos al documento finalmente obtenido (véase la sección 4.3.3.).

Documentos iniciales   escaneado   codificación   cambio de formato

Registro electrónico

Ilustración 4- Cambio de formato electrónico.

#### 4.3.3. Del papel o el microfilm a la imagen escaneada.

##### Escáneres y faxes.



##### Nociones básicas

El escaneado de documentación permite volcar el contenido de páginas de papel a un fichero informático, que contiene una imagen del documento inicial con resolución variable.

La resolución de un escáner se mide en puntos por pulgada (dpi). En la actualidad, los escáneres permiten fácilmente una resolución de 300 o 600 dpi en color.

Un fax está constituido por un escáner, un sistema de transmisión de datos por mediación de la línea telefónica y una impresora de destino. La resolución del escáner de un fax es de una calidad bastante más baja (entre 100 y 200 dpi)

##### Escáner.

Deben observarse una serie de reglas para garantizar un buen resultado

En el escaneo de documentos, es conveniente respetar algunas reglas básicas; la calidad del resultado dependerá de ello en buena medida.

Los textos sólo deben conservarse en forma de fichero de imagen cuando sean pequeños o cuando el reconocimiento óptico de caracteres no sea posible (p.e. firma...). En el resto de casos, es preferible su codificación.

##### Escanear un documento.

##### Posibilidad

- todos los elementos necesarios del documento inicial deben ser transferidos al mismo soporte.
- las relaciones entre el documento y el resto del sistema de archivos deben ser mantenidas (por ejemplo, otros documentos relacionados).
- la calidad inicial del documento (contraste, tamaño de los caracteres...) debe de ser suficiente para permitir la mejor impresión posible tras una conversión a un formato electrónico.
- antes del escaneado de documentos, debe realizarse un test de muestra.
- incluso si el escaneado está subcontratado, el organismo peticionario debe controlar la calidad e integridad de los documentos digitalizados.
- los formularios deben ser estudiados específicamente para facilitar su migración hacia soportes electrónicos (tamaño de los tipos de letra, emplazamiento de los campos...)

#### 4.3.4. De la imagen escaneada a los formatos codificados.

##### Reconocimiento de caracteres.

El OCR permite generar ficheros textuales a partir de imágenes de texto.

##### Reconocimiento óptico de caracteres.



##### Nociones básicas

El reconocimiento óptico de caracteres (OCR), permite a un ordenador el "releer" un texto. El programa utiliza un fichero de imagen que representa el texto que debe leerse (p. e. un fichero generado por un escáner), analiza los dibujos de los caracteres y permite construir un fichero en formato textual, posteriormente editable mediante cualquier tratamiento de texto.

El reconocimiento no es perfecto. Cuando el ordenador duda entre un carácter específico, lo advierte para que un operador pueda después releer el texto e indicar la letra no reconocida. Más raramente, el ordenador puede leer caracteres de una manera incorrecta. Las utilidades de corrección automática permiten optimizar esta fase de corrección, aunque tras el reconocimiento automático debe efectuarse siempre otro control que puede realizarse con la ayuda de programas informáticos.

El porcentaje usual de reconocimiento de un texto tipográfico a partir de una imagen, viene a ser de un 95% (es decir, de dos a tres errores por línea de texto), porcentaje que puede variar en función del lenguaje de la fuente.

Tras haber escaneado un documento que contenga texto e imágenes, es posible extraer el texto gracias a un sistema de reconocimiento óptico de caracteres, lo que permite volver a trabajar el texto, utilizar solamente algunas partes de él, o indizarlo para facilitar la consulta.

### **Faxes.**

El reconocimiento de caracteres es poco eficaz en los faxes.

### **Archivos de faxes.**

#### **Nociones avanzadas**

La utilización de sistemas de reconocimiento óptico de caracteres en faxes es generalmente difícil a causa de su débil resolución y de la escasa calidad del resultado en el papel.

Una solución consiste en conservar los faxes en su formato digital original (ITU-I Grupo III o ITU-T Grupo IV). Aunque este formato no permite el volver a tratar los textos, suministra una versión comprimida de las imágenes y los textos que ahorra espacio de memoria.

### **La vectorización.**

Es eficaz en los gráficos y ahorra memoria.

Los gráficos que aparecen en el documento original pueden ser vectorizados para ahorrar espacio.

En algunos casos, como por ejemplo la digitalización de planos, la vectorización aporta ventajas añadidas. Por supuesto, la vectorización no es eficaz más que con gráficos e imágenes constituidas por contornos.

### **Vectorización de imágenes**



#### **Nociones básicas**

Los gráficos vectoriales están constituidos por elementos simples (rectas, curvas, rectángulos...). En vez de conservar los gráficos en forma de imagen constituida por puntos -como en el caso de las fotografías-, la vectorización identifica los elementos básicos que constituyen el gráfico. Las ventajas, al igual que ocurre para con el reconocimiento óptico de caracteres, son de dos órdenes:

- el gráfico vectorial ocupa mucha menos memoria que la imagen (sólo se conservan los tipos de elementos y sus coordenadas de inicio y finalización).
- es más fácil la edición del gráfico o la utilización de subconjuntos de éste.

### **Codificación.**

Unas pocas normas aseguran buenos resultados.

Durante la fase de codificación de los datos (OCR o vectorización de gráficos), deben seguirse algunas reglas básicas, propuestas en los recuadros de este capítulo. Esto facilita el tratamiento de la información electrónica en todos los estadios posteriores de su ciclo de vida.

## Codificación de textos y gráficos

### Posibilidad

- los elementos no codificados de un documento (por ejemplo, fotos o imágenes), deben ser conservados por separado en otro formato apropiado.

### 4.3.5. De un formato electrónico a otro.

Hay dos casos en los que presenta interés la transformación de un documento de un formato electrónico a otro diferente:

- la migración de un documento a un formato temporalmente más duradero para facilitar su relectura y consulta a largo plazo.
- añadir estructura a un “texto plano” para obtener un documento estructurado o una base de datos consultable con mayor comodidad

#### **Conversión a un formato estándar.**

Su principal ventaja es asegurar una mayor perdurabilidad.

Cuando se trata de realizar una conversión de formato de información electrónica, es importante tomar precauciones para no perder datos de forma involuntaria: las especificaciones de los formatos propietarios y de los formatos estándares no se corresponden siempre. Los ejemplos ofrecidos a continuación ilustran esta dificultad.

Los formatos estándares para la conservación y acceso a la información electrónica se desarrollan en los capítulos 5 y 6

## Pérdidas de información.

### Ejemplo

#### **Ejemplo de pérdida de información en un documento.**

Un documento presenta el siguiente texto:

“Es conveniente dar por concluido este proyecto(1)”

y una nota a pie de página añade:

“(1) Salvo si los presupuestos inicialmente previstos son finalmente liberados”,

el documento, inicialmente en formato propietario de tratamiento de textos, se graba en un nuevo formato estándar: si no se ha tomado ninguna precaución, el proceso de conversión puede ignorar las notas a pie de página y cambiar así por completo el sentido de la recomendación inicial.

#### **Ejemplo de pérdida de información en una base de datos.**

Una base de datos de un programa de contabilidad se encuentra en un formato específico, propio de la aplicación. Al estar aquella vinculada a un sistema operativo y una plataforma informática determinada, se decide extraer su información para conservarla en un formato independiente. Si no se toma ninguna precaución, pueden perderse los vínculos existentes entre las cuentas antes y después de una modificación del plan contable.

### **Estructurar la información.**

Es posible añadir una estructura a un "texto plano" para obtener un documento estructurado o una base de datos.

El segundo caso de migración de un formato electrónico a otro se da cuando a partir de informaciones no estructuradas, en texto plano, como ocurre con un texto escaneado, se pretende articular una estructura (tabla de materias, índice...) con la finalidad de facilitar su consulta. En la actualidad, existen aplicaciones que pueden ayudar a redefinir la estructura de un documento: por ejemplo, mediante la utilización de diversos tipos de fuentes tipográficas para diferentes niveles de encabezamiento.

Durante la reestructuración de documentos escaneados, es importante reunir los elementos que se encuentran separados por saltos de página o de columna, o por la inserción de un gráfico, tabla, etc.

Todo esto es igualmente aplicable a las bases de datos formateadas en texto plano (viejas bases de datos conservadas en un formato de bajo nivel, o escaneado de anuarios o formularios). En este caso, la estructura se encuentra muchas veces implícita, indicada por separadores (tabuladores o puntos y comas) o por la situación de los campos en la página escaneada (columna, situación del campo de un formulario en papel...).

Este último ejemplo es especialmente importante en tanto que no existe una norma oficial para los formatos de datos estructurados, y que a menudo es necesaria la conservación de datos en un formato de bajo nivel poco estructurado (véase en particular la sección 5.2.4. sobre los formatos de datos).

### **4.3.6. Valoración de la información electrónica.**

#### **Comprobar la documentación.**

La mayor parte de los problemas aparecen en la conversión y en la transmisión de la información.

Conservar la información tiene poco interés si ya no es posible su utilización cuando ésta es necesaria. Cuando la información deba ser transformada en un documento para dar prueba de una actividad, su valoración es particularmente importante: es necesario valorar con cuidado su autenticidad, integridad y valor.

#### **Expurgo de documentación.**

La valoración de los documentos que ya no poseen ninguna utilidad o ningún valor, es una tarea importante.

#### **Expurgo**

##### **Nociones avanzadas**

Uno de los aspectos más importantes de la valoración es la identificación de los documentos que deben destruirse. La conservación total ni es posible ni útil. Es necesario destruir los documentos desde el momento en que éstos ya no poseen ninguna utilidad o ningún valor, lo que además redundará en una mayor facilidad de acceso a los que se mantengan.

Esto también es válido para aquellos items integrados en bases de datos cuyo valor a largo plazo es escaso. Su supresión y el mantenimiento del equilibrio harán más rápida la velocidad de proceso y de acceso a los datos.

A la hora de seleccionar los documentos que deben destruirse, es necesario recurrir a la competencia de los archiveros.

#### **Líneas directrices sobre expurgo.**

Definición del mejor modo de eliminar documentos por parte de archivos regionales y nacionales

#### **Líneas directrices sobre expurgo.**

##### **Ejemplo.**

La Unidad de Estrategia de la Información (ISU) del gabinete del primer ministro, junto con la oficina de archivos de Tasmania (Australia) ha desarrollado unas líneas directrices sobre procedimiento de eliminación de documentos:

- identificación de los documentos con valor indefinido o a largo plazo.
- los documentos temporales deben tener un período de conservación.

- constitución de “autoridades permanentes de expurgo” que otorguen una autorización legal continua para la eliminación de documentos tras un plazo específico.

El anexo nº1 relativo al expurgo, reglamenta el expurgo de los documentos administrativos comunes a la mayor parte de organismos gubernamentales.

La sección 11 del anexo 1, autoriza el expurgo de documentación efímera, rutinaria o duplicada, no destinada a archivos oficiales, o que no debe ser objeto de un registro oficial, a saber:

- notas y cálculos utilizados únicamente para preparar otros documentos, como correspondencia, informes, documentos presupuestarios, programas y listas.
- documentación preliminar no destinada a ser conservada en los archivos del organismo gubernamental y cuyo contenido está integrado en el sistema de archivos o archivado como un documento común. Puede tratarse de borradores y primeras redacciones. Sin embargo, a fin de justificar o apoyar el desarrollo de una determinada política, puede ser necesario guardar borradores de grupos de trabajo cuando hayan experimentado cambios importantes. Los pequeños cambios de redacción no son considerados importantes.
- copias de documentos comunes o documentos conservados también en otro lugar dentro del organismo

### Controles de la información

#### Posibilidad

Pueden efectuarse diferentes tipos de control sobre la información electrónica:

- su conformidad a normas y estándares definidos.
- protección contra la escritura.
- legibilidad de los soportes.
- comparación con la información incluida en la documentación (por ejemplo, al intentar imprimir los primeros documentos).
- integridad del contenido.

Sin embargo, no todo puede ser sujeto a comprobación. En ocasiones el problema no llega a ser detectado hasta la fase de acceso y de utilización de un documento. Si la información ha sido suministrada por otro organismo o servicio, debe requerírseles para que intenten reconstruir la información inicial cuando esto sea posible. La mayor parte de los problemas se presentan durante las fases de conversión (descritas en esta sección) o durante la transmisión a otro organismo (véase la sección 4.5), más que durante la fase de conservación de la información

#### Formatos propietarios.

¿Qué hacer cuando una aplicación genera sus datos con formatos propietarios?.

Si la aplicación genera datos en un formato propietario, puede ser necesaria la conservación del sistema completo para poder acceder a la información (aplicación, plataforma informática, documentación e incluso personal capacitado para poner todo en funcionamiento). Debe estudiarse con cuidado el coste de esta solución en función del valor intrínseco de la información.

Otra alternativa es el transferir la información en un formato de bajo nivel (texto plano para un documento o archivo secuencial plano para una base de datos). En este caso, algunas informaciones, en particular algunos elementos estructurales, corren el riesgo de perderse.

Una tercera solución es el “migrar” la información a otro formato más estandarizado. El desarrollo (o compra) del instrumental necesario supone un cierto coste. La solución puede ser interesante cuando los datos deban ser conservados respetando íntegramente la estructura. El formato de conversión debe ser el más normalizado para garantizar la mayor perdurabilidad posible.

**Posibilidades.** Existen varias soluciones para conservar los datos a largo plazo. La mejor elección depende de varios criterios.

Debe señalarse que aunque el coste de conversión de la información electrónica hacia un nuevo formato pueda ser elevado, el no hacerlo puede suponer un coste aún más importante.

Formatos propietarios.		
Posibilidad		
¿La aplicación admite formatos estándares?	Sí	Utilizar formatos estándares con preferencia a los propietarios existentes por defecto
No		
¿Existen herramientas de conversión?	Sí	Procurarse una herramienta de conversión para “migrar” a formatos estándares
No		
¿Se puede aceptar perder una parte de la estructura?	Sí	“Migrar” a un formato de fichero plano
No		
¿Es más barato desarrollar una herramienta de conversión que conservar el sistema antiguo?.	Sí	Desarrollar una herramienta de conversión para “migrar” a formatos estándares
No		
Conservar la aplicación y la plataforma con sus archivos hasta que se encuentre otra solución		

Ilustración 5- ¿Qué hacer con los formatos propietarios?.

#### 4.4. Gestión y clasificación de la información electrónica.

##### Administrar la información electrónica:

- Registrar nueva información electrónica.
- Establecer un responsable.
- Coordinar procesos.
- Clasificar la información electrónica.
- Decidir sobre las transferencias.

La finalidad de la gestión de la información electrónica es la continua preservación de su **fiabilidad, autenticidad, integridad y comprobabilidad**. Para ello es necesario que el contexto de la información esté bien definido. Cuando el contenido, contexto y estructura son suficientes para constituir una prueba de una actividad, la información se convierte en un documento.

La administración del ciclo de vida de la información electrónica lleva consigo una serie de tareas suplementarias. La responsabilidad en la gestión de un conjunto determinado de información electrónica puede ser transferida a otro organismo o servicio (véase la sección 4.5. sobre transferencias).

La gestión de la información electrónica incluye las tareas siguientes:

- **Registro de los accesos a la información electrónica.** Esto necesita una puesta al día retrospectiva de la información electrónica existente.
- **Designación.** Cuando hay varios organismos o servicios involucrados, la información electrónica debe adscribirse al organismo o servicio indicado e identificarse un gestor de la información electrónica.
- **Seguimiento.** Esta tarea consiste en coordinar las diferentes fases de tratamiento (recepción de la información electrónica, conversión, conservación, utilización, transferencia a otra organización).
- **Clasificación.** Se trata de facilitar la búsqueda de una información particular en soporte electrónico (véase a continuación).
- **Decisión de efectuar transferencias a otro organismo o servicio.**

La gestión de la información electrónica implica a un número creciente de grupos de personas. Las herramientas de flujos de tareas (workflow) pueden ser de una gran ayuda a la hora de administrar los trabajos de cooperación dentro de un grupo.

##### Clasificación de la información.

La clasificación de la información es una de las tareas de gestión más importantes.

Una de las tareas más importantes, y quizás la más compleja, consiste en clasificar la información. El sistema de codificación debe ser lo suficientemente claro para ser comprensible por otros organismos (sobre todo si la responsabilidad en la información electrónica debe ser transferida a otro servicio u organismo).

La estructura permite localizar una información concreta en un documento o una base de datos, mientras que la clasificación permite recuperarla dentro del conjunto de las informaciones conservadas.

Es necesario utilizar el mismo sistema general de clasificación sea cual sea el soporte de la información, aunque pueden añadirse algunos criterios suplementarios de indización a fin de tener en cuenta las especificidades de la información electrónica.

#### La Comisión Europea.

##### Ejemplo

Algunos servicios de la Comisión Europea utilizan un sistema de clasificación básico que hace referencia al funcionamiento de esta institución, a la administración, al personal y al presupuesto.

### **Sistema de clasificación.**

Existen varios criterios para clasificar la información electrónica.

Son diversos los criterios que pueden ser utilizados para clasificar e indizar la información electrónica. Los criterios siguientes son particularmente útiles en el caso de los mensajes electrónicos.

- Tipo de documento.
- Fechas (producción, caducidad).
- Autor(es) (personas, servicios).
- Firmante.
- Destino (personas, servicios).
- Copias (personas, servicios).
- Número de información electrónica, versión.
- Adscripción (fecha, expediente, servicio...).
- Asunto.
- Proyecto o actividad.
- Palabras claves.
- Lenguaje.
- Número de páginas.
- Estatuto (oficial, no oficial), confidencialidad.
- Documentos relacionados, enlaces hacia otros documentos.
- Otras categorías definidas por el usuario.

### **Descripción archivística de un documento.**

#### Posibilidad

El Consejo Internacional de Archivos ha redactado unas normas internacionales para la descripción de archivos (ISAD/G), normas que será necesario tener en cuenta en toda estrategia de clasificación

La identificación de la unidad de descripción es denominada "área de mención de identidad". Se recomienda no modificarla durante la vida del documento.

### **Correo electrónico.**

Según su contenido, debe de ser tratado como una carta electrónica o una discusión informal.

### **Correo electrónico**

#### Nociones avanzadas

El correo electrónico puede ser administrado de diferentes maneras en función de su contenido:

- Debe ser contemplado como una "carta electrónica" cuando el contenido pueda ser archivado y considerado como una declaración "oficial".
- Debe ser contemplado como una discusión informal (por ejemplo, una llamada telefónica), cuando el correo no tenga una utilidad posterior; es el caso de una respuesta automática explicando que la persona "no se encuentra disponible en ese momento, y que leerá su correo en cuanto vuelva".

Como el correo electrónico admite tanto cartas electrónicas oficiales como discusiones informales, en ocasiones es difícil decidir su gestión. La capacitación de los archiveros en materia de valoración contribuirá en gran medida a definir nuevas reglas de gestión del correo electrónico.

Los archivos australianos dan algunos ejemplos de documentos de corta vida y rutinarios cuyo registro no se considera necesario (véanse los ejemplos en el cuadro de la sección 4.3.6.).

#### 4.5. Transferencias.

Al final de la parte activa de su ciclo de vida, los documentos pueden ser remitidos al archivo. No todos los documentos merecen ser conservados. Aquellos sin utilidad o sin valor deben de ser eliminados con la ayuda de los archiveros (véase lo referente al expurgo en el cuadro de la sección 4.3.6.).

##### Dos tipos de transmisión:

- física
- transferencia de la responsabilidad

Existen dos tipos principales de transmisión:

- La transmisión física de una información electrónica.
- La transferencia de la responsabilidad.

La información se transmite generalmente en forma de documento (después de haber sido transformada en documento). Éste puede ser remitido a otra organización u otro servicio, o transferido a los servicios de archivo.

Es posible separar las funciones de responsable de la información electrónica (gestión) de las de conservador de ésta. La información electrónica puede ser conservada en su servicio originario o en un servicio especializado. La facilidad que los medios de comunicación electrónicos otorgan a la hora de copiar o transferir, posibilita la separación de los mencionados roles.

De cualquier manera, las comprobaciones deben efectuarse **ANTES** de transferir la información electrónica.

#### Realizar con éxito una transferencia física.

##### Posibilidad

Las claves del éxito a la hora de realizar una transferencia física descansan en algunas reglas básicas:

- la información debe de ser suministrada en su integridad (incluyendo el contexto).
- las responsabilidades deben de ser claramente definidas tanto en el organismo que suministra la información electrónica como en el receptor.
- durante cada transferencia, los dos organismos (el que suministra la información electrónica y el que la recibe) deben asegurarse de que la información no ha sido alterada sin aprobación previa.

#### Reino Unido.

##### Ejemplo

El *Public Record Office*, en el Reino Unido, plantea actualmente una nueva estrategia para la transferencia de la documentación electrónica.

La transferencia se realiza únicamente por mediación de una red electrónica segura. El documento es, por tanto, independiente del soporte utilizado por el suministrador.

El organismo que recibe los documentos que deben conservarse (en este caso el *Public Record Office*), selecciona un único tipo de soporte para archivarlos, y conserva un control sobre esta única tecnología con la finalidad de garantizar una conservación duradera.

El conjunto transferido comprende varios elementos que son conservados conjuntamente.

- El propio documento electrónico.
- una revisión.

- los metadatos.
- un visualizador que funcione en un PC.
- un navegador que también funcione en un PC.

Se considera que los microordenadores PC tienen la suficiente compatibilidad para garantizar que el documento sea legible en el futuro. Cada documento es por lo tanto autónomo e independiente de todo problema de formato.

### **Comprobar los documentos.**

Los tests garantizan transferencias correctas.

Las comprobaciones realizadas antes de la transferencia permiten garantizar la legibilidad de los documentos.

### **Comprobaciones antes de la transferencia.**

#### Posibilidad

La lista reseñada a continuación ejemplifica los pasos que el organismo transmisor debe seguir para comprobar la información antes de su transferencia. Esto debe de hacerse en estrecha colaboración con el organismo receptor (servicio de archivos u otro).

1. Hacer dos copias de los datos.
2. Comparar los datos de la documentación suministrada con la ayuda de un programa estadístico.
3. Identificar y documentar los errores en la documentación.
4. Pueden también realizarse otras comprobaciones cruzadas, como verificaciones entre documentos y entre variables, a fin de verificar la consistencia de los datos.
5. Consultar al suministrador de la información electrónica si la identificación de los códigos plantea problemas o si hay errores e inconsistencias en los datos.
6. Documentar los ficheros físicos indicando todas las dificultades encontradas.

## 5.- Conservación de la información electrónica a largo y a corto plazo.

### Soportes y formatos de ficheros.

Dos elementos claves para la conservación a largo plazo.

Aunque la durabilidad física de los soportes en los que se almacena la información electrónica es un aspecto importante, también lo es la tecnológica: la conservación física de la documentación electrónica no sirve para nada si equipos y aplicaciones no pueden leer los datos para transformarlos en información útil.

Siendo soporte y contenido diferentes en la documentación electrónica, abordaremos sucesivamente las normas y estándares para los soportes y para los formatos de ficheros, centrándonos especialmente en la perdurabilidad de los soportes y en la madurez y duración previsible de los estándares existentes.

### 5.1. Soportes para el almacenamiento de datos.

#### Existen tres tipos de soportes. -

Además del papel:

- microfilm,
- soporte magnético,
- soporte óptico.

Pueden utilizarse varios soportes. Unos están más capacitados para una conservación a corto plazo, mientras que otros están orientados hacia una conservación a largo plazo.

Además del papel, puede considerarse que existen tres grandes tipos de soportes: microfilms, soportes magnéticos y soportes ópticos, con numerosos subtipos. No serán descritos otros tipos de soportes menos conocidos (p.e. cintas de papel).

#### Soportes de almacenamiento.



##### Nociones básicas

- **Microfilms:** se encuentran muy extendidos en los archivos, pero no permiten fácilmente ni modificaciones ni la consulta de la documentación. Existen algunos estándares, y su duración es excelente.
- **Soportes magnéticos:** tecnológicamente, son algo antiguos. Se basan en la polarización de partículas magnéticas en un sentido u otro para almacenar cada bit. Se utilizan por lo general en cintas de acceso secuencial a los datos.
- **Soportes ópticos:** son los soportes más recientes. Utilizan la desviación (o no) de un rayo luminoso en diminutas muescas realizadas en el soporte para indicar la presencia o no de un bit. Se trata en general de discos que permiten un acceso directo (acceso más rápido a la información que con una cinta) y altas capacidades de almacenamiento.

La cinta magnética y el microfilm son muy utilizados para la conservación a largo plazo, aunque también lo son cada vez más los soportes ópticos, particularmente bien adaptados para este tipo de conservación.

Para una conservación a más corto plazo existe una mayor variedad de soportes, al ser la perdurabilidad del soporte menos importante.

### **Soportes magnéticos.**

Se trata generalmente de cintas o cartuchos.

### **Soportes magnéticos**

#### Posibilidad

Los diversos soportes magnéticos, bastante estandarizados, tienen una variada esperanza de vida. Los principales soportes son:

**Disquete:** el disquete 3 ½ se encuentra muy estandarizado y se utiliza en diferentes plataformas (PC, pero también Mac y Unix). Su poca capacidad -por lo general no permite conservar más de 1,44 Mb- y escasa longevidad, limitan su utilidad a la conservación a muy corto plazo y al intercambio de archivos.

**Cartucho magnético:** es muy utilizado para el almacenamiento de datos en sistemas de medio tamaño. Existen cartuchos de cuarto de pulgada y de media pulgada fabricados por IBM.

**Cinta magnética:** la cinta de 1600 bpi es legible prácticamente en cualquier unidad de cinta, y ha sido reconocida por X/Open como formato de intercambio. La cinta 6250, con una capacidad de 112,5 Mb, es muy utilizada en archivos antiguos. Las cintas deben de ser rebobinadas cada dos años y reescritas cada diez o quince en un nuevo soporte (del mismo tipo o diferente).

**Casete de vídeo de 8 mm.:** aunque normalizado, no existe más que un suministrador principal. La capacidad típica actual es de 2,3 Gigas. El casete debe reescribirse cada dos años.

**Casete DAT:** normalizado, el casete DAT (Digital Audio Tape) es utilizado principalmente para grabación audio, aunque también se usa en las tecnologías de la información. El casete DAT debe de ser reescrito cada dos años.

Por el momento, existen varios formatos completamente propietarios, sobre todo en el caso de disquetes de alta densidad y de discos duros fijos. La memoria de las tarjetas de crédito, aunque de limitada capacidad, puede convertirse también en un importante soporte de conservación en el futuro.

Los soportes ópticos están evolucionando rápidamente y podrían generalizarse para una conservación a largo plazo. Sin embargo, la actual superioridad de los discos magnéticos en tiempo de acceso da a éstos ventaja a la hora de almacenar grandes volúmenes de información.

### **Disco óptico.**

Permite almacenar una gran cantidad de información y un acceso directo, al contrario que la cinta.

### **Soportes ópticos**

#### Posibilidad

Los principales soportes ópticos son descritos a continuación:

**CD-ROM:** normalizado, es una buena solución para los nuevos archivos.

Hay diferencias entre los CD-ROM estampados (enfocados a la producción de un gran número de copias y muy fiables) y los CD-ROM grabados uno a uno (más baratos para un pequeño número de copias, pero menos fiables, salvo si a continuación se recubren de un baño específico. Los CD-ROM deben reescribirse cada diez o veinte años y a diferencia de las cintas magnéticas, no necesitan un medio ambiente de almacenamiento controlado.

**DVD** (en inglés, *Digital Versatile Disk*- disco digital polivalente-): este nuevo tipo de disco podría convertirse en el futuro en un excelente soporte de archivo. Tiene una gran capacidad (4,7 a 18 Gigas), y debería beneficiarse de una gran difusión. Los lectores de DVD pueden leer igualmente los CD-ROM (aunque sólo los lectores DVD más recientes tienen pueden leer CD-ROM grabados).

**Disco WORM no regrabable** (en inglés: *Write Once Read Many*- única escritura y muchas lecturas-). No existe un estándar, y se encuentran menos extendidos. Deben reescribirse cada diez o veinte años.

**Disco óptico regrabable:** son discos que utilizan en su mayor parte la tecnología magneto-óptica, que combina aspectos de ambas tecnologías de almacenamiento para obtener una mayor rapidez, capacidad de almacenamiento y la posibilidad de regrabar los datos varias veces. Existen pocos estándares en la actualidad. Estos discos deben de ser regrabados cada diez o veinte años.

A menudo, se utilizan programas propietarios para acceder a los datos almacenados. Es importante verificar que es posible un fácil acceso a los formatos de los archivos y a las estructuras jerárquicas.

La rápida evolución de las tecnologías y la limitada perdurabilidad de los soportes electrónicos, aconsejan regrabaciones periódicas de los archivos electrónicos; aunque supongan un coste añadido, permiten resolver muchos problemas planteados por formatos no estandarizados y la propia evolución de la tecnología. No obstante, la mayor parte de soportes magnéticos y ópticos tienen unidades de detección y corrección que permiten reparar errores automáticamente.

Durante cada reescritura, debe optarse por:

- guardar los ficheros antiguos tal y como están.
- migrar hacia un soporte o formato más moderno (véase igualmente la sección 4.3.5., sobre conversión de formatos electrónicos).

No sólo deben tenerse en cuenta los elementos financieros, sino también la accesibilidad, legibilidad, perdurabilidad y la preservación de la autenticidad.

### **Conservar la documentación electrónica.**

Es importante el respetar algunas reglas en lo referente al medio de conservación.

### **Conservación a largo plazo.**

Ejemplo.

Se han desarrollado numerosas discusiones dentro de la ISO, de la ANSI y del CIA sobre la mejor manera de conservar la documentación electrónica. A continuación se enumeran algunos ejemplos prácticos:

- Temperatura media de conservación: 18° C +- 5°.
- Humedad relativa de conservación: 40% +- 5%.
- Frecuencia de regrabación: 10 años.

ISO. Organización Internacional de Normalización.

ANSI. Instituto nacional norteamericano de normalización.

CIA. Consejo Internacional de Archivos.

## 5.2. Formatos de archivos.

### Diversos

#### formatos:

imágenes de mapa de bits, gráficos vectoriales, textos, datos, programas, vídeo y audio, etc.

Existen muchos tipos de normas y estándares en función del tipo de datos que se desee conservar. La mejor manera de proceder es seleccionar un conjunto común de estándares desde el principio, para facilitar así la circulación de la información. Preferiblemente, se utilizarán los mismos formatos para la conservación a corto y a largo plazo.

### Diferentes tipos de formatos.



#### Nociones básicas

Los diversos formatos pueden reagruparse en algunas grandes familias, en función de los objetos que traten:

- **imágenes de mapa de bits:** están constituidas por un conjunto de puntos. El escaneado de un documento, por ejemplo, produce este tipo de imagen. Ocupan un espacio significativo y suelen comprimirse antes de su almacenamiento. Pueden utilizarse como fuente para posteriores codificaciones (de texto o gráficos vectoriales), o grabarse tal cual (p.e., en el caso de una fotografía). El tipo de compresión utilizado define el tipo de formato de archivo de imagen. Los formatos de fax constituyen un caso específico de imágenes comprimidas.
- **gráficos vectoriales:** cuando se utilizan gráficos o imágenes constituidas únicamente por contornos, es posible ahorrar mucho espacio utilizando un formato vectorial que no conserva más que las coordenadas de los vectores utilizados (segmentos de recta, arcos de círculo, etc.). Sin embargo, este formato no es utilizable en el caso de las fotografías.
- **textos:** comprenden, en general, tres aspectos diferentes:
  - el propio texto, que agrupa caracteres codificados.
  - la estructura del texto (por ejemplo, el título, los capítulos, los términos que se desean resaltar, las listas, etc.).
  - los aspectos de presentación (términos resaltados en negrita o en rojo, etc.).
- **datos:** existen diversas posibilidades de guardar datos con sus características originales. p.e., en la actualidad, no existe ningún medio económicamente realista de mantener la capacidad de cálculo de las hojas de cálculo. Para aplicaciones basadas en modelos de información y proceso, las posibilidades de conservación mejoran. La extracción de bases de datos de acuerdo a las normas y modelos explicados en sus instrucciones, es la manera correcta de proceder.
- **programas:** menos independientes de las plataformas que los datos. Las mismas plataformas se vuelven obsoletas cada vez con mayor rapidez.
- **audio, vídeo** y el resto de objetos que pueden incluirse en un documento.

### Orientaciones futuras.

#### Nociones avanzadas

Los documentos y las bases de datos deberían acercarse cada vez más; en el futuro, a la noción de documento compuesto e incluso de documento orientado a objeto, es decir, un documento que reagrupa varios elementos autónomos (texto, imágenes, sonido, vídeo), ligados internamente.

Mientras tanto, los estándares en desarrollo (OLE de Microsoft, OpenDoc de IBM y Apple, o incluso el lenguaje JAVA de Sun) no son lo suficientemente estables para permitir hoy en día la utilización de documentación electrónica que integre sus propias instrucciones de tratamiento (lectura, navegación, etc.).

### 5.2.1. Imágenes de mapa de bits.

#### Compresión de imágenes.

Puede ser: sin pérdida/con pérdida.

Existen dos tipos de compresión de imágenes:

**Compresión sin pérdida:** la imagen, después de haber sido comprimida y descomprimida, es exactamente idéntica a la original. Esto impone proporciones de compresión bajas, en torno a 2:1.

**Compresión con pérdida:** en este caso, las informaciones menos útiles de la imagen no son conservadas. El ojo humano distingue mucho mejor algunas partes de las imágenes que otra. Este tipo de formato permite proporciones de compresión más elevadas, dependiendo del porcentaje de degradación aceptable.

#### Existen muchos formatos de imágenes de mapa de bits.

Es necesaria la elección de uno normalizado y perdurable.

#### Formatos de imágenes de mapa de bits

##### Posibilidad

Los principales formatos de imágenes (incluidos los de fax y vídeo), se describen a continuación:

**TIFF.** (*Tag Image File Format*): se utiliza a menudo en los ficheros generados por escáneres. Existen varias posibilidades según el número de colores escogido. No genera pérdida, pero no permite más que una escasa compresión. La última versión 6.0 de TIFF permite opciones multipágina. Para que el formato conserve una correcta compatibilidad, no deben realizarse compresiones añadidas (por ejemplo, algoritmos de compresión Packbits 32773 del ITU-T, LZW o incluso JPEG).

**GIF.** (*Graphics Interchange Format*): originado en CompuServe, el formato se encuentra muy extendido en Internet. Existen dos especificaciones GIF 87A y GIF 89A. Los navegadores permiten generalmente leer ambos formatos.

**JPEG.** (*Joint Photographic Experts Group*): esta norma internacional es utilizada cada vez más a menudo (también en Internet). Se trata de un formato de compresión con pérdida que permite un alto nivel de compresión. Se trata de una buena opción en términos de memoria y de perdurabilidad.

**Faxes:** existen dos formatos de ficheros de faxes, según se utilice una línea telefónica normal o una línea RSDI.

**Vídeo:** existen dos normas de formatos de vídeo: MPEG-1, orientado a los ordenadores y la multimedia. MPEG-2, más reciente, orientado hacia la televisión digital (incluido el sonido).

**Otros formatos de imágenes:** la utilización de otros formatos propietarios para imágenes (como las BMP o PCX) puede ser peligrosa, pues su conservación no está garantizada a largo plazo. EL Photo-CD de Kodak no se utiliza demasiado en documentos.

## 5.2.2. Archivos de gráficos vectoriales.

### Formatos de gráficos vectoriales.

Permiten conservar la estructura del gráfico.

### Formatos de gráficos vectoriales.

**CGM** (Computer Graphics Metafile): formato normalizado para los gráficos vectoriales, es una buena garantía de perdurabilidad.

**Formatos específicos de algunos tipos de aplicaciones:** algunas aplicaciones necesitan un acercamiento específico a estándares que les son propios. Es el caso, por ejemplo, de los SIG (Sistemas de Información Geográfica) o del CAD (diseño asistido por ordenador).

**Otros formatos de imágenes:** Es preferible evitar la utilización de otros formatos propietarios (como PICT en los MacIntosh o Windows Metafiles de Microsoft, y otros muchos formatos de programas), puesto que no aportan garantía alguna de conservación a largo plazo.

## 5.2.3. Archivos de texto.

### Texto.

Un texto que incluye la estructura, pero no la presentación, es independiente de plataformas.

Existen varios tipos de archivos de texto, según se conserven estructura y/o presentación:

- Texto plano, que no contiene más que un conjunto de caracteres: es un fichero de bajo nivel difícil de manejar a causa de la pérdida de estructura.
- Texto plano asociado a una estructura: muy útil a la hora de su manejo, suministra un fichero independiente del tipo de plataforma que tenga el equipo.
- Texto completo (caracteres, estructura y presentación): no es independiente del equipo utilizado para leerlo (color o blanco y negro; texto previsto para pantalla o impresora...).

### Juegos de caracteres normalizados.

Existen como respuesta a necesidades y culturas variadas.

### Juegos de caracteres codificados

#### Nociones avanzadas

Las principales familias de códigos para la representación de caracteres son tres.

- El ISO 646, equivalente al ASCII. Esta norma codificada en 7 bits no permite acceder a determinados caracteres europeos (por ejemplo, los acentos).
- Juegos de caracteres de 8 bits. Se trata de una ampliación del ISO 646. Hay dos normas útiles para la Unión Europea: ISO/IEC 8859-1, para la Europa Occidental y el ISO/IEC 8859-7, para Grecia. Existen otros juegos para el árabe, el hebreo y el cirílico.
- Juego de caracteres universal (UCS) (*Universal Character Sets*). Este juego normalizado (ISO/IEC 10646), ofrece la mayoría de los caracteres e ideogramas utilizados en todo el mundo mediante dos bits (UCS-2, conocido también con el nombre de Unicode) o de 4 bits (UCS-4). Otro código, con un número variable de bits, denominado UTF, permite el intercambio de caracteres multibits entre diferentes máquinas, puesto que los caracteres de control de 8 bits nunca aparecen en las secuencias de bits (por ejemplo, el código de control de fin de cadena es a menudo 00. En UTF, ningún otro carácter lleva 00 en sus bits, mientras que el carácter "A", por ejemplo, se codifica como 00 64 en UCS-2).

**Presentación.**  
Es preferible conservar sólo la estructura de un texto al estilo asociado con cada elemento.

Muchas características de la presentación de un texto dependen de la plataforma utilizada: por ejemplo, no tiene sentido ofrecer un carácter intermitente para una impresión en papel. Sin embargo, la tecnología evoluciona rápidamente y los textos conservados podrán ser leídos en el futuro con medios aún no concebidos en el presente.

Preservar la presentación (p.e., fuentes tipográficas o macros de un texto) plantea problemas a largo plazo. La mejor solución consiste en no conservar más que la estructura del texto (p.e. "esto es un título de primer nivel"), más que el estilo asociado (p.e., los títulos de primer nivel se representan en Arial/Helvética 16 puntos), dejando a cargo del programa de visualización la elección de un estilo para las diferentes partes del texto.

**Texto estructurado.**  
Existen algunas normas y numerosos formatos propietarios.

#### Texto estructurado.

##### Posibilidad

Existen varias normas y estándares que permiten salvaguardar un texto estructurado.

**SGML** (*Structured General Mark-up Language*). No se trata de una estructura propiamente dicha, sino más bien un lenguaje de marcación (markup language). Esta norma internacional es utilizada cada vez con mayor profusión. Permite conservar un texto y su estructura, pero sin la presentación. Existen normas adicionales, como:

- DSSSL: semántica y especificaciones de las hojas de estilo (*Document Style Semantics and Specification Language*).
- SPDL: lenguaje de descripción de página normalizado basado en el PostScript (*Standard Page Description Language*).
- SDIF: formato de intercambio de documentos SGML (*SGML Document Interchange Formats*).
- Intercambio de información de fuentes tipográficas (FIPS, *Font Information Interchange*).
- DTD: definición de la estructura de los documentos (existen varios según los tipos de documentos) (*Documento Type Definition*).

**HyTime**: extensión de SGML que permite la inclusión de objetos multimedia en documentos compuestos.

**HTML**: lenguaje de marcación Hipertexto (*Hyper Text Mark-up Language*). Este lenguaje de marcación se presenta como una versión simplificada del SGLM. Especialmente extendido en Internet, es aún sin embargo poco estable y escasamente adaptado a documentos largos. Para la conservación a largo plazo de los documentos, es preferible la utilización de SGML.

**ODA**: (*Office Document Architecture*). Esta norma internacional permite agrupar en un mismo fichero el texto, la estructura y la presentación. Ha sido desarrollada para ser utilizada en el ámbito de la ofimática, y no es independiente de la plataforma utilizada.

**RTF** (*Rich Text Format*): formato utilizado principalmente por el conjunto de aplicaciones Office de Microsoft. No aporta ninguna garantía de estabilidad ni de perdurabilidad, por lo que es preferible utilizar filtros de exportación o herramientas de conversión a formatos estándares.

**PostScript:** norma de descripción de página creada por Adobe, muy utilizada para enviar o imprimir textos con su presentación. Debe reemplazarse por normas abiertas.

**PDF** (*portable document format*): Permite únicamente consultar documentos en diversas plataformas. Existen muchos otros formatos propietarios.

### EDI y mensajería electrónica.

Dos aplicaciones específicas que utilizan estructuras específicas.

### Aplicaciones específicas

#### Nociones avanzadas

La estructura de los textos puede ser específica, con campos suplementarios para facilitar la indización. Es el caso concreto de los **mensajes electrónicos**.

Ya han surgido algunas normas para facilitar el intercambio de informaciones estructuradas, generalmente para intercambios comerciales. Se trata del **EDI** (intercambio de datos informáticos-*Electronic Data Interchange*), principalmente con la norma **Edifact**. Además de las definiciones genéricas de EDI, han sido creadas otras EDI particulares para diversos sectores (banca, automóvil...). Aunque se trata de algo interesante para intercambios formales entre organismos, su normalización es lenta y no siempre se tiene en cuenta la rápida evolución tecnológica (imágenes, sonido y vídeo).

### 5.2.4. Datos y programas.

#### Datos y programas.

Plantean un problema complejo, puesto que no existe un formato estándar ampliamente utilizado.

Hoy en día no existe un formato estándar de alto nivel para los ficheros de hojas de cálculo y bases de datos. Esto trae como consecuencia que para tener la seguridad de poder leer los datos tras un largo período, es necesario disponer de una herramienta capaz de leer el formato antiguo, o conservar el propio programa.

No debe confundirse el interfaz entre el programa y los datos (por ejemplo, SQL en el caso de una base de datos) y el formato del fichero donde se conservan los datos.

El problema estriba en que los programas son menos independientes de las plataformas que los datos.

Cuando no existe un formato estándar para un tipo de datos, la mejor solución consiste en:

- utilizar un formato propietario muy extendido que pueda leído por numerosos programas.
- prever una estrategia de migración (o de conservación del programa con sus datos). La sección 4.3. da algunas pautas al respecto.

La conservación de los programas plantea un problema similar. Es entonces necesario conservar el programa fuente o una plataforma capaz de hacerlo funcionar.

## Diferentes tipos de datos

dan lugar a diferentes posibilidades: hoja de cálculo, base de datos, contabilidad y gestión, formulario, firma electrónica.

## Ficheros de datos

### Posibilidad

Los datos pueden ser conservados en diferentes tipos de ficheros. Los principales se detallan a continuación.

**Hoja de cálculo:** el formato de fichero EXCEL puede ser considerado como un estándar propietario de hecho. Quizás la inclusión de tablas en SGML sea la primera etapa hacia una norma abierta.

**Base de datos:** aunque cada vez en mayor medida las bases de datos utilicen SQL como lenguaje de consulta, éstas carecen de un formato de alto nivel para su conservación. La mejor solución consiste en utilizar una base de datos muy extendida o salvaguardar los datos como texto plano con separadores de campo (lo que permitirá a otra aplicación recrear la base de datos). En este último caso, debe salvaguardarse igualmente la estructura de la base de datos.

**Edifact:** (*Electronic Data Interchange For Administration, Commerce and Transport*- intercambio electrónico de datos para la administración, el comercio y el transporte): varios tipos de mensaje permiten el intercambio de archivos de contabilidad y facturas.

**STEP:** (*Standard for The Exchange of Product data*-norma para el intercambio de datos sobre productos): este conjunto de normas facilita el intercambio de datos de productos industriales. Utiliza el lenguaje EXPRESS para la representación e intercambio de datos sobre productos. Son normas muy utilizadas.

**IDEF** (*Function and Information Modelling*-modelización de funciones y de información). IDEF0 es la metodología utilizada para determinar el ámbito de cada protocolo de aplicación en la norma STEP. Se utiliza también en la ingeniería de procesos de trabajo. IDEF3 está consagrada a la mejora de los procesos de trabajo y suministra formatos gráficos estándares para los diagramas de flujo de tareas.

**Formulario:** en este caso, es posible no conservar más que la información introducida y una copia de un formulario en blanco. Existen iniciativas para alcanzar un estándar para los formularios (FIMS, HTML 3.0...).

**Firma electrónica:** existen dos formatos de firmas encriptadas: DES y RSA. Ambas pueden ser utilizadas para autenticar documentos. La sección 6.3.1. describe los procedimientos de encriptación.

## 6. Acceso y difusión de la información.

### 6.1. Hacia la sociedad de la información.

#### La sociedad de la información.

Da a la información electrónica un papel de primer orden.

La Comisión Europea ha adoptado el concepto de “sociedad de la información”, una sociedad en la que la información electrónica jugará un papel clave.

#### La sociedad de la Información.

##### Nociones avanzadas

Jacques Delors -por entonces Presidente de la Comisión Europea-, publicó en enero de 1994 su *Libro Blanco sobre el crecimiento, la competitividad y el empleo*.

Al mismo tiempo, encomendó a Martin Bangemann, Comisario Europeo de Industria, la preparación de un informe con la ayuda de un grupo de alto nivel formado por dirigentes de industrias de las tecnologías de la información, las telecomunicaciones y los “media”. Este informe, “Europa y la sociedad de la información”, tomaba en consideración no solo las infraestructuras, sino también servicios, aplicaciones, y los aspectos sociales.

En febrero de 1995, una primera reunión del G7 destinada a la sociedad de la información planteó el establecimiento de once proyectos piloto, incluyendo el del “gobierno en línea”, que propone acercar los gobiernos a los ciudadanos gracias a las nuevas tecnologías.

La Comisión continúa construyendo la sociedad de la información en Europa, y ha contribuido ya a sensibilizar al público y a los responsables de la toma de decisiones. Con esta finalidad, ha puesto en marcha diferentes iniciativas, como su “plan de acción” que enumera las acciones más importantes, en especial las medidas de tipo legal requeridas para conseguir la puesta en funcionamiento de la sociedad de la información en Europa, cumpliendo cuatro funciones principales, a saber:

- servir de herramienta de orientación para los estados miembros y otras instituciones europeas (claridad y transparencia).
- suministrar informaciones detalladas sobre el desarrollo del marco reglamentario necesario.
- suministrar un servicio de información abierto a todas las partes interesadas, especialmente al público, sobre la política de la UE, y permitir de esta manera un diálogo constructivo entre la Comisión y todos los segmentos pertinentes de la sociedad, y
- suministrar una herramienta de gestión interna para la Comisión.

Es fundamental que la información electrónica no sólo sea conservada, sino también accesible lo más fácilmente posible, para favorecer así su mejor circulación: además de los organismos públicos, el ciudadano puede acceder con mayor comodidad a las informaciones públicas en las que está interesado. Pero incluso tras su uso corriente, la información electrónica puede utilizarse en el marco de investigaciones estadísticas, científicas o históricas.

### **Los archiveros y la sociedad de la información.**

Los archiveros son los guardianes de la memoria colectiva de la sociedad de la información.

La profesión de archivero, así como otras relacionadas con la información, forman parte fundamental y activa de la moderna sociedad de la información. Son requeridas a fin de conservar la memoria colectiva de la sociedad de la información -el Forum DLM sobre documentación electrónica (Bruselas, 18 a 20 diciembre 1996) fue el primer acercamiento interdisciplinar en este sentido-.

### **Vida privada y apertura.**

Un conflicto de intereses recogido en las legislaciones nacionales y comunitarias.

Una de las dificultades a la hora de difundir información electrónica, es identificar la información que debe conservarse o agruparse, pues ésta puede ser tratada de un modo no previsto por su autor: por ejemplo, los componentes de una base de datos pueden utilizarse en investigaciones demográficas con criterios particulares de selección.

Existe un permanente conflicto de intereses entre el derecho a la vida privada y el de apertura. La legislación fija la frontera entre ambas, por lo que es importante verificar tanto las directivas europeas como la legislación nacional.

## **Información electrónica y personas con problemas de visión.**

### Nociones avanzadas

Las personas con problemas de lectura pueden recurrir a dos métodos diferentes para leer información electrónica:

1) La información electrónica no supone un problema para ciegos y personas con problemas de visión, puesto que éstos pueden utilizar programas informáticos estándares (tratamientos de texto, hojas de cálculo, bases de datos...) mientras otro programa que aumenta las pantallas o suministra elementos suplementarios (braille, voz sintética) trabaja en segundo plano.

Estos "lectores de pantalla" que aumentan el contenido textual de una pantalla o que lo transforman temporalmente en discurso sintético o en braille (por ejemplo, un dispositivo con puntos tangibles alineados que reproducen caracteres braille bajo control del ordenador), están disponibles en la actualidad en todas las plataformas informáticas.

La lectura de documentos largos y no estructurados sigue siendo sin embargo difícil, ya que las personas con problemas de visión no pueden tener una visión general de una página o de un documento.

2) Ambos grupos pueden leer mucho más rápido mediante la utilización de documentos estructurados y de programas especiales de lectura, (=navegadores 3D) que envían directamente la información audio, hablada y en braille a dispositivos de salida (sintetizador o línea de lectura en braille). Además, el texto en la pantalla puede ser reformateado y presentado línea a línea para que su lectura sea así fácil para las personas con problemas de visión. Estos programas de lectura existen para SGML y HTML.

## **6.2. Normas y estándares para el acceso a los datos.**

### **Difusión de la información.**

Deben seleccionarse normas y estándares que faciliten el acceso a los datos.

Algunos estándares y normas están mejor adaptados para hacer accesible la información al mayor número de personas posible. Algunos, ya fueron enumerados en el capítulo 5 en relación a los estándares de conservación. La selección ahora ofrecida presenta los más adaptados para la difusión de la información. Conservación y difusión pueden realizarse, si se considera necesario, mediante formatos diferentes.

## Estándares para la difusión.

Posibilidad

### **Soportes para la difusión de información:**

- disquete 3 ½ de alta densidad 1,44 Kb.
- CD-ROM estándar.
- en el futuro, el DVD (*Digital Versatile Disk*) puede constituir una posibilidad añadida.

### **Protocolos de Internet para la comunicación e intercambio de documentos en la red.**

- TCP/IP (*Transfer Control Protocol, Internet Protocol*). Ambos protocolos permiten intercambios por medio de una red. Son muy utilizados en todo el mundo en las redes Internet e Intranet.
- HTTP (*Hypertext Transport Protocol*), para los servidores Internet.
- FTP (*File Transport Protocol*), para los servidores de ficheros.
- SMTP (*Simple Mail Transfer Protocols*) para la mensajería electrónica.
- MIME (*Multipurpose Internet Multimedia Extensions*), permite la utilización de varios formatos en el mismo texto (por ejemplo, texto con diferentes juegos de caracteres, imágenes...).

### **Formatos para los documentos.**

- SGML para documentos compatibles. El formato HTML es mejor para documentos cortos, como las páginas de bienvenida de lugares de Internet (*home page*). Los documentos largos pueden realizarse directamente en formato SGML en los servidores de Internet.
- Formatos propietarios de tratamientos de texto (Microsoft Word, WordPerfect). Aunque estos formatos no estén recomendados para una conservación a largo plazo, constituyen estándares de hecho muy utilizados en el ámbito Windows de los Pcs.

### **Interfaces de acceso a bases de datos.**

- SQL 2 (*Structured Query Lenguaje*) para las bases de datos relacionales.
- ISAM (*Indexed Sequential Access Method*) para los ficheros secuenciales indizados (interfaz de bajo nivel).

## **6.3. Seguridad.**

### **6.3.1. Derechos de acceso.**

Existen numerosos tipos de acceso a la información electrónica para los usuarios:

- derecho de acceso a la página inicial.
- derecho de acceso a la totalidad o a parte de un documento.
- derecho de visualización y de impresión o únicamente derecho de visualización.
- etc.

### Convertir un fichero en anónimo.

Una tarea más compleja de lo que parece.

Los derechos de acceso deben ser estudiados con cuidado para conservar el derecho a la vida privada y el anonimato. existen dos modos de transformar un fichero en anónimo para llevarlo a dominio público:

- Supresión de ciertos campos (por ejemplo, nombres).
- Combinar informaciones para producir estadísticas.

#### ¿Anónimo o no?.

##### Ejemplo

Deben tomarse algunas precauciones para asegurar que el procedimiento utilizado para convertir un fichero en anónimo no permite búsquedas selectivas que recuperen un pequeño número de documentos.

Supóngase, por ejemplo, la supresión de los apellidos de una base de datos. No obstante, pueden localizarse personas por criterios profesionales dentro de un lugar determinado: mediante búsquedas cruzadas, es posible encontrar una persona individualizada y descubrir informaciones añadidas sobre ella (como sus cifras salariales): pueden, por ejemplo, existir pocos cardiólogos en un pueblo de 1000 habitantes...

### 6.3.2. Encriptación y autenticación.

#### Encriptar los datos:

Para asegurar la confidencialidad.  
Para autenticar el documento.

La encriptación de datos puede tener dos finalidades muy diferentes:

- No permitir la lectura de un documento más que a determinadas personas.
- Garantizar que el documento ha sido correctamente realizado por una persona particular y no ha sido modificado por alguna otra.

En materia de encriptación, es importante tomar en consideración las especificidades nacionales, puesto que la legislación en materia de encriptación es diferente en cada país: pueden existir restricciones a la exportación de programas de encriptación (como en los USA) , o restricciones en su uso (como en Francia).

No se trata simplemente de acceder a la información o de garantizar un cierto nivel de confidencialidad. Es igualmente necesario asegurar que la información a la que se accede está bien identificada. La autenticación de los documentos es un elemento importante para obtener una información correcta.

#### Existen dos tipos de algoritmos de encriptación:

- de clave simple, tipo DES.
- de doble clave, tipo RSA.

#### Algoritmos de encriptación.

##### Nociones avanzadas

Existen dos grandes familias de algoritmos de encriptación:

#### **Algoritmos de clave simple (simétricos):**

Tipo DES (*Data Encryption System*) (NIST FIPS 46-1, cercano al ISO 8227-DEA).

En este caso, la misma clave de encriptación sirve para codificar y decodificar los datos. Emisor y receptor de un mensaje codificado disponen de la misma clave.

#### **Algoritmos de doble clave (asimétricos):**

Tipo RSA (*Rivest, Schamir & Adelman*, apellidos de sus inventores).

En este caso, existen dos claves para cada persona.

Una que se guarda en secreto (la clave privada) y otra (clave pública) que se difunde de modo general. Un mensaje codificado con una de estas claves no puede ser descodificado más que con la segunda clave.

En el caso de un **mensaje confidencial**, el emisor envía el mensaje codificado con la clave pública del receptor. Sólo éste podrá descifrarlo con la ayuda de su clave privada.

En el caso de un **mensaje autenticado**, el emisor envía el mensaje codificado con su propia clave privada. Toda persona que tenga acceso a su clave pública podrá descifrar el mensaje con la garantía de que ha sido correctamente enviado por el emisor y de que nadie ha podido modificarlo.

El programa **PGP** (*Pretty Good Privacy*) de Philip Zimmermann es un programa de encriptación de tipo RSA de dominio público en Internet. Y ello, a pesar de la prohibición existente en los USA de la exportación de algoritmos que utilicen claves lo suficientemente largas como para permitir una buena seguridad.

Existen servidores de autenticación. Permiten el obtener claves públicas personales mediante un procedimiento que garantiza su autenticidad.

#### 6.4. Acceso a los datos.

**Hay varias maneras de permitir a los usuarios el acceso a los datos.**

Existen varias maneras de dar acceso a la información electrónica:

- Salas de lectura electrónica, es decir, salas puestas a disposición del público, con equipos que permiten la lectura de la documentación electrónica.
- Copia de la información electrónica sin modificación alguna, para su utilización en el propio domicilio.
- Realización y mantenimiento de una "copia de consulta" de la información electrónica en un nuevo formato más fácil para los usuarios (el acceso puede hacerse por consulta en el sitio, por mediación de una copia dada al usuario, o en línea).
- Utilización de un modelo de acceso general basado en metadatos para generar automáticamente una visión de la información electrónica útil para el usuario (el acceso puede hacerse por consulta en el sitio, por mediación de una copia dada al usuario, o en línea).

Los dos últimos casos son válidos para una política de difusión en Internet.

#### Alemania.

##### Ejemplo

En Alemania, se diferencia entre bases de datos suministradas en un formato adaptado para la consulta por parte de los investigadores o "copias de investigación", y las "copias para archivo", conservadas en un formato de fichero plano para evitar los problemas de los formatos estándares.

**Difundir la información:** puesta a disposición de la información (difusión pasiva) / envío selectivo de información (difusión activa).

La difusión de la información necesita también una política de información y de toma de conciencia del grupo de potenciales usuarios. Existen dos estrategias que pueden ser complementarias:

- Dejar disponible la información, dejando al usuario la iniciativa de su localización gracias a las herramientas de navegación en línea (difusión pasiva).
- Enviar una determinada información a un grupo seleccionado de usuarios (difusión activa).

Ambas estrategias pueden combinarse dando detalles a un grupo determinado sobre el lugar en el que pueden localizar la información.

La puesta en práctica de una política de difusión es indispensable para permitir a los usuarios el acceso a la información.

## **7. Conclusiones.**

Esta guía ha propuesto algunos ejemplos de las soluciones actualmente utilizadas para tratar la información electrónica, así como algunas sugerencias para definir una estrategia en este ámbito. No existe una única política válida para todos los países y todos los casos. Del mismo modo, es necesario que cada cual defina su propia estrategia en este ámbito con todos los involucrados. Si esta guía aporta alguna ayuda en esta tarea, habrá cumplido su propósito.

## **8. Anexos.**

Anexo 1- Terminología.

Anexo 2- Preguntas en suspenso.

Anexo 3- Para entender la normalización.

Anexo 4. Ejemplos de tests para definir una estrategia en relación a la información electrónica.

Anexo 5. Modelo ideal: ¿Qué metadatos es necesario crear?.

Anexo 6. Modelo ideal: ¿Cómo seleccionar normas y estándares correctos?.

Anexo 7. Índice.

Anexo 8. Índice de las ilustraciones.

Anexo 9. Bibliografía.

### **8.1. Terminología.**

Las definiciones siguientes son utilizables dentro del marco de esta guía. Existen definiciones dadas en las respectivas legislaciones nacionales que deben también ser tenidas en cuenta.

#### **Información.**

Ítem o elemento de conocimiento que puede ser transmitido.

#### **Dato.**

Representación de una información elemental en un formato que permite su proceso.

#### **Datos legibles por máquina (DLM).**

Datos en un formato adecuado para su recuperación, proceso y transmisión mediante un ordenador.

#### **Documento.**

Grupo consistente y coherente de datos, estructurado para presentar un argumento o para servir de prueba de la ejecución de una acción.

**Base de datos.**

Grupo de datos estructurado para facilitar su consulta y posteriores tratamientos.

**Records/Enregistrement.**

Conjunto consistente de datos registrados en un soporte<sup>3</sup>.

**Documento electrónico.**

Documento cuya información está registrada en un formato adecuado para la consulta, tratamiento y transmisión por un equipo informático.

**Soporte.**

Elemento que permite el registro, almacenamiento y recuperación de los documentos.

**8.2. Preguntas en suspenso.****1. ¿Cuál es el valor legal de la información electrónica?**

La noción de "original" utilizada respecto a los documentos en soporte papel se vuelve compleja en el caso de la información electrónica. Cada vez es más fácil duplicar los documentos, de tal manera que ya no puede diferenciarse entre original y copia. Además, un documento electrónico puede estar constituido únicamente por enlaces o relaciones con otros documentos electrónicos.

Una de las soluciones para abordar este problema es la encriptación con claves públicas y privadas (véase la sección 6.3.2. en relación a la encriptación de datos). El autor encripta su documento con su clave privada, y cualquier persona puede recuperar la clave pública en un servidor de autenticación para descifrar el documento con la garantía de que éste no haya sido modificado desde su encriptación.

**2. El significado de los términos puede variar de un país a otro.**

La traducción de los diferentes términos no es suficiente, pues el problema del vocabulario es ciertamente complejo. Una solución consistiría en crear una lista de conceptos con los términos adecuados en cada lengua.

**8.3. Para entender la normalización.****• Tres niveles en el grado de madurez de una tecnología: estándares de hecho, PAS, norma.**

Para permitir la libre competencia y garantizar una adecuada compatibilidad de los datos sea quien sea el suministrador, es importante especificar solamente las interfaces y no los propios productos. Así, varios suministradores pueden proponer productos compatibles. Datos y aplicaciones ganan en duración de conservación.

Existen varios niveles que definen la madurez de una tecnología desde los productos propietarios hasta las normas abiertas:

- **los estándares de facto** (*de facto standards*). Cuando un producto adquiere una gran importancia dentro del mercado, la compatibilidad de otras aplicaciones y datos es medida en relación al producto, y la evolución de la compatibilidad depende del fabricante (por ejemplo, el tratamiento de textos Word de Microsoft).
- **Las especificaciones disponibles públicamente** (*PAS- Publicly Available Specifications*). en ocasiones, varias empresas representativas en el mercado pueden reunirse en consorcio para definir un estándar de interfaz. Esta definición permite el desarrollo de diversos productos compatibles entre ellos (por ejemplo, las especificaciones de X/Open o de IETF).

---

<sup>3</sup>"Un morceau particulier d'information enregistrée générée, collectée ou reçue lors du commencement, de la conduite ou de l'achèvement d'une activité institutionnelle ou personnelle, et qui comprend un structure, un contenu ainsi qu'un contexte suffisants pour fournir une preuve ou une évidence de cette activité" (CIA).

- **Las normas** (*de jure standards*). Los organismos oficiales permiten garantizar un consenso en una determinada especificación. En este caso, la especificación estándar se denomina “norma” (por ejemplo, las normas de juego de caracteres ISO).

Una tecnología, a medida que va madurando, da lugar al surgimiento de estándares de hecho, de PAS y finalmente de normas.

- **Organismos internacionales de normalización en el ámbito de la informática: ISO/IEC, ITU-T.**

Existen varios organismos oficiales internacionales:

- ISO: Organización Internacional de normalización; abarca un gran número de ámbitos.
- CEI o IEC: Comité Electrotécnico Internacional.

Estos dos organismos han creado un comité conjunto para realizar normas en el ámbito del tratamiento de la información.

-UIT o ITU: Unión Internacional de Telecomunicaciones. Su comité ITU-T se ocupa específicamente del ámbito de las telecomunicaciones (se trata de la nueva denominación del CCITT).

El procedimiento de elaboración de normas abarca varias etapas. En ocasiones, es posible empezar a utilizar una norma desde la penúltima etapa denominada “proyecto de norma internacional” (*Draft International Standard-DIS*).

- **Organismos europeos de normalización en el ámbito informático: CEN/Cenelec, ETSI.**

A nivel europeo, existen organismos equivalentes:

- CEN: Comité Europeo de Normalización.
- Cenelec: Comité europeo de normalización electrotécnica.
- ETSI: Instituto Europeo de Normalización de las Telecomunicaciones.

Los dos primeros se ocupan en común de las tecnologías de la información. Crean las Normas Europeas (**EN**) y las Prenormas Europeas (**ENV**). Se trata a menudo de adaptaciones de normas internacionales. Existe sin embargo una diferencia en cuanto al nivel legal: la decisión 87-95 del Consejo de las Comunidades Europeas, de 22 de diciembre de 1986, hace obligatoria la referencia y cumplimiento de las normas europeas, mientras que las normas internacionales son de cumplimiento voluntario.

Estos organismos oficiales desarrollan o establecen normas, pero también perfiles (conjuntos de estándares con diferentes opciones para permitir una mejor interoperatividad). Los perfiles internacionales desarrollados por ISO/IEC se denominan ISP (*International Standardised Profiles*).

Existen también otros organismos desarrolladores de especificaciones:

- X/Open, que reúne sus especificaciones en la X/Open Portability Guide (XPG).
- IETF (*Internet Ingeeniering Task Force*), que desarrolla especificaciones Internet tras realizar “llamadas para sugerencias” (*Request for Comments- RFC*).
- NIST del gobierno de los USA (*National Institute for Standards and Technologies*), que desarrolla perfiles denominados FIPS (*Federal Information Processing Standards*).
- Muchos otros organismos, trabajan también en numerosos aspectos de los estándares para el tratamiento de la información.

El cuadro siguiente presenta las referencias de todas las normas citadas en esta guía.

Denominación	Norma o perfil internacional	Norma o perfil europeo	Otras especificaciones	Notas
<b>Soportes de almacenamiento (secciones 5.1 y 5.2)</b>				
Disquete 3 ½	ISO/IEC 9529-1	EN 9529-1		
	ISO/IEC 9529-2	EN 9529-2		
Cartucho ¼	ISO 8462-1			
	ISO 8462-2			
Cartucho ½				
Cinta 1600 bpi	ISO/IEC 3788:1976			
Cinta 6250 bpi				
Casete 8 mm	ISO/IEC 11319			
	ISO/IEC 12246			
Casete DAT				
CD-ROM	ISO 9660			
	ISO 10149			
WORM				
TMO				
DVD				En preparación
<b>Formatos de ficheros de imágenes de mapa de bits y gráficos (sección 5.2)</b>				
Imágenes TIFF				
Imágenes GIF				
Imágenes JPEG				
Fax grupo III	UIT-T Grupo III			Antiguo CCITT
Fax grupo IV	UIT-T Grupo IV			Antiguo CCITT
Vídeo MPEG-1				
Vídeo MPEG-2				
Gráficos CGM	ISO/IEC 8632			
Gráficos CAD				
Gráficos GIS				
Juegos de caracteres (sección 5.2.3.)				
7-bits	ISO 646			
8-bits	Europa	ISO/IEC 8859-1		
Occidental				
8-bits Grecia	ISO/IEC 8859-7			
Multibit	ISO/IEC 10646			
<b>Texto estructurado (sección 5.2.3.)</b>				
SGML	ISO/IEC 8879			
DSSSL	DIS 10179			
SPDL	ISO/IEC 10180			
SDIF	ISO/IEC 9069			
Font Information Interchange	ISO/IEC 9541			
DTD normalizado	ISO/IEC 12083			
HyTime	ISO/IEC 10744			
HTML			W3C HTML 3.0	
ODA/ODIFF	ISO 8613	EN 41509		
	FOD 26	EN 41515		
<b>Formatos de datos (sección 5.2.4)</b>				
EDIFACT	ISO/IEC 9735	EN 29735		
STEP/ Express	ISO 10303			
IDEFO&3			IDEF	
FIMS				
<b>Protocolos de comunicaciones (sección 6.2)</b>				
HTTP			IETF RFC	
FTP			IETF RFC	
			Lenguaje de consulta a bases de datos	
SQL	ISO/IEC 9075			Versión II
ISAM				
<b>Algoritmos de encriptación (sección 6.3.2)</b>				
DAS	ISO 8273			
DES			NIST FIPS 46-1	Cercano al DAS
RSA				

#### **8.4. Lista de tests para los DLM.**

Este anexo presenta las diferentes cuestiones que pueden plantearse a la hora de definir una estrategia para la información electrónica. Las cifras entre paréntesis reenvían a las secciones de la guía que tratan este asunto. Las próximas revisiones incorporarán ejemplos con experiencias y cláusulas de compras.

El organismo no deberá tener obligatoriamente una respuesta para cada punto, ni forzosamente tratar todos los puntos (por ejemplo, la puesta en marcha o la adopción de un thesaurus). La lista de los puntos no tratados, irresueltos, no fijados o rechazados, es también una información importante que forma parte integrante de la estrategia.

##### I. Estrategia general.

###### A. Identificación de todos los implicados (4.1).

Puesta en marcha de un grupo estratégico multidisciplinar

###### B. Lista de términos y conceptos comunes (8.1;8.2)

###### C. Necesidades de los usuarios: identificación y seguimiento (4.1)

###### D. Política sobre el valor legal de los documentos(8.2)

###### E. Política de información y de formación de los servicios (4.1)

###### F. Tecnología antigua (soportes, documentos...) (5.1;5.2)

##### II. Gestión de la información electrónica.

###### A. Identificación de las responsabilidades (durante cada transferencia) (3.2)

1. *Responsabilidad en la gestión de la información electrónica.*

2. *Responsabilidad en la conservación de la información electrónica.*

###### B. Identificación y registro de la información electrónica importante (4.1;4.2)

(tamaño y delimitación respecto a otra información electrónica).

###### C. Política de documentación contextual de la información electrónica (2.1:4.2)

###### D. Definición de un plan de evaluación (4.3.6)

1. *Procedimiento para la información electrónica de proveniencia dudosa (4.2)*

2. *Procedimiento de aprobación para la destrucción o alteración de datos (4.2)*

###### E. Definición de reglas para el sistema de clasificación (4.1;4.4).

1. *Realización de una clasificación por palabras claves (thesaurus)*

###### F. Estrategia de encriptación (4.3.2)

1. *Confidencialidad.*

2. *Autenticación.*

###### G. Política de transferencia física de la información electrónica (4.5).

1. *Lista de elementos que deben transferirse*

2. *Lista de los controles que deben efectuarse en cada transferencia.*

3. *Soporte utilizado para las transferencias físicas, soporte original, red asegurada...*

### III. Conservación (preservación) de la información electrónica.

A. Decisión de conservar los "originales" de la información electrónica sólo en este soporte (4.2), o también en copia en soporte papel.

B. Elección de los soportes de conservación. (5.1).

1. *Conservación a corto plazo.*

2. *Conservación a largo plazo.*

3. *Medio ambiente de conservación (temperatura, humedad, periodización en la realización de copias de seguridad...).*

C. Elección de los formatos de ficheros conservados (5.2)

1. *Ficheros de imágenes de mapa de bits (5.2.1).*

a) *Tipo de compresión utilizada (5.2.1).*

b) *Conservación de los faxes (4.3.4.; 5.2.1).*

2. *Ficheros gráficos (5.2.2.)*

a) *Ficheros gráficos específicos (CAD, GIS)*

3. *Ficheros audio, vídeo y multimedia (5.2.1)*

4. *Ficheros de texto (5.2.1).*

a) *Juegos de caracteres (por defecto, aceptables).*

b) *Formatos de texto estructurado.*

c) *Conservación o no de la presentación.*

5. *Ficheros de datos (5.2.4).*

a) *Bases de datos (bajo nivel o propietario).*

1) *Ficheros de compatibilidad/gestión.*

b) *Hojas de cálculo (bajo nivel o propietario).*

c) *Formularios (5.2.4).*

d) *Documentos compuestos (4.3).*

6. *Programas (5.2.4).*

D. Política de conservación de viejos sistemas y programas (4.3.6).

(si es necesario) Mantenimiento, documentación, experiencia...

### IV. Estrategia de recuperación de documentación en soporte papel.

A. Opciones para la conservación de documentación en soporte papel (4.2).

B. Evaluación de la calidad de los documentos seleccionados para ser escaneados (4.3.3).

C. Estrategia para el reconocimiento de caracteres (4.3.4).

1. *Opciones en la utilización del reconocimiento óptico de caracteres (4.3.4).*

2. *Procedimiento en el reconocimiento óptico de caracteres (4.3.4).*

D. Estrategia de vectorización de los gráficos (4.3.4).

E. Procedimiento para la unificación de los diversos elementos que componen una información electrónica (texto, imágenes, gráficos...) (4.3.4)

### V. Conversión o conservación del formato de datos (4.3.3).

A. Política de documentación de los formatos de datos utilizados por los sistemas y las aplicaciones (2.4.2).

B. Opciones de conversión o conservación de los formatos antiguos (4.3.6).

C. Opciones para añadir una estructura a los textos o a los datos recuperados (4.3.4).

D. Opciones de construcción de una base anónima para la difusión (6.3.1).

1. *Procedimiento de control de la protección del anonimato.*

E. Procedimiento para el análisis de las pérdidas de información debidas a la conversión (4.3.3).

VI. Acceso y utilización de los datos.

A. Política de derechos de acceso (en visualización e impresión) (6.3.1).

B. Puesta a disposición de las informaciones (difusión pasiva) (6.4).

1. *Sala de lectura electrónica* (6.4).

a) *Estándares de consulta* (6.2).

2. *Copia de información electrónica sin modificaciones para su difusión* (6.4).

a) *Soporte de las copias* (6.2).

3. *Copia de la información electrónica en un determinado formato para su difusión* (6.4).

a) *Soportes, protocolos, lenguajes y formatos* (6.2)

b) *Estrategia Internet* (6.2).

4. *Modelo de acceso para generar automáticamente un formato de fácil utilización para el usuario* (6.4).

a) *Soportes, protocolos, lenguajes y formatos* (6.2).

b) *Estrategia Internet* (6.2).

C. Política de promoción (difusión activa) (6.4).

### **8.5. Modelo ideal: ¿qué metadatos es necesario crear?**

Este anexo presenta un ejemplo de posibles metadatos basados en la propuesta realizada por Dublin Core en diciembre de 1996. Pueden encontrarse informaciones añadidas a este respecto en la dirección siguiente de Internet: [http://www.purl.org/metadata/dublin\\_core](http://www.purl.org/metadata/dublin_core).

Esta propuesta no es la única manera de proceder, y no supone sino una serie de ejemplos que pueden ayudar al lector a definir una estrategia propia. Los quince elementos son optativos y la lista es extensible.

#### TÍTULO.

La denominación dada al recurso por el creador o el editor.

#### CREADOR.

(Autor o creador). La(s) persona(s) responsable(s) en primera instancia del contenido intelectual del recurso. Por ejemplo, los autores para los documentos escritos, los artistas, los fotógrafos y los ilustradores para los recursos visuales.

#### TEMA.

(Tema y palabras claves). El tema del recurso, o las palabras claves o las frases que describen el tema o contenido del recurso. La definición de este elemento tiene como meta el promover la utilización de vocabularios y palabras claves controlados. Podría incluir datos de clasificación según un esquema normalizado (por ejemplo, los códigos de clasificación de la Biblioteca del Congreso o los de la clasificación Dewey) o vocabularios controlados según un esquema normalizado (como los encabezamientos de temas médicos o los descriptores del thesaurus sobre arte y arquitectura).

#### DESCRIPCIÓN.

Se trata de una descripción textual del contenido del recurso, incluyendo resúmenes en el caso de documentos o descripciones de contenido cuando se trata de recursos visuales. Las colecciones futuras de metadatos podrían incluir una descripción informática del contenido (análisis espectral de un recurso visual, por ejemplo), que no puede ser integrable en los sistemas de redes actuales. En este caso, el campo podría contener un enlace con esta descripción, con preferencia a la propia descripción.

#### EDITOR.

La entidad responsable de la puesta en disposición del recurso en su forma actual, como un editor, una facultad universitaria o una entidad institucional. La especificación de este campo tiene por finalidad el identificar a la entidad que da acceso al recurso.

#### COLABORADORES.

La(s) persona(s) u organismos(s) -además que la(s) especificada(s) como "Creador"- que aportaron contribuciones intelectuales importantes al recurso, pero de un modo secundario respecto a las personas o entidades definidas como "Creador" (por ejemplo, redactores,. transcritores, ilustradores y organizadores).

#### FECHA.

Aquella en la que se puede disponer del recurso en su forma actual. Es recomendable presentarla en forma de un número de ocho cifras en secuencia año-mes-día, de la siguiente manera: YYYYMMDD, tal y como está definido por la norma ANSI X3.30-1985. De esta manera, la fecha del día en el que se escribió esto sería la de 19961203, o el 3 de diciembre de 1996. Existen también otras posibilidades, pero en el caso de que se utilicen, deben identificarse sin ninguna ambigüedad posible.

#### TIPO.

La categoría del recurso (página de inicio, novela, poema, documento de trabajo, prueba, informe técnico, ensayo, diccionario). El tipo de recurso debería ser elegido de entre una lista de tipos. En la dirección (URL) siguiente, se encuentra una propuesta preliminar de tipos: <http://www.roads.lut.ac.uk/Metadata/DC-ObjectTypes.html>.

#### FORMATO.

La representación de los datos del recurso, como texto/html, ASCII, fichero Postscript, aplicación ejecutable o imagen JPG. La definición de este elemento tiene por finalidad el suministrar las informaciones necesarias para permitir a las personas o a las máquinas el tomar decisiones sobre la utilización de los datos codificados (por ejemplo, qué equipos y qué programas son necesarios para visualizarlos o ejecutarlos).

Al igual que ocurre con el tipo de recurso, el formato será asignado a partir de listas como los tipos de soportes registrados en Internet (tipos MIME). En principio, los formatos pueden incluir soportes como los libros, publicaciones seriadas y otros soportes no electrónicos.

#### IDENTIFICADOR.

Cadena o número utilizado únicamente para identificar el recurso. Las URL y URN (si se da el caso) son algunos ejemplos de identificador para los recursos interconectados. Otros identificadores únicos a nivel mundial, como los números estándar internacionales de los libros (ISBN) u otros números oficiales, pueden también ser utilizados aquí.

#### FUENTE.

Si fuera posible, la obra impresa o electrónica de donde proviene el recurso. Por ejemplo, la codificación html de un soneto de Shakespeare podría identificar la versión papel del soneto que sirvió de base para su transcripción en versión electrónica.

#### LENGUAJE.

Lengua(s) del contenido intelectual del recurso. En el caso de que sea posible, el contenido de este campo debería coincidir con los tres códigos de caracteres Z39.53 para los lenguajes escritos. Ver: <http://www.sil.org/sgml/nisoLang3-1994.html>

#### RELACIÓN.

Relación con otros recursos. La definición de este elemento tiene por finalidad el suministrar un medio de expresar las relaciones entre los recursos que tienen relaciones formales con otros, pero que ellos mismos son recursos discretos. Por ejemplo, imágenes en un documento, capítulos en un libro o partes de una colección. Actualmente se está trabajando en la elaboración de una definición oficial de la "relación".

#### ALCANCE.

Se trata de las características espaciales y/o temporales de este recurso. En la actualidad se está trabajando en una definición oficial del alcance. Usuarios y desarrolladores deben entender que la utilización de este elemento se encuentra aún en una fase experimental.

#### DERECHOS.

Un enlace hacia una nota de derechos de autor, una declaración de gestión de los derechos o un servicio que informaría sobre las condiciones de acceso al recurso. La especificación formal de los derechos está actualmente en fase de elaboración. Usuarios y desarrolladores deben entender que la utilización de este elemento se encuentra aún en una fase experimental.

### 8.6. Modelo ideal: ¿Cómo elegir los estándares correctos?

Este anexo presenta el estándar propuesto para cada tipo de datos. Pueden utilizarse también otras posibilidades utilizando las informaciones dadas en esta guía. Se considera conveniente el definir un estándar recomendado. Otra posibilidad es el establecimiento de una lista de estándares aceptables para cada tipo de fichero. Esta propuesta no es la única manera de proceder, y no supone sino la plasmación de una serie de ejemplos que podrán ayudar al lector a definir su propia estrategia.

Tipos de datos	Estándar recomendado	Comentarios
Juegos de caracteres	ISO/IEC 8859-1	Otra posibilidad para los países de Europa Occidental es la de Unicode (ISO/IEC 10646) si son necesarios otros juegos de caracteres.
Texto estructurado	SGML	
Imágenes de mapa de bits	JPEG	
Faxes	ITU-T Grupo III	
Gráficos vectoriales	CGM	
Audio y vídeo	MPEG II	
CAD/CAM	STEP	
Contabilidad/Facturación	Edifact	
Otros ficheros de bases de datos	Fichero plano, coma de separación	No existe ningún formato estándar para las bases de datos. El fichero plano permite la conservación a largo plazo si la estructura de la base de datos se encuentra bien documentada.
Ficheros encriptados	RSA	
Programas	fuente o versión compatible con los PC	No existe ningún estándar para los programas compilados. El código binario Java, que no depende de la plataforma, puede contribuir a la conservación a largo plazo de las aplicaciones Java.
Soportes para la conservación a largo plazo	DVD	El DVD es relativamente nuevo, pero debería ser profusamente utilizado en el futuro. Su gran capacidad y su fácil utilización pueden ayudar a hacer del DVD el soporte de los archivos. Varias organizaciones han escogido este soporte. Se recomienda elegir un soporte o un pequeño número de soportes para facilitar su futura utilización.

## 8.7. Índice de términos.

---

### A

Acceso a los datos, 47  
Algoritmos de encriptación, 46  
Anonimato, 46  
ASCII, 39  
Autenticación  
servidores de autenticación, 47

---

### B

Base de datos, 13; 42

---

### C

Caracteres, juegos de, 39  
ASCII, 39  
UCS, juego de caracteres  
universal, 39  
Cartucho magnético, 35  
Casete DAT, 35  
Casete de video de 8 mm, 35  
CCITT, 51  
CD-ROM, 35  
CEI, 51  
CEN, 51  
Cenelec, 51  
CIA, 7  
Cinta magnética, 35  
Compresión de imágenes, 38  
Consejo Internacional de Archivos,  
7  
Contabilidad, 42  
Contabilidad/Gestión (ficheros),  
42  
Contenido, 13  
Contexto, 13  
Correo electrónico, 31

---

### D

Datos, 11; 41  
Datos legibles por máquina, 16; 49  
Derechos de acceso, 45  
DES, encriptación, 46  
Destrucción, 21  
Difusión, 47  
Difusión activa, 47  
Difusión pasiva, 47  
DIS (proyecto de norma  
internacional), 51  
Disco óptico regrabable, 35; 36  
Disco WORM, 35; 36  
Disquete, 35  
DLM, 49  
Documento, 14

documento orientado a objeto,  
37

Documento electrónico  
contenido, 13  
contexto, 13  
estructura lógica, 13  
DSSSL, 40  
DTD, 40

---

### E

EDI, 41  
Edifact, 41; 42  
EN (norma europea), 51  
Encriptación, 46  
DES, 46  
PGP, 47  
RSA, 46  
ENV (pre-norma europea), 51  
Escáner, 24  
Especificaciones disponibles  
públicamente, 50  
Estándares de facto, 50  
Estructura lógica, 13  
ETSI, 51  
EXPRESS, 42

---

### F

Facturas, 42  
Fase de creación, 10  
Fase de diseño, 10  
Fase de mantenimiento, 10  
Fax, 25; 37  
formatos de ficheros de fax, 38  
FIMS, 42  
FIPS, 40  
Firma electrónica, 42  
Formatos de Vídeo, 38  
Formulario, 42  
Fotografía, 37  
FTP, 45  
Fuentes tipográficas, 40

---

### G

GIF, 38  
Gráficos vectoriales, 37

---

### H

Hoja de cálculo, 42  
HTML, 40  
HTTP, 45  
HyTime, 40

---

### I

IDEF, 42  
IDEF0, 42  
IDEF3, 42  
IETF, 50  
Imágenes de mapa de bits, 37  
Internet, 47  
ISAM, 45  
ISO, 51  
ISP, 51  
ITU, 51  
ITU-T, 51

---

### J

JPEG, 38

---

### L

lenguaje de marcación, 40

---

### M

Mensaje  
mensaje autenticado, 47  
mensaje confidencial, 47  
mensajes electrónicos, 31; 41  
clasificación, 31  
Microfilms, 34

---

### N

NIST, 51  
Normas, 51  
normas europeas, 51

---

### O

OCR, 24  
ODA, 40  
OLE, 38  
OpenDoc, 38

---

### P

PAS, especificaciones disponibles  
públicamente, 50  
PDF, 41  
Perfiles, 51  
PGP, 47  
Photo-CD, 38

PostScript, 41  
Prenormas Europeas, 51  
Presentación, 13  
Programas, 41  
Proyecto de norma internacional,  
51

---

## **R**

Reconocimiento óptico de  
caracteres., 24  
*RFC*, 51  
RSA, encriptación, 46  
RTF, 40

---

## **S**

SDIF, 40  
SGML, 40  
Soportes, 34

Soportes magnéticos, 34  
Soportes ópticos, 34  
SPDL, 40  
SQL, 41; 45  
STEP, 42

---

## **T**

Tecnología magneto-óptica, 35; 36  
Texto, 39  
Texto plano, 26  
TIFF, 38

---

## **U**

UCS, juego de caracteres universal,  
39  
UIT, 51

---

## **V**

Vectorización, 25  
Vida privada, 44  
Vídeo, 38

---

## **W**

Workflow, 14; 30  
WORM, 35; 36

---

## **X**

X/Open, 51  
XPG, 51

## **8.8. Índice de ilustraciones.**

Ilustración 1: ¿Documento o base de datos? (14)

Ilustración 2: Ciclo de vida de los datos legibles por máquina (17)

Ilustración 3: Conversión de un documento en soporte papel a formato electrónico (22)

Ilustración 4: Cambio de formato electrónico (23)

Ilustración 5: ¿Qué hacer con los formatos propietarios? (29)

## **8.9. Bibliografía.**

La presente bibliografía se limita a proponer algunas obras que han servido de base para la redacción de la presente guía. No han sido citados numerosos documentos internos de organismos nacionales y comunitarios.

Les archives dans l'Union européenne.  
Rapport du groupe d'experts sur les problèmes  
de coordination en matière d'archives.  
Commission européenne, secrétariat général.  
Luxembourg EUR-OP, Luxembourg, 1994,  
104 p., ISBN 92-826-8234-X –  
Catalogue n° CM-83-94-741-FR-1

Actes du DLM-Forum sur les données  
lisibles par machine. Bruxelles, 18 au 20  
décembre 1996, dans: INSAR – Courrier  
européen des archives, supplément II, 1997,  
EUR-OP, Luxembourg, 376 p.,  
ISBN 92-828-0112-8 –  
Catalogue n° CM-AC-97-SO1-FR-C  
(FR, DE, EN)  
sur Internet:  
<http://www.echo.lu/dlm/en/proc-index.html>

EPHOS – European Procurement  
Handbook for Open Systems.  
Référence: EUR 14021

Guide pour la gestion des documents  
électroniques du point de vue archivistique.  
Conseil international des archives, comité sur  
les documents électroniques, Paris, 1997.  
Sur Internet:  
<http://www.archives.ca/ica/p-er/francais.html>

B. Bauwens, F. Evenpoel et J. Engelen,  
«Standardisation as a Prerequisite for  
Accessibility of Electronic Text Information for  
Persons who cannot use Printed Material»,  
IEEE Transactions on Rehabilitation Engineering,  
vol 3, n° 1, p. 84-89, 1995.

Chr. Reeves et T. Wesley, «Guidelines for  
Accessible Web page design», Brochure  
publiée par le Harmony Consortium (1997).  
Elle se trouve également sur Internet:  
[http://www.esat.kuleuven.ac.be/teo/harmony/  
guidelines](http://www.esat.kuleuven.ac.be/teo/harmony/guidelines)