

## Conservación de documentos electrónicos a largo plazo

---

Actualmente la mayor parte de la información que se maneja en una administración está recogida en documentos electrónicos. El trabajo diario de cualquier funcionario se realiza también con documentos electrónicos. Muchos de los trámites que se realizan o de los procedimientos en los que se participa son realizados de manera electrónica. Podríamos seguir enumerando aspectos en los que la documentación en formato electrónico forma parte de las tareas cotidianas de cualquier funcionario público.

Y sin embargo, muchas de las leyes y normas que rigen estos procedimientos fueron creadas para un entorno en el que la documentación estaba basada en papel. Algunos ejemplos de estas leyes y normativas son la Ley de Procedimiento Administrativo o la Ley de Protección del Patrimonio Histórico Español

Esto hace que haya muchos aspectos de la legislación que deban ser revisados a la luz de la realidad de la documentación electrónica. Aspectos como el “archivo a perpetuidad” de ciertos documentos como las actas municipales, adquieren una complejidad enorme cuando éstas se crean en formato electrónico.

### 1. *Visión histórica*

Desde que existe la escritura, ésta ha sido utilizada para dejar constancia de aquellos hechos que debían tener efectos a largo plazo. Esta escritura se realizó primero sobre soportes resistentes al tiempo, como las tablillas de barro mesopotámicas, o los jeroglíficos egipcios sobre paredes de piedra.

Con el paso del tiempo, fueron haciéndose habituales los documentos que reflejaban actos administrativos, y que debían conservarse mientras perdurasen los efectos de esos actos administrativos. Al mismo tiempo, los soportes sobre los que se plasmaban estos documentos iban evolucionando hacia materiales más flexibles y manejables, como el papiro primero y el pergamino después.

Desde el principio existía la necesidad de conservar los documentos creados, lo que condujo a la aparición de toda una serie de técnicas que permitían conservar los documentos a lo largo de largos periodos de tiempo.

En el caso de los escritos sobre materiales perdurables, el propio soporte de la escritura hacía que ésta se conservase durante un periodo de tiempo muy amplio. A esto se unió el empleo de ciertas técnicas que ayudasen a esta conservación, como controlar la humedad del lugar en que se guardaban las tablillas, logrando la conservación de estos documentos durante siglos. Tanto es así, que hoy en día podemos leer perfectamente tablillas de barro sumerias del siglo 30 a.C.

La utilización de soportes menos resistentes para los documentos más habituales, hizo que fuera preciso mantener estos documentos a salvo de factores externos como

plagas de insectos, ya que dado su origen vegetal o animal, eran muy vulnerables a la acción de éstos. Esta protección solía consistir en guardar el documento en algún recipiente que lo mantuviese a salvo de estos organismos.

Durante la Edad Media aparecieron nuevos tipos de tinta que eran más perdurables sobre el soporte que las empleadas en la Antigüedad. Sin embargo, con el tiempo estas nuevas tintas empleadas han causado grandes daños en los propios códices debido a su composición ácida. En muchos casos el deterioro del soporte hacía necesario copiar el documento original en un nuevo soporte, de manera que se salvase la información contenida en él

Ya en la Edad Moderna, aparecen los grandes archivos, como lugar en el que almacenar un gran número de documentos relativos a una temática particular, como el Archivo General de Indias o el Archivo de Simancas, lo que permitía custodiar adecuadamente los documentos.

Como se ve, la conservación de los documentos ha sido una preocupación constante desde el origen de los mismos.

## **2. Situación actual: El documento electrónico**

En nuestros días asistimos a un cambio tecnológico que ha hecho que se produzca un cambio muy importante en cuanto a los documentos que empleamos. Los documentos han dejado de estar mayoritariamente soportados en papel, para convertirse en documentos electrónicos. Éstos necesitan 2 factores para ser accesibles: el formato de los mismos y el soporte.

Haciendo una analogía con el documento tradicional en papel, pergamino o papiro, para poder leer un documento necesitamos tener el soporte y entender el lenguaje en que el documento esta escrito. Este formato (lenguaje) del documento, ha dependido hasta ahora de la aplicación informática que se utilizara para crearlo.

Al tratarse de un mundo en el que las aplicaciones empleadas para crear documentos (como procesadores de texto, hojas de cálculo, presentaciones, etc), los sistemas operativos o las mismas compañías que crean las aplicaciones cambian con frecuencia, nos encontramos con muchos documentos creados con herramientas muy comunes en una época, pero que no pueden ser utilizados en la actualidad. Un ejemplo paradigmático de esto pueden ser lo procesadores de texto que se utilizaban mayoritariamente hace 10 años hoy tienen una presencia minoritaria en el mercado.

Esto hace que sea más difícil acceder a un documento electrónico creado en 1985 que a una tablilla sumeria del siglo XII a.C.

### **2.1. ISO 19005-1: PDF/A**

En esta situación, se detecto la necesidad de contar con un formato que garantizase la conservación de los documentos electrónicos a largo plazo. Anteriormente ISO ya había

estandarizado distintos formatos que partían de una especificación de PDF, como es el caso del PDF/X.

El formato de documento para conservación a largo plazo fue aprobado por el comité técnico 46 (TC46) de ISO con el nombre ISO 19005-1 en Septiembre de 2005. Este formato también es conocido como PDF/A, al estar basado en un subconjunto del formato PDF.

La versión actual de estándar PDF/A está construida a partir de la especificación 1.4 del formato PDF de la empresa Adobe Systems Inc. Actualmente ISO está trabajando en una nueva versión del estándar que lo amplíe a la versión 1.7 del formato PDF.

La necesidad de crear un estándar para la preservación de los documentos electrónicos a largo plazo viene dada por la existencia de un gran número de formatos que podrían ser empleados para este fin. Es necesario por tanto definir un estándar que garantice la interoperabilidad de los sistemas de almacenamiento.

Hay algunas consideraciones que pueden ser útiles a la hora de entender la necesidad del estándar.

Por un lado, existen hoy en día múltiples formatos de documento que podrían utilizarse para conservar los documentos a largo plazo. Entre estos formatos podríamos considerar el TIFF, los formatos nativos de las aplicaciones de creación de documentos, el formato PDF tal y como lo publica Adobe, el XML y algunos otros.

Un formato de archivo de documentos electrónicos a largo plazo debe cumplir con algunos requisitos básicos:

- Debe ser un formato independiente del dispositivo, de manera que el documento pueda ser accedido por cualquier usuario ahora o en el futuro. No sería aceptable que el formato dependiese del sistema operativo, la resolución del dispositivo de visualización, o cualquier otro parámetro cuya evolución futura desconocemos.
- Debe ser un formato autocontenido, de manera que toda la información necesaria para representar el contenido del documento se encuentre en el propio documento, y no sea preciso acceder a contenidos externos que quizá en el futuro no estén disponibles. Un ejemplo común de este punto lo tenemos con las fuentes de texto. Si para representar el documento empleamos las fuentes del sistema, no podemos garantizar que en el futuro el documento se represente de manera fidedigna al no poder asegurar que esas fuentes estén disponibles.
- Autodocumentado, es decir, que toda la información necesaria para localizar el documento, catalogarlo, etc esté contenida en el propio documento, y no almacenada en recursos externos.
- El formato debe impedir las restricciones, de manera que el contenido esté accesible sin protecciones de ningún tipo. La existencia de protecciones, como contraseñas, compromete el acceso futuro a la información.

- El formato debe estar publicado.

Además de estos criterios, partir de un formato que esté ampliamente aceptado por la comunidad garantiza una rápida adopción del mismo.

## 2.2. PDF versus PDF/A

En principio, puede parecer que PDF es un candidato idóneo para convertirse en el estándar para almacenamiento a largo plazo de documentos electrónicos. Sin embargo, algunas de las capacidades de PDF como formato podrían impedir que los documentos generados fuesen accesibles en el futuro.

A modo de ejemplo enumeramos algunas de estas características:

- Los documentos en formato PDF pueden contener contraseñas, que impedirían la apertura del documento por todo aquel que no la conozca. Con el transcurso del tiempo, esta contraseña puede perderse y el documento quedar inaccesible.
- Los documentos PDF pueden contener anexos, y éstos pueden ser cualquier tipo de archivo. Esto quiere decir que si los anexos no están en un formato perdurable, la información contenida en el documento PDF estará incompleta al no poder accederse a la que contienen los anexos.
- Los documentos PDF no necesariamente son autocontenidos. Puede haber dependencias de las fuentes del sistema o de contenido accesible a través de enlaces en el documento.
- Existen múltiples herramientas de creación de documentos PDF en el mercado, y el nivel de cumplimiento de la especificación por parte de todas ellas no siempre es óptimo. Esto puede provocar en el futuro problemas para el acceso a documentos PDF que no se ajusten escrupulosamente a la especificación

Por tanto, es necesario contar con una especificación del formato PDF estándar y que asegure el acceso al documento en el futuro. Debe garantizarse que los documentos serán visualizados de manera consistente y predecible, por mucho tiempo que pase o muchos cambios tecnológicos que haya.

PDF/A es un estándar internacional publicado por ISO que especifica el formato adecuado para la preservación de documentos electrónicos a largo plazo. Ha sido desarrollado con la participación de multitud de gobiernos, organismos y compañías de todo el mundo, por lo que cuenta el respaldo de la mayor parte de la comunidad.

Hay aspectos relacionados con la conservación de documentos electrónicos que el estándar PDF/A no cubre. Entre estos aspectos podemos destacar, entre otros, el soporte físico en que se conservará el documento, el sistema operativo, las condiciones de almacenamiento, el equipo informático necesario, etc.

### **3. Descripción del formato PDF/A**

A continuación presentamos un breve resumen de algunas características importantes del formato PDF/A. A grandes rasgos podemos clasificar en 3 grandes grupos las diferencias del formato PDF/A respecto de la especificación pública del formato PDF publicada por Adobe: Características obligatorias, características prohibidas y recomendaciones.

Serían obligatorias todas aquellas características que son imprescindibles para garantizar el acceso al documento en el futuro. Estas características aseguran que el documento sea representado de manera fidedigna en cualquier dispositivo. Entre estas características están la incrustación de fuentes, la identificación de las anotaciones o la preservación del aspecto visual del documento.

En el formato están prohibidas todas aquellas características que puedan comprometer el acceso futuro al documento. Entre estas características están el cifrado y protección del documento, los objetos dinámicos incrustados (cuya reproducción está íntimamente ligada al sistema operativo), las fuentes propietarias o aquellas que no se pueden incrustar, o los enlaces externos (que podrían no estar disponibles en el futuro).

Finalmente, hay algunas características que figuran en el estándar como recomendaciones, como es el caso de la utilización de metadatos para identificar la información contenida en el documento. En caso de emplearse metadatos, éstos deben estructurarse en formato XMP.

En función de las características que cumpla el documento, se establecen 2 niveles de cumplimiento con el estándar: 1a y 1b.

El nivel 1b, el más bajo, garantiza que el documento es visualmente idéntico al que creó el autor. En este nivel puede no haber etiquetado del contenido.

En el nivel 1a, el mayor, se incluyen todas las características del nivel 1b y además se obliga a que el documento esté etiquetado, lo que permite añadirle mucha información sobre su estructura. Esto permite, por ejemplo, que un lector pueda conocer en que orden debe leerse o visualizarse el documento. Por ejemplo, un lector automático puede leer las columnas como tales en lugar de horizontalmente.

### **4. Consideraciones para crear documentos PDF/A**

A la vista de todo lo expuesto anteriormente, hay una serie de aspectos que deben tenerse presente a la hora de generar documentos que vayan a convertirse en PDF/A. Es importante tener presente que hay ciertas características que podrán forzarse a la hora de

hacer la conversión a PDF/A, pero hay otras que deben ser tenidas en cuenta en el momento de crear el documento.

Algunos de estos aspectos que deben ser tenidos en cuenta son:

- No deben utilizarse fuentes que no puedan ser incrustadas en el documento, por ejemplo, aquellas de las que no tengamos derechos de uso.
- Los gráficos que se empleen en el documento deben estar en un formato independiente del dispositivo.
- No debe emplearse la transparencia.
- No debe haber enlaces a contenido externo que complemente la información contenida en el fichero. Estos enlaces externos se convertirán en texto cuando el documento se convierta en PDF/A.
- Los formularios que se empleen, son convertidos en texto cuando el documento se convierte en PDF/A. Esto quiere decir que se sustituyen los campos dinámicos por su contenido.
- No deben emplearse acciones que modifiquen el contenido del documento, ya que en PDF/A todo el contenido es estático. Un ejemplo de este tipo de acciones podría ser un campo que contuviese la fecha del sistema.
- La firma digital en la actual revisión del formato PDF/A no está soportada como tal. Si se convierte un documento PDF firmado digitalmente en PDF/A, la información de la firma se convertirá en información estática que se representará en el PDF/A, pero no será un campo de firma digital. En caso de que sea necesario conservar la información sobre la firma digital de manera que sea posible consultarla en el futuro habrá que almacenar esa información en una estructura que sí se mantenga dentro del PDF/A. El mecanismo que parece más adecuado es mantener toda esa información en una estructura de datos XMP que se conserve con el documento PDF/A.

## **5. Evolución futura del estándar**

Como se ha comentado anteriormente, la versión actual del estándar PDF/A está basada en la especificación 1.4 del formato PDF que publica Adobe, mientras que la versión más reciente del formato PDF es la 1.7 (publicada al mismo tiempo que apareció Acrobat 8). Actualmente se está trabajando en la extensión 2 del estándar ISO 19005 que estará basado en la especificación 1.6 y añade algunas características importantes como son:

- Soporte para compresión de imagen en formato JPEG2000
- Mejor soporte para firma digital
- Soporte para fuentes OpenType

- Gráficos 3D
- Contenido de audio y de video
- Consistencia con otros formatos estándar como PDF/X, PDF/E o PDF/UA

Esta nueva versión del estándar será la 19005-2, es decir, será una extensión de la 19005-1 actual. Posteriormente se avanzará con la creación del formato 19005-3.

En Enero de 2007, en paralelo con el trabajo de ISO en el formato 19005-2, Adobe decidió ceder a AIIM el formato PDF para su conversión en un estándar ISO. En estos momentos ISO está trabajando sobre la especificación 1.7 de PDF y considerando convertir PDF en un estándar ISO.

Existen asimismo otros estándares en desarrollo por parte de ISO que también utilizan la especificación PDF como punto de partida, como es el caso de PDF/E o PDF/UA.