

17 de noviembre de 2004

## LISTA DE FORMATOS ACEPTABLES PARA EL INTERCAMBIO DE DOCUMENTOS EN LA UNIVERSIDAD DE CÁDIZ, Y CÓMO GENERARLOS A PARTIR DE LAS HERRAMIENTAS MÁS HABITUALES

### 1. OBJETIVOS DEL PRESENTE DOCUMENTO

El lunes 11 de octubre de 2004 se publicó en el Boletín Oficial de la Universidad de Cádiz n.º 15 la **Normativa de Intercambio de Información Institucional en la UCA**, por la que los documentos oficiales intercambiados o publicados digitalmente deben estar disponibles en formatos abiertos, siendo tarea del CITI el mantener una lista actualizada de formatos abiertos para diferentes tipos de documentos, a los cuales deberán atenerse las publicaciones institucionales, así como difundir las instrucciones sobre cómo generarlos. Es pues el objetivo de estas líneas el informar de cuáles son los formatos de documentos aceptables y de cómo generarlos a partir de los programas de uso más habitual hoy día en nuestra universidad.

Este documento no es inamovible, y podrá variar en el tiempo según avance la tecnología o los usos y costumbres informáticas en la Universidad de Cádiz. Se ha pretendido ser flexible en favor de los usos actuales y comodidad de los usuarios, para lo cual la normativa en cuestión se ha relajado un tanto en aquellos puntos en los que se ha creído necesario.

### 2. TERMINOLOGÍA

Para el presente documento se empleará la siguiente catalogación:

- **Formato cerrado o propietario:** Sus derechos pertenecen a una o varias empresas. Su especificación no está publicada, o lo está parcialmente. Habitualmente se requieren técnicas de ingeniería inversa para que otras herramientas lo utilicen. Ejemplos: DOC (MS-Word), XLS (MS-Excel).
- **Formato publicado:** Sus derechos pertenecen a una o varias empresas, pero su especificación está publicada totalmente y por tanto existe la posibilidad de su uso por parte de una aplicación no perteneciente a la empresa propietaria. Este empleo no está restringido. Sin embargo, el futuro del formato depende, al igual que en el caso de los formatos cerrados, de los avatares y decisiones de la empresa propietaria. Ejemplos: PDF y Postscript de Adobe, RTF de Microsoft.
- **Formato abierto:** Pertenece a alguna organización sin ánimo de lucro, con el propósito de regular el formato. Éste está completamente publicado y suele existir una implementación de referencia que ayuda a desarrollar software totalmente compatible. Ejemplos: HTML de W3C, SX\* de OpenOffice.org.

### 3. REQUISITOS

1. **Formato abierto:** Es lo que da más garantía de perdurabilidad a los datos, ya que está controlado por una organización sin ánimo de lucro que se dedica a su normalización.

2. **Existencia de lectores y generadores** que deben ser:
- libres.** La elección de los formatos debe enmarcarse dentro de la tendencia hacia el software libre, pues así no se fuerza a los destinatarios a adquirir productos comerciales, ni se fomenta la copia ilegal de software.
  - de buena calidad.** Los usuarios no deben verse forzados a utilizar herramientas de baja calidad, pues esto, además de los perjuicios obvios, podría generar el incumplimiento de la normativa.
  - muy difundidos.** Es deseable evitar la necesidad de capacitación y adaptación a nuevas herramientas, así como generar a los destinatarios más trabajo del mínimo necesario.
  - disponibilidad en las plataformas más usadas en puestos de trabajo.** No debe limitarse la elección de la plataforma del puesto de trabajo del usuario, al menos en el caso del personal docente. Por lo menos deben existir implementaciones adecuadas de las herramientas necesarias para las más utilizadas: MS Windows, GNU/Linux y Macintosh.

## 4. FORMATOS MÁS IMPORTANTES PARA DOCUMENTOS DE TEXTO

Los documentos de texto son los más utilizados para el intercambio institucional, con gran diferencia; la versión actual del presente informe se centra en documentos de este tipo, aunque posteriormente será ampliado a otro tipo de documentos.

### 4.1. Formatos que cumplen con todos los requisitos

- **Texto simple.** También conocido (incorrectamente) como «texto plano» o «ASCII». Este formato se compone de una secuencia de caracteres en alguna codificación normalizada por la organización ISO, como ISO-Latin1, ISO-Latin9 o Unicode (UTF-8).
  - **Adecuado para:** mensajes de correo electrónico.
  - **Inconvenientes:**
    - Es un formato muy poco potente, no permite el resaltado del texto.
    - Al no estar Unicode suficientemente extendido aún, la visualización correcta del texto depende de los códigos de caracteres empleados por emisor y receptor; aunque es fácil cambiar entre códigos, el proceso puede ser complicado y tedioso para un usuario normal.
- **HTML.** Evidentemente, se excluyen de este formato extensiones no especificadas por los estándares publicados por W3C, que es el consorcio encargado de ello.
  - **Adecuado para:** publicación en la web, formularios en páginas web, intercambio de documentos formateados sencillos.
  - **Inconvenientes:**
    - El mismo documento puede verse distinto en distintos equipos o visores (navegadores).
    - Algunas herramientas de generación de HTML muy extendidas generan código HTML propietario no compatible con la especificación estándar.
    - Muchas veces un documento se compone de varios ficheros, por ejemplo si tiene gráficos, por lo que resulta complicado para el usuario común el intercambio o publicación.

### 4.2 Formatos que cumplen con todos los requisitos salvo el 2.c (estar muy difundidos)

Tratamos aquí el caso de los formatos que cumplen todas las condiciones especificadas en el apartado 3, salvo la existencia de programas lectores y generadores que estén muy difundidos.

- **SXW.** Se trata del formato nativo de OpenOffice.org para documentos de texto. Basado en XML, ha sido escogido por QASIS como base para la creación de un estándar para formatos utilizados en

las *suites ofimáticas*. A pesar de su poca difusión, ésta no es tan baja como para no poder considerarse como alternativa (se estiman unos 40 millones de usuarios en el mundo, y se cree que este número aumentará), y con una política de formación adecuada que capacite a los miembros de la UCA para su empleo, es el formato más adecuado para documentos compartidos susceptibles de modificación.

- **DocBook, T<sub>E</sub>X/L<sup>A</sup>T<sub>E</sub>X, etc.** Lenguajes para la composición de texto y la generación de documentos en formatos adecuados para ser impresos o visualizados, como Postscript, PDF o HTML. Aunque muy extendidos en el ámbito académico, científico y de la publicación profesional, sus características los hacen poco adecuados para el uso general en la oficina.

### 4.3 Formatos publicados de interés

- **RTF.** Siglas de *Rich Text Format*, o *formato de texto enriquecido*, es un formato publicado propietario de Microsoft, que ha evolucionado desde la versión inicial 1.0 hasta la 1.7. Es un formato de texto, en el sentido de que el documento se guarda en texto plano, compuesto de caracteres y órdenes de control, y el texto del documento en texto sencillo.

- **Inconvenientes:**

- Aunque Microsoft proporciona documentación técnica sobre RTF, no existe una normativa que especifique este formato.
- RTF no puede guardar macros, lo que también puede verse como una ventaja en cuanto a seguridad.
- Las imágenes se guardan sin comprimir, lo que hace que los documentos con gráficos empujados ocupen mucha memoria de almacenamiento.
- Hay problemas de compatibilidad entre versiones: un documento RTF generado con una versión de MS-Word puede no poder abrirse desde una versión anterior.
- A veces hay problemas con material tabular (tabuladores, filas y columnas).
- Por último, parece que Microsoft va a abandonar este formato en favor de los esquemas de referencia de XML en las futuras versiones de MS-Office.

- **Esquemas de referencia XML de Microsoft.** Con la versión 2003 de MS-Office, Microsoft anunció un nuevo formato basado en XML, junto con la documentación técnica de referencia y la disponibilidad de licencias gratuitas para su uso libre. Esos esquemas describen cómo se almacena la información cuando los documentos se guardan como XML. Sin entrar en detalles técnicos, básicamente en teoría cualquier fabricante de programas, ya comerciales o de software libre, podría crear procesadores de texto que entendieran estos formatos.

A pesar de ello, no se va a recomendar este formato para el caso que nos ocupa debido a los siguientes

- **Inconvenientes:**

- No todas las aplicaciones de MS-Office pueden hacer uso del formato XML: PowerPoint sólo guarda las presentaciones en formatos binarios propietarios, no XML.
- Algunas características avanzadas de MS-Office 2003 sólo están disponibles cuando los documentos se guardan en su formato binario propietario.
- Hoy por hoy, este formato sólo se encuentra en la última versión de MS-Office, la 2003, y sólo para la plataforma Windows; se incumplen pues los requisitos 2 *a*, *c* y *d*.
- Microsoft no ha asegurado que vaya a publicar los esquemas de referencia XML para futuras versiones de MS-Office.
- XML no es el formato «nativo» o predeterminado de MS-Office 2003: el usuario tiene

que preocuparse de guardar el documento en este formato diciéndolo expresamente.

- Los documentos guardados en este formato XML pueden aún contener datos binarios codificados que sólo se pueden manejar desde Windows y Office 2003.
- Los esquemas XML no garantizan necesariamente la fidelidad de los documentos entre distintas aplicaciones de oficina.
- Algunos aspectos legales de la licencia no están claros.

Todos los formatos enumerados hasta ahora son de tipo editable, es decir, la persona que recibe la información puede editarla y modificarla con facilidad. Sin embargo, en la mayoría de las ocasiones la información difundida esta orientada a ser visualizada y no necesita ser modificada por la persona que la recibe.

- **PDF.** *Portable Document Format*, o *formato transportable de documentos*. Fue creado en 1996 por la compañía Adobe, derivado de otro formato suyo, Postscript (PS). Tiene la propiedad de que el aspecto del documento es el mismo en pantalla que cuando se imprime en papel, y es el mismo en cualquier plataforma. Existen lectores libres y gratuitos de muy buena calidad muy difundidos. Sobre todo el Adobe Acrobat Reader, gratuito (pero no libre) y disponible para prácticamente todas las plataformas.

Recientemente Adobe ha anunciado el desarrollo de otro formato: **XDP** (*XML Data Package*). No es éste lugar ni tiempo para describirlo, es muy prometedor pero sólo el futuro dirá si tiene éxito. XDP añade a PDF la posibilidad de incluir componentes XML reescribibles. De todas formas parece estar más orientado a formularios electrónicos que a documentos de propósito general.


#### 4.4 Formatos recomendados

Teniendo en cuenta estas premisas, **se recomienda el uso de los siguientes formatos:**

- Para documentos que sólo tengan que visualizarse, y no modificarse: **PDF**.
- Para documentos que deban ser modificables electrónicamente por el usuario, como por ejemplo impresos de peticiones o trabajo en grupo: el **formato nativo de OpenOffice.org (SXW)**, que también es el de StarOffice, el paquete de ofimática de Sun Microsystems.

### 5. CÓMO PRODUCIR DOCUMENTOS EN LOS FORMATOS RECOMENDADOS

#### 5.1 CÓMO PRODUCIR DOCUMENTOS EN PDF DESDE MS WINDOWS

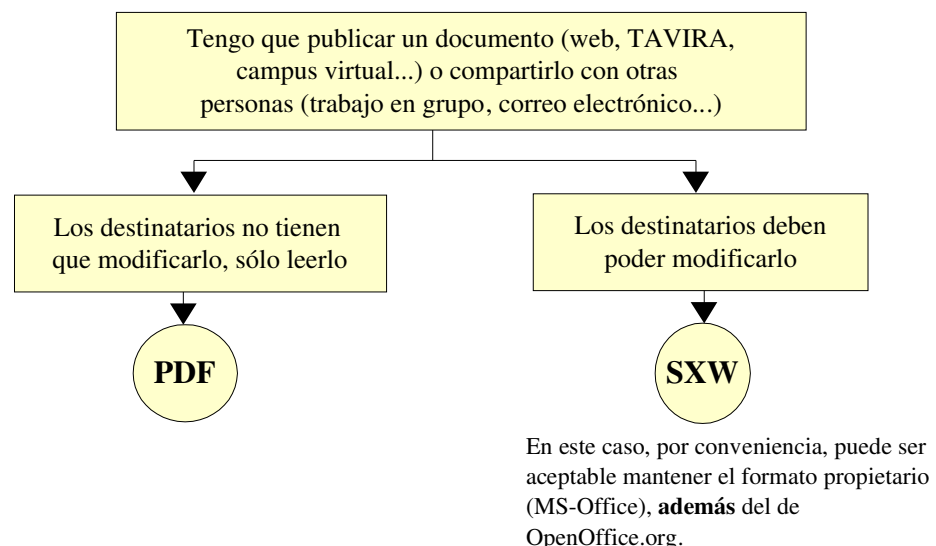
1. **Usando OpenOffice.org.** A partir de la versión 1.1 de este programa, descargable desde su página web [es.openoffice.org](http://es.openoffice.org), no hay más que pulsar un botón con el símbolo de Adobe (dependiendo de la versión o configuración, una especie de **A** roja sobre un cuadrado blanco:  o un documento con la base en rojo y las letras **PDF** en blanco) o escoger del menú Archivo el elemento Exportar a formato PDF...
2. **Usando MS-Office.** Este programa no admite la conversión o exportación a PDF, por lo que hay que emplear otra herramienta. Por ser libre, se recomienda **PDFCreator**, accesible desde <http://sourceforge.net/projects/pdfcreator>. Este programa instala en Windows una cola de «impresora» virtual tal que lo que se manda a imprimir a ella, se convierte a un fichero en formato PDF, con lo cual se pueden generar tales documentos no sólo desde MS-Word sino desde cualquier aplicación de Windows desde la que se pueda imprimir.

## 5.2 CÓMO PRODUCIR DOCUMENTOS EN FORMATO DE OpenOffice.org

1. Evidentemente empleando desde el principio esta herramienta, escribiendo el documento con ella (o con StarOffice), y guardándolo en su formato nativo. OpenOffice.org es una *suite ofimática* libre que está compuesta por un procesador de textos, una hoja de cálculo, un programa de presentaciones, otro de dibujo y diagramas, un editor de ecuaciones y otro de páginas web. Está en español y otros idiomas, tiene diccionarios y ayudas, admite varias fuentes de datos, hay versiones para MS-Windows, GNU/Linux, Sun Solaris y recientemente para Mac-OS/X, exporta a PDF y a Macromedia Flash, y en definitiva, salvo para casos muy especiales, puede sustituir perfectamente a MS-Office.
2. Si el documento ya ha sido escrito en un formato de MS-Office, lo abriremos con OpenOffice.org o StarOffice, que lo importará, y a continuación podemos guardarlo en el formato nativo de OpenOffice.org, o en otro de los que acepte. OpenOffice.org tiene filtros de importación capaces de aceptar los formatos de MS-Office desde la versión 97 a la 2003 (XML).

(StarOffice es la versión comercial de OpenOffice.org, mantenida por Sun Microsystems. Son idénticos salvo que StarOffice trae más tipos de letras y galerías de símbolos, plantillas, etc., y ofrecen soporte y mantenimiento.)

## 6. GUÍA RÁPIDA: CÓMO ESCOGER EL FORMATO MÁS ADECUADO



## 7. CONCLUSIÓN

El CITI, por el presente documento, según mandato de la Normativa de Intercambio de Información Institucional en la UCA, publicada en el BOJA n.º 15 el 11/X/04:

- **Recomienda** los siguientes formatos de documentos de texto:
  - **PDF** para documentos de lectura exclusiva.
  - **SXW** (OpenOffice.org/StarOffice) para documentos compartidos o que deban ser modificados.
- **Acepta:**
  - **Texto sencillo** para situaciones donde el formato visual sea muy poco importante, como mensajes de correo electrónico o borradores.
  - **HTML/XHTML** en versiones fijadas por el W3C, actualmente 4.01 y 1.0 respectivamente;

para páginas web sobre todo. Observe que MS-Word o MS-Frontpage no generan HTML estándar.

- **Rechaza:**
  - **DOC** y otros formatos propietarios binarios (WP de WordPerfect, etc.).
  - Esquemas de referencia XML de Microsoft.
  - **RTF**.
  - Otros formatos que, aun siendo libres o publicados, sean de muy poca difusión en general, como Postscript, DjVu, DVI, DocBook, T<sub>E</sub>X/L<sup>A</sup>T<sub>E</sub>X...

## 8. BIBLIOGRAFÍA BÁSICA

- *Comparative assessment of Open Documents Formats— Market Overview*  
*Valoris, ENTR/02/21-IDA/MIDDLEWARE-XML, Specific agreement nº 3 – IDA.20030523, Final Assessment Report*
- *Microsoft Comments to IDA regarding the Valoris Report “Comparative Assessment of Open Document Formats – Market Overview”*
- *Comments on the document ENTR/02/21-IDA/MIDDLEWARE-XML: “Comparative assessment of Open Document Formats Market Overview” as submitted by Valoris.Sun Microsystems.*
- *Política de formatos de documentos*  
*Diego di Pascua, para el Servicio Central de Informática, Universidad de la República (Uruguay)*
- *Aplicaciones utilizadas para el ejercicio de potestades. Criterios de Seguridad, Normalización y Conservación.*  
*Ministerio de Administraciones Públicas*

*Los nombres de compañías y productos mencionados aquí pueden ser marcas registradas de sus respectivos dueños.*