



Política de formatos de documentos

**Servicio Central de Informática
Universidad de la República**

Versión 1.0.0

Versión 1.0.0 – Noviembre 2003 – Diego Di Pascua (SeCIU)

Resumen

Recomendación

Se recomienda la utilización de los formatos listados a continuación.

Para documentos originales:

1. Texto plano ASCII
2. HTML
3. Formatos nativos de OpenOffice.org (sxw, sxc, sxd, sxi, sxm, sxd)

Para publicaciones (debiendo mantener el original en alguno de los formatos anteriores):

1. Texto plano ASCII
2. HTML
3. PDF

Proceso

Se fijaron determinados objetivos referidos a la estrategia de la organización para los formatos utilizados (ver 'Objetivos estratégicos').

A partir de ellos y de otras características deseables, se generaron los requisitos para los formatos a elegir (ver 'Requisitos').

Los requisitos propuestos no son fáciles de alcanzar. En la actualidad, no existen opciones que no tengan alguna desventaja (ver 'Recomendación'). Para paliar este hecho, se debió separar los casos de los formatos para los documentos originales con las versiones publicadas. Si bien puede utilizarse el mismo formato para ambos casos (utilizando texto plano o HTML), de querer utilizarse los formatos nativos del OpenOffice.org para la edición debe publicarse una versión en otro formato, dada la poca difusión de los formatos de dicha herramienta.

Para ello, se recomienda la utilización de PDF, siempre que el texto plano y el HTML no sean suficientes.

Coherentemente con lo anterior, ya que el PDF no es un formato abierto, debe mantenerse el original en alguno de los formatos recomendados para los documentos originales.

Tabla de contenido

Terminología.....	5
Contexto.....	6
Objetivos estratégicos.....	7
Requisitos.....	8
Recomendación.....	9

Organización del documento

Terminología

Especifica la catalogación utilizada para los formatos de documentos.

Contexto

Realiza una breve introducción a la realidad referida a los formatos de documentos.

Objetivos estratégicos

Fija los objetivos que acompañen la estrategia de la institución respecto a sus documentos.

Requisitos

A partir de los objetivos anteriormente indicados y de otras características deseables, genera los requisitos que debe cumplir un formato.

Recomendación

Efectúa la recomendación de formatos, dadas las premisas marcadas anteriormente.

Terminología

Para el presente documento se usará la siguiente catalogación:

- **Formato cerrado:** Sus derechos pertenecen a una o varias empresas. Su especificación no está publicada, o lo está parcialmente. Habitualmente se requiere de técnicas de ingeniería reversa para que otras herramientas lo utilicen plenamente. Ejemplos: doc, xls.
- **Formato publicado:** Sus derechos pertenecen a una o varias empresas. Su especificación está publicada totalmente, y existe la posibilidad de utilizar el formato por parte de una aplicación no perteneciente a la empresa propietaria del mismo. El uso del formato no está restringido. El futuro del formato, al igual que en el caso de los formatos cerrados, depende de los avatares y decisiones de la empresa propietaria. Ejemplos: pdf, ps, rtf.
- **Formato abierto:** Pertenece a alguna organización sin fines de lucro, con el propósito de regular el formato. Fomenta su uso por parte del público en general, sin esperar retribución por ello. Tiene publicado completamente el formato, y habitualmente existe una implementación de referencia que ayuda a desarrollar software totalmente compatible. Ejemplos: html, sxw, sxc.

Contexto

Los formatos de documentos que se han utilizado desde los comienzos de la computación han cambiado reiteradamente. Se sucedieron evoluciones, desapariciones y creaciones de nuevos formatos, muchas veces verdaderos estándares de hecho del momento.

Inicialmente, los documentos se generaban utilizando texto plano. Para lograr mayor riqueza, posteriormente aparecieron diversas alternativas. En algunos casos, se trató de formatos abiertos (por ej., HTML), mientras que en otros se trató de formatos cerrados (por ej., los nativos de WordStar, WordPerfect, 123, Quattro Pro) o publicados (por ej. PDF, RTF).

En la actualidad, la gran difusión del Microsoft Office ha causado que sus formatos nativos se hayan convertido en la opción utilizada por la mayoría de los usuarios en todo el mundo.

Por otra parte, los formatos van sufriendo cambios, lo cual muchas veces genera incompatibilidades entre documentos cifrados con distintas versiones de un mismo formato.

Todos estos cambios muchas veces repercuten negativamente en la perdurabilidad de los documentos, y generan la necesidad de invertir reiteradamente en herramientas que permitan al usuario manejarse con las nuevas especificaciones.

Objetivos estratégicos

La existencia de una política de formatos de documentos debe explicitar los objetivos estratégicos buscados y la forma de alcanzarlos. Dichos objetivos son propuestos en este punto.

Documentos perdurables

Una de las características más deseables para la documentación de una organización de las características de la Universidad de la República es la perdurabilidad.

Como fue expresado anteriormente, los formatos evolucionan y desaparecen con el tiempo. En el caso de los formatos que no son abiertos, muchas veces esta evolución depende de factores ajenos al interés público.

Una vez ocasionado el problema de no disponer de herramientas para el manejo adecuado y completo de un conjunto de documentos, la migración de los mismos a otro formato, de ser posible, seguramente genere mucho trabajo y obtenga un resultado limitado.

Esto es especialmente cierto en casos en que la especificación de los formatos no está disponible completamente, o cuando simplemente no existen implementaciones que los contemplen adecuadamente.

Por lo tanto, los formatos deben elegirse previendo que el documento generado pueda ser utilizado en el futuro con el contenido y la calidad originales.

Acompañar tendencia hacia software libre

La utilización de un determinado formato genera la necesidad para todos los destinatarios de disponer de una herramienta que lo interprete adecuadamente.

Esto, sumado a la tendencia hacia el software libre, implica que deban existir herramientas libres que procesen sin limitaciones los formatos elegidos.

No propiciar la piratería de software

En la misma línea de lo anterior, la elección de formatos no debe fomentar la piratería de software. Esto se da, en los hechos, cuando se fuerza a utilizar herramientas de costo elevado para modificar e, incluso, visualizar correctamente un documento.

Requisitos

A partir de los objetivos estratégicos marcados y de otras características deseables, se generan los requisitos indicados en este punto para los formatos a elegir.

1. Formato abierto

Para dar ciertas garantías de perdurabilidad de los datos, deben utilizarse formatos abiertos. Como se desprende de lo expresado en el punto anterior, ésta es la única forma de resguardo frente a los avatares y decisiones de las empresas comerciales.

2. Existan lectores y generadores:

a. libres

La elección de formatos debe enmarcarse dentro de la tendencia hacia el software libre. Como consecuencia inmediata, no se fuerza a adquirir productos comerciales a los destinatarios (internos y externos), con el consiguiente fomento de la piratería informática.

b. de buena calidad

La decisión resultante no debe forzar a los usuarios a utilizar herramientas de baja calidad, lo cual, además de los perjuicios obvios, podría generar el incumplimiento de la política.

c. muy difundidos

Es deseable evitar la necesidad de capacitación y adaptación a nuevas herramientas, así como generar requerimientos especiales a los destinatarios.

d. disponibles en las plataformas más usadas como estación de trabajo (Windows, Macintosh, Linux)

No debería limitarse la elección de la plataforma para la estación de trabajo del usuario. Por lo menos, deben existir implementaciones de herramientas adecuadas para las más utilizadas: Windows, Macintosh, Linux.

Recomendación

Los formatos que cumplen con todos los requisitos son:

- **Texto plano ASCII**

Este formato refiere a una secuencia de caracteres en codificación ASCII de 7 u 8 bits. No incluye otras representaciones de caracteres.

- **HyperText Markup Language (HTML)**

La recomendación no incluye extensiones propietarias, no especificadas por los estándares publicados.

El HTML está regulado, desde 1996, por la organización W3C.

Ambos formatos, si bien cumplen con todo lo requerido, tienen algunas limitaciones que los hacen inadecuados para ciertos casos.

En el caso del texto plano, es un formato poco potente, acotado a transmitir información textual sin destaque.

En cuanto al HTML, tiene los siguientes inconvenientes:

- el mismo documento puede verse distinto en equipos distintos
- algunas herramientas muy usadas generan códigos propietarios, no compatibles con la especificación estándar; si bien esto no es un problema del formato, en los hechos puede ser difícil cumplir la norma
- muchas veces un documento se compone de varios archivos (caso típico: gráficos), por lo que es complicado para un usuario común realizar tareas como la publicación o intercambio de documentos sin perder parte en el proceso

Otra opción, que no tiene todos los problemas anteriores, es utilizar los formatos nativos del OpenOffice.org. Sin embargo, estos no cumplen con el requisito de que existan herramientas muy difundidas (si lo hace con el resto de los requisitos).

En aras de que la política no sea excesivamente restrictiva, entonces, se deben tomar en cuenta nuevos elementos.

Se nota, entonces, que para publicar un documento no es imprescindible la utilización de un formato abierto, si se mantiene el original en formato abierto. Para la publicación, puede usarse un 'formato publicado', que cumpla con el resto de los requisitos establecidos:

- lectores y generadores
 - libres
 - de buena calidad
 - muy difundidos
 - disponible en plataformas más usadas como estación de trabajo (Windows, Macintosh, Linux)

Con esta nueva premisa, puede utilizarse el **Portable Document Format (PDF)**, de Adobe.

El PDF es un formato creado por Adobe en 1993, derivado del Postscript. A partir de su creación, se han sucedido varias versiones, cuyas especificaciones son fácilmente accesibles.

Para el mismo, existen productos libres y gratuitos de excelente calidad, muy difundidos (en Windows, para leer, sobre todo el Acrobat Reader, que es gratuito, no libre), y disponibles en las plataformas más importantes

Teniendo un formato apto para la publicación, para generación y modificación de los documentos originales, así como para trabajo en grupo, puede eliminarse la necesidad de cumplir el requerimiento 2.c ('herramientas muy difundidas') que, acotado a este entorno de trabajo, es de menor importancia que el resto de los requisitos.

Por lo tanto, restan los siguientes:

- formato abierto
- lectores y generadores
 - libres
 - de buena calidad
 - disponible en plataformas más usadas (Windows, Macintosh, Linux)

En estas nuevas circunstancias, los formatos nativos del **OpenOffice.org (sxw, sxc, sxd, sxi, sxm, sxd)** son adecuados, por lo que se recomienda su uso.

Los formatos nativos de OpenOffice.org son basados en XML, y su especificación es fácilmente accesible.

OpenOffice.org es la implementación de referencia, pero también son nativos. en el terreno comercial, de StarOffice, perteneciente a Sun Microsystems.

Por qué no RTF

El Rich Text Format es un 'formato publicado', propietario de Microsoft. Según la catalogación realizada en el presente documento, se encuentra dentro de la misma categoría que el PDF. Por lo tanto, no cumple con los requisitos para ser usado como formato de documentos originales.

Muchas veces se maneja como una alternativa a los formatos cerrados. Sin embargo, no se recomienda su uso para publicaciones, ya que existen pocas herramientas que manejen adecuadamente las distintas versiones de RTF. Esto genera que el documento RTF generado por una herramienta no siempre es utilizable por otra.

Resumen de recomendación

Para documentos originales:

- Texto plano ASCII
- HTML
- Formatos nativos de OpenOffice.org (sxw, sxc, sxd, sxi, sxm, sxd)

Para publicaciones (debiendo mantener el original en alguno de los formatos anteriores):

- Texto plano ASCII
- HTML
- PDF