

# PRESERVACIÓN DIGITAL, la gran olvidada en las bibliotecas digitales

Mientras que la información en papel puede durar cientos o miles de años, la información en formato digital no durará más de unas décadas. Esto es debido principalmente al envejecimiento de los soportes digitales, a la obsolescencia de los formatos, del software o del hardware, así como a la falta de compatibilidad entre sistemas. Con la preservación digital se buscan soluciones para el acceso, recuperación y utilización de todos los materiales digitales almacenados.

Principales bibliotecas digitales en la actualidad.

www.cervantesvirtual.com

BIBLIOTECA VIRTUAL  
**MIGUEL DE CERVANTES**

Destacamos

San Luis de León

Santa Teresa de Jesús

BIBLIOTECA DIGITAL HISPÁNICA

20 de junio de 2008

Últimas noticias

Sindicar noticias **RSS** **BUSCADOR**

The Perseus Digital  
Gregory Crane, Editor-in-Chief  
Configure Display · Help Library · Tools · Publication

PERSEUS IS CHANGING! PLEASE VISIT PERSEUS

Perseus Digital

A graph of the places and dates men...

## ¿Qué es la preservación digital?

Ya en 1999, Baeza-Yates -en su libro *Information Retrieval*- definió una biblioteca digital como una colección de objetos digitales; las descripciones de esos objetos; un conjunto de usuarios; y sistemas que ofrecen una serie de servicios como la captura, indexación, catalogación, búsqueda, navegación, recuperación, entrega, archivo y *preservación*.

Por tanto, lo primero que debemos abordar es una definición de preservación, y la primera duda que se nos plantea es si por preservación digital nos referimos a la conservación de los documentos tradicionales en papel por medio de objetos digitales, o a la preservación de los propios documentos digitales. Realmente, es ambas cosas y con la misma importancia, aunque generalmente no se considera la versión digital un recurso de preservación primario.

### Diferencia entre preservación y copias de seguridad

Muy a menudo se confunden dos conceptos cuya finalidad es bien distinta: la preservación digital y la copia de seguridad. No es lo mismo lo que se ha de conservar para el futuro -y para obtener derivados-, que lo que se ha de guardar por precaución.

Por una parte, las copias de seguridad son una protección contra hechos catastróficos, tales como roturas de disco o pérdida de datos por apagones. Lo que se guarda como copia de seguridad en una biblioteca digital es tanto la información publicada en el servidor (recursos digitales junto con información catalográfica), como los recursos digitales en proceso de edición.

Sin embargo, por otra parte, la preservación digital se ocupa de guardar y conservar los recursos digitales que necesitaremos en el futuro. El ejemplo más claro lo tenemos en el caso de las imágenes gráficas. Interesa hacer copias de seguridad de las imágenes JPG (comprimidas de menor calidad) que colgamos en el servidor, pero también conviene hacer copias para preservar las imágenes TIFF de alta calidad que no publicamos, pero que generaron las JPG.

Si bien las copias de seguridad, al igual que las de preservación, se basan en la redundancia de la información mediante grabaciones periódicas, ni la organización de estas grabaciones ni la temporización es la misma. Las copias de seguridad se pueden hacer mediante copias integrales, incrementales o rotativas, y la periodicidad es generalmente alta (diaria), mientras que en el caso de las copias de preservación el método suele ser la grabación integral del material una vez finalizado y el copiado del mismo una vez al año renovando los soportes (CDs, DVDs, etc.).

**Library**  
 ef, Tufts University  
 s - Collaborations - Support Perseus

**Search**

**4.0 FOR THE LATEST VERSION**

**Library**

500 1000 1500  
 tioned in this collection

**About Perseus**  
 Perseus is an evolving digital library, engineering interactions through time, space, and language. Our primary goal is to bring a wide range of source materials to as large an audience as possible. We

**SH NATIONAL CORPUS**

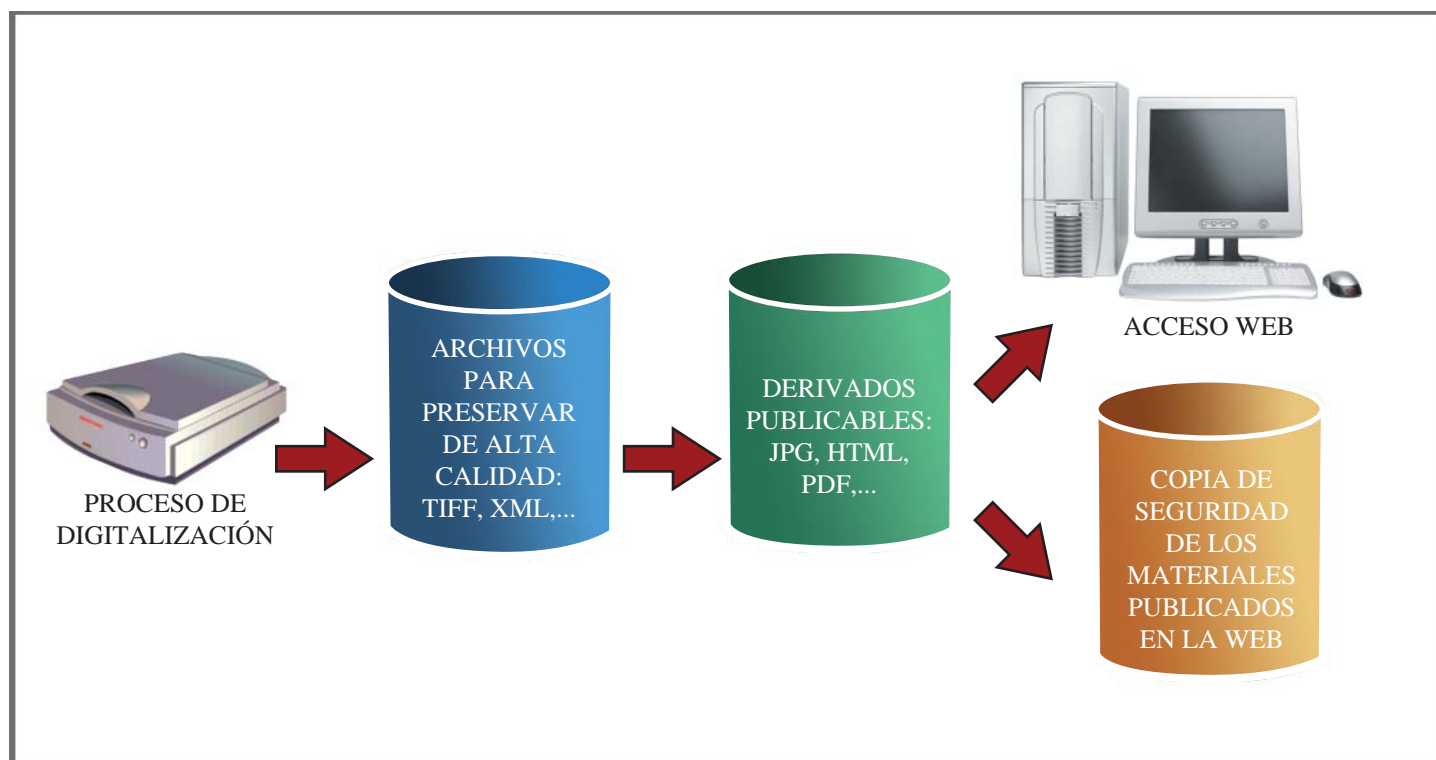
**The Corpus Using Obtaining Xaira/SARA FAQ Archive Contact Us A-Z**

**About the BNC**  
 The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of current British English, both spoken and written.  
[\[more\]](#)

**Search the**

**Results of your search**  
 The word was  
 This 47 column track for the party  
 1983 BNC 038342 (1983) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1984 BNC 038342 (1984) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1985 BNC 038342 (1985) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1986 BNC 038342 (1986) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1987 BNC 038342 (1987) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1988 BNC 038342 (1988) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1989 BNC 038342 (1989) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1990 BNC 038342 (1990) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1991 BNC 038342 (1991) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1992 BNC 038342 (1992) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1993 BNC 038342 (1993) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1994 BNC 038342 (1994) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1995 BNC 038342 (1995) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1996 BNC 038342 (1996) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1997 BNC 038342 (1997) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1998 BNC 038342 (1998) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 1999 BNC 038342 (1999) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".  
 2000 BNC 038342 (2000) as reported in QUARTY --, the office  
 New York, in the building form of the "Young One".

" lemma="the  
 lemma="cost">cost  
 "NN1" lemma="issue">is.  
 "gal">legal </w><w type="1  
 " lemma="should">should  
 at of">out of </w><w type="1  
 </w><w type="NN1" lemma="1  
 nt to">pursuant to </w><w



Diseño del procedimiento de producción de objetos en una biblioteca digital.

### Definición de las políticas y técnicas de preservación

Es primordial establecer una política de preservación digital, que se conciba como una responsabilidad institucional y donde se dé respuesta a las siguientes preguntas: ¿qué guardar y por qué guardarlo?, ¿dónde guardarlo?, ¿hasta cuándo guardarlo?, ¿cómo encontrarlo después?, ¿cómo hacer que se mantenga inalterado? y ¿cómo evitar que se vuelva obsoleto?

Cada autor tiene su propio enfoque sobre la preservación digital, pero casi todos coinciden en que las principales técnicas que se deben seguir son las siguientes:

- a) *Preservación de los sistemas originales.* Es el sistema más sencillo, ya que consiste en mantener en funcionamiento el ordenador con el que los objetos digitales fueron creados, almacenados, y pueden ser consultados. Sin embargo, cada vez será más difícil encontrar repuestos y las prestaciones del equipo se volverán obsoletas.
- b) *Emulación.* La emulación permite que el software original sea usado sin que el sistema original siga existiendo. Se debe preservar el emulador, el sistema operativo, la aplicación y los datos, y la pérdida de alguno de estos componentes hace inaccesible la información.

- c) *Migración.* La migración consiste en convertir la información a nuevos formatos. Es una medida contra la obsolescencia.
- d) *Replicado y rejuvenecimiento.* Ya que aquellos datos importantes de los que sólo existe una copia en un ordenador son altamente vulnerables, el replicado consiste en hacer copias periódicamente y almacenarlas en lugares seguros. El rejuvenecimiento consiste en la transferencia de la información a nuevos medios de almacenamiento, ya que los soportes de almacenamiento en los que se graba la información digital son efímeros.
- e) *Arqueología digital.* Se llama arqueología digital a la recuperación de la información a partir de fuentes de datos dañadas, fragmentadas o arcaicas. Es el remedio cuando no se han tomado las debidas precauciones y la información se ha estropeado.

### Preservación de los textos digitales. El uso de XML y TEI

Una biblioteca digital abarca una amplia variedad de formatos digitales y sus múltiples combinaciones. De este modo, podemos encontrar libros digitales en hipertexto, imágenes digitales, facsímiles digitales (imagen más texto), grabaciones de audio digital, grabaciones de vídeo digital y montajes de vídeo, audio y texto sincronizados. Además, cada tipo de objeto digital requiere un plan de preservación diferente, propio y adecuado. En el caso de los textos digitales es imprescindible meditar detenidamente

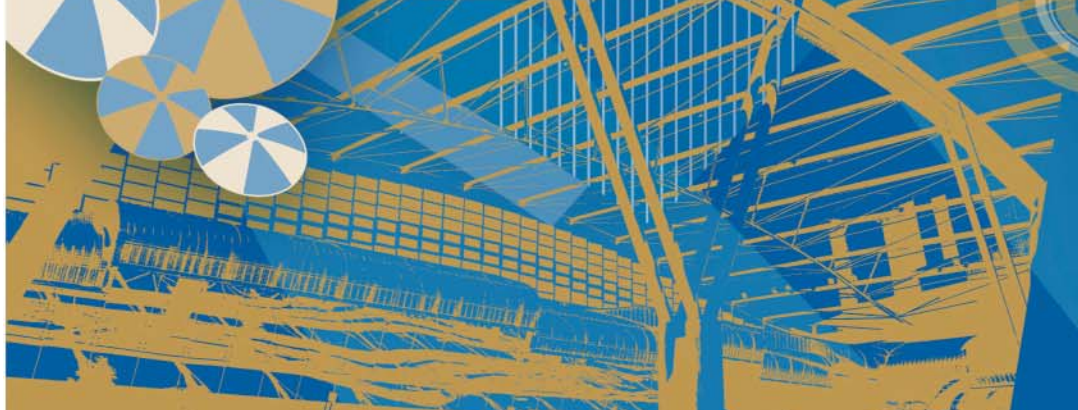
AYUNTAMIENTO DE MÁLAGA

# 2016.

**málaga**  
ciudad europea  
de la cultura

2016. 2016. 2016. 2016. 2016. 2016. 2016. 2016. 2016.

2016. 2016. 2016. 2016. 2016. 2016.



CARACTERÍSTICAS	HTML	SGML	XML
Se espera que dure	+	+	+
No es un lenguaje propietario	+	+	+
Se centra en la estructura	-	+	+
Es ampliamente usado	+	-	E
Es fácilmente convertible	+	+	+
Es directamente soportado por los navegadores	+	-	E
Es extensible	-	+	+
Es fácil de usar	+	-	+

Tabla de comparación de formatos HTML, SGML y XML (+=Sí, -=No, E=Se espera).

antes de elegir tanto el formato de codificación como el lenguaje de marcado.

Una de las obsesiones de muchos editores es la tecnología WYSIWYG (*What You See Is What You Get*), es decir, la apariencia final del documento. Sin embargo, el formato elegido finalmente debe favorecer la preservación de los textos, por tanto, es necesario que sea durable, fácilmente convertible a los nuevos formatos que aparezcan y que sea ampliamente usado por la comunidad bibliotecaria. Lo ideal, en este caso, es que se trate de un estándar, ya que garantizará la existencia de programas que lo soporten, por tanto, se recomienda huir de los formatos propietarios de los fabricantes de software, optando por formatos abiertos de dominio público.

Actualmente, el XML *-Extensible Markup Language-* es el lenguaje de marcado más aceptado para los textos en las bibliotecas digitales frente a otros como el SGML (más complejo) o el HTML (no adaptable o ampliable).

Tras la elección de un formato abierto, extensible, descriptivo y que permita almacenar y recuperar la información metatextual, el siguiente paso es la selección del lenguaje de marcado. Aquí podemos decidir entre crear nuestro propio lenguaje de codificación o adoptar y adaptar uno ya existente. Como no es cuestión de reinventar

la rueda, la mejor opción es elegir un lenguaje que esté siendo ampliamente utilizado y modificarlo, bien ampliándolo bien simplificándolo, según nuestras necesidades.

Actualmente, el lenguaje más utilizado es el TEI *-Text Encoding Initiative-* y su simplificación, el TEI-Lite. No se trata de un estándar, pero sin duda es el lenguaje de codificación más aceptado y utilizado en las mejores bibliotecas digitales (Oxford Text Archive, Biblioteca Virtual Miguel de Cervantes, Women Writers Project, The Perseus Digital Library, etc.).

### Vale la pena el esfuerzo

Para concluir, podemos hacernos eco de unas palabras de Peter Shillingsburg, en *Scholarly Editing in the Computer Age: Theory and Practice*, donde resume la importancia de todos los aspectos relacionados con la preservación digital que hemos comentado anteriormente. Así pues, según Peter Shillingsburg, "el editor que use un sistema de codificación universal para desarrollar una edición electrónica con una aplicación multiplataforma habrá creado una herramienta disponible para cualquier lector con acceso a un ordenador y habrá asegurado la longevidad de ese trabajo editorial para las generaciones de software y hardware venideras. Vale la pena el esfuerzo". ■

AUTOR: Sánchez Quero, Manuel. Universidad de Alicante.

TÍTULO: *Preservación digital, la gran olvidada en las bibliotecas digitales.*

RESUMEN: En este artículo se revisa brevemente tanto la definición del término preservación digital como la diferenciación entre los conceptos de *preservación* y de *copias de seguridad*. Su autor también realiza un recorrido por las principales técnicas de preservación para, finalmente, analizar la importancia de elegir correctamente el formato y lenguaje de marcado para garantizar la pervivencia de los textos digitales.

MATERIAS: Bibliotecas Digitales / Almacenamiento y Conservación de Colecciones / Nuevas Tecnologías / Digitalización.