

Seminario de Repositorios Institucionales y su Interoperatividad con el Repositorio Nacional

**“Preservación Digital Sustentable”
Cd. de México, 11 de septiembre 2017**

Juan Voutssás

- **Instituto de Investigaciones Bibliotecológicas y de la Información,**
- **Universidad Nacional Autónoma de México**

Problemática de los documentos digitales

- Sin tratamiento, la obsolescencia tecnológica hace que los documentos digitales sean inaccesibles en un lapso muy corto, a pesar de que permanezcan. Por lo mismo, no podemos dejar sus cadenas de bits inamovibles eternamente; esto los llevaría a ser inaccesibles con el tiempo. Para poder preservarlos, debemos cambiarlos de tiempo en tiempo.
- A pesar de los cambios, debemos poder garantizar que se mantienen auténticos, íntegros, exactos y confiables, ya que: 1) Se ha conservado su forma y contenido; 2) Tienen información que ayuda a verificar su identidad e integridad; 3) Se han protegido de acciones no autorizadas; 4) Se han protegido de pérdida o corrupción accidental.
- **InterPARES** = The International Project on Authentic Records on Electronic Systems. 2002. "Appendix 6. How to Preserve Authentic Electronic Records". In: "The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project". University of British Columbia. Sitio oficial del proyecto. Véase el apéndice 6, disponible en: <http://www.interpares.org/Vbook/index.cfm>

Características de los documentos digitales

- Los documentos digitales no existen como entidades físicas; su contenido, estructura y forma no están ya íntimamente entrelazados; están constituidos por componentes digitales vinculados, perfectamente distinguibles entre ellos, no necesariamente contiguos dentro del mismo objeto;
- Todos los documentos u objetos digitales residen eventualmente en un cierto soporte o medio físico –disco magnético, cd-rom, cinta, dvd, memoria de estado sólido, etc.– y están codificados en ese soporte por medio de señales en forma de cadenas de bits –magnéticas, ópticas, optomagnéticas, etc.–. Esto es, todos tienen objeto físico.
- Su naturaleza de acceso electrónico y remoto dificulta proteger la propiedad intelectual y la privacidad;

Características de los documentos digitales

- **Todos los documentos u objetos digitales requieren de uno o varios programas o aplicaciones informáticas para que esas cadenas de bits sean agrupadas y reinterpretadas bajo una lógica y unas reglas generalmente definidas por el tipo de los datos. Esto es, conforman un objeto lógico;**
- **Los documentos digitales, por su misma naturaleza, tienen una mayor fragilidad intrínseca;**
- **Por su misma naturaleza, es más difícil establecer la versión original o definitiva de un documento.**
- **Los documentos u objetos digitales deben contar –además de los datos o contenido– con un identificador claro y específico de cada uno de ellos, metadatos y eventualmente procedimientos.**

Objetos digitales:

- **“Objeto digital” u “objeto de información”** es cualquier entidad documental: texto, imagen, sonido, etcétera, que ha sido codificada numéricamente bajo algún formato y ensamblada junto con algún conjunto de metadatos de tal forma que puede ser almacenada, buscada, encontrada, y usada dentro de un sistema informático.
- **“Objeto digital”** es un objeto documental representado numéricamente por una agregación o secuencia de valores numéricos en forma de una o más cadenas de bits las cuales –bajo ciertas reglas de representación– contienen datos dentro de ese objeto, así como metadatos acerca de las propiedades del mismo y –cuando sea necesario–, los métodos o procedimientos para realizar operaciones sobre el objeto. El significado de las cadenas de bits como objeto documental está en función de la interpretación que hagan de ellas los programas de su entorno digital.
- **“Objeto digital”** es una entidad en la que uno o más archivos de contenido y sus metadatos correspondientes están unidos física y/o lógicamente mediante el uso de un contenedor digital.

Retos básicos de los documentos digitales:

- **Producir, compilar, gestionar, y distribuir documentos digitales accesibles y confiables.**
- **Preservarlos a pesar de los cambios y la obsolescencia tecnológica.**
- **Preservarlos íntegros, auténticos, confiables y accesibles.**

Factores que afectan a la Preservación (1):

- Los factores que afectan la preservación a largo plazo de los documentos digitales son múltiples y pueden agruparse de varias formas para su estudio. He aquí una agrupación en seis categorías: factores tecnológicos, culturales, sociales, legales, documentales, y económicos.¹
- Los factores **tecnológicos** son los que tienen que ver con aspectos de obsolescencia tecnológica, formatos, seguridad informática, etc.
- Los factores **culturales y sociales** presentan la componente humanista del problema, y tienen que ver con la falta de sensibilidad hacia la importancia de los documentos, el contexto social adecuado, la accesibilidad futura, los usuarios, la reducción de la brecha digital, transparencia, las restricciones a su uso, etc.
- ¹ Voutssas-M., Juan. 2009. “Preservación del Patrimonio Documental Digital en México”. México: UNAM: CUIB. 207 p. ISBN: 978-607-02-0583-5. Disponible en: http://132.248.242.6/~publica/archivos/libros/preservacion_patrimonio.pdf

Factores que afectan a la Preservación(2):

- Los factores que afectan la preservación a largo plazo de los documentos digitales son múltiples y pueden agruparse de varias formas para su estudio. He aquí una agrupación en seis categorías: factores tecnológicos, culturales, sociales, legales, documentales, y económicos.¹
- Los factores **legales** permiten enfocar las problemáticas documentales acerca de derechos de propiedad intelectual, secreto industrial, privacidad y protección de datos personales, jurisdicciones de los documentos, obligaciones de informar y preservar, etc.
- Los factores **documentales** contemplan las metodologías necesarias para la producción, selección, ordenamiento, gestión, conservación, preservación y recuperación documental.
- Los factores **económicos** tienen que ver con los aspectos financieros de la preservación: financiamiento, costos, presupuestos actuales, sustentabilidad, viabilidades futuras, etc.
- ¹ Voutssas-M., Juan. 2009. "Preservación del Patrimonio Documental Digital en México". México: UNAM: CUIB. 207 p. ISBN: 978-607-02-0583-5. Disponible en: http://132.248.242.6/~publica/archivos/libros/preservacion_patrimonio.pdf

Gestión y Preservación de documentos digitales:

- **Implementación de estrategias, políticas, sistemas y procedimientos de gestión documental digital.**
- **Implementación de estrategias, políticas, sistemas y procedimientos de preservación documental digital.**

Estrategias:

- Para ello las organizaciones requieren de estrategias de dos dimensiones; es necesario diferenciar entre unas *estrategias de gestión* y unas *estrategias de preservación*. Esto obedece de origen a planteamientos ampliamente difundidos por UNESCO en sus *Directrices para la Preservación del Patrimonio Digital* las que ofrecen un amplio marco de referencia para el establecimiento de estrategias de preservación para cada organización y sus tipos de documentos digitales (UNESCO, 2003).

Algunos modelos:

- “Sustainability of Digital Formats - Planning for Library of Congress Collections” - <http://www.digitalpreservation.gov/formats/index.shtml>
- El estándar ISO 19005-1:2005 (Gestión de documentos - Formato de archivo de documentos electrónicos para conservación a largo plazo, Parte 1: Uso de PDF 1.4 PDF/A-1-2) - <https://www.iso.org/standard/38920.html>
- UNESCO. “Carta sobre la preservación del patrimonio digital”. 2003. - http://portal.unesco.org/es/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html

Algunos modelos:

- El documento “Trustworthy Repositories Audit & Certification” del RLG-NARA Digital Repository Certification Task Group.
- La iniciativa NDIIPP - *The National Digital Information Infrastructure and Preservation Program* de la Biblioteca del Congreso de los E.U.A.
- El estándar ISO 16363:2012 - Define la práctica recomendada para organismos que evalúan la confiabilidad de los repositorios digitales, aplicable a toda la variedad de ellos; puede utilizarse también como base para la certificación.
- InterPARES 2 Project. 2008. “*Experiential, Interactive and Dynamic Records, Appendix 14: Chain of Preservation Model*”:
http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_part_5_modeling_task_force.pdf

Conservación y Preservación digitales:

- **Conservación Digital:** Tiene que ver con el soporte y forma del documento digital, y su alcance es siempre el corto-mediano plazo. Sinónimo de “mantenimiento”. (objetos físicos)
- **Preservación Digital:** Tiene que ver con el contenido del documento, y su alcance siempre es el largo plazo. (objetos lógicos).
- “Glosario Inter pares de Preservación Archivística Digital”. 2014. UNAM : Instituto de Investigaciones Bibliotecológicas y de la Información. Disponible en: http://iibi.unam.mx/archivistica/glosario_preservacion_archivistica_digital_v4.0.pdf

Estrategias de Conservación:

- Estas estrategias tienen que ver con las acciones tomadas para anticipar, prevenir, detener o retardar el deterioro del **soporte** de obras digitales a **corto-mediano plazo** con objeto de mantener su permanencia y accesibilidad, así como la estabilización tecnológica, migración a nuevos soportes, sistemas y formatos digitales para garantizar la trascendencia de los contenidos. Sinónimo de “mantenimiento”. Selección, uso y cuidado de los soportes digitales, siguiendo estándares internacionales.
- Transferencia de datos periódicamente hacia nuevos soportes digitales: (**réplica, refrescado**); actualización de formatos, formatos auto-contenidos, (Teoría de los Formatos Universales de Preservación)
- *ISO 18925:2013 “Imaging materials - Optical disc media - Storage practices”*
- *Guidelines for Physical Digital Storage Media. Version 1.0. 2011. Libraries and Archives Canada. Sitio Oficial de la Organización. Disponible en: <http://www.collectionscanada.gc.ca/>*

Características de la información para preservación:

- Aunque varían de sector en sector, en términos generales se distinguen seis características que la información digital debe tener y mantener con el tiempo:¹
- permanencia (la información existe),
- accesibilidad (la información es reinterpretable y usable),
- disponibilidad (la información puede ser encontrada y accedida),
- confidencialidad (privacidad),
- autenticidad (integridad), y
- aceptabilidad (no repudio).
- Todas ellas contribuirán a la usabilidad de la información

¹ Voutssas-M., Juan. 2010. "Preservación Digital y Seguridad Informática". (pp.133-138) En: "Investigación Bibliotecológica". México: UNAM: CUIB, vol. 24, núm. 50, ene/abr. 2010. pp.127-155. ISSN: 0187-358-X. Disponible en: <http://132.248.242.6/~publica/archivos/50/ibi002305007.pdf>

Estrategias de Preservación:

- ***Estas estrategias consisten básicamente en “un conjunto de principios, políticas, reglas y procedimientos así como las herramientas y mecanismos utilizados para implementarlas y que han sido adoptadas por una institución o programa para mantener a largo plazo los componentes digitales y su información relacionada, así como para reproducir documentos y/o agregaciones de ellos producidos mediante la interpretación de controles externos aplicados a los documentos seleccionados para su preservación”.***
- Glosario InterPARES.

Estrategias de Preservación:

- **Conway (1996) estableció: “el mundo digital transforma el concepto tradicional de preservación, que ya no consiste en proteger la integridad física del objeto, sino en especificar la producción y mantenimiento del objeto cuya integridad intelectual es la principal característica”.¹**
- **La ALA –American Library Association– define ahora la preservación digital como: “la reinterpretación exacta de contenido autenticado a largo plazo por medio de la combinación de políticas, estrategias y acciones que garanticen el acceso futuro a contenidos digitales a pesar de los desafíos de la obsolescencia tecnológica y lo efímero de los soportes”.²**
- **El ICA –Consejo Internacional de Archivos– define preservación digital como: “el proceso específico para mantener los materiales digitales durante y a través de las diferentes generaciones de la tecnología a través del tiempo, con independencia de donde residan”.³**
- ¹ Conway, Paul. 1996. “Preservation in the Digital World”. CLIR – Council on Libraries and Information Resources Reports. Disponible en: <https://www.clir.org/pubs/reports/conway2/index.html>
- ² ALA - American Library Associations and ALCTS - Association for Library Collections and Technical Services. 2007. Disponible en: <http://www.ala.org/alcts/resources/preserv/defdigpres0408>
- ³ ICA - International Council on Archives. ICA Terminology Database. Disponible en: <http://www.ciscra.org/mat/mat/term/3062>

Estrategias de Conservación:

- **1) Clara asignación de responsabilidades**
- **2) Adquiera la infraestructura técnica adecuada**
- **3) Mantenimiento, actualización y cambio de sistemas**
- **4) Transferencia de datos periódicamente hacia nuevos soportes de almacenamiento**
- **5) Adopción de condiciones apropiadas para el almacenamiento de soportes**
- **6) Redundancia de medios y ubicaciones**
- **7) Seguridad informática en los sistemas**

Estrategias de Preservación:

- **1. Preferir estándares**
 - **2. Usar formatos auto-descriptivos,**
 - **3. Restringir el número de formatos que pueden usarse**
 - **4) Migrar, emular, etc.**
- Voutssás M., Juan. “Estrategias de Conservación y Preservación de Documentos de Archivo Digitales: Elementos Técnicos”. Capítulo in-extenso en libro: “Los Archivos Digitales : Una visión integradora”. pp. 169-235. Puebla, México: BUAP. 2014. 239 p. ISBN: 978-607-487-785-4. Disponible en: http://iibi.unam.mx/voutssasmt/documentos/archivos_digitales_3_corto.pdf

Características de la metainformación para preservación (metadatos)(1):

- **Contenido** - tipos y géneros de contenidos, multimedialidad (variedad de tipos de datos en un mismo documento), multidimensionalidad (múltiples aproximaciones o enfoques a un documento), etc.;
- **Contexto** - información acerca de cómo, cuándo, dónde, por quién y en qué circunstancias se produce el documento; por qué se produce; para qué se usa, su relación con otros documentos en un corpus;
- **Proveniencia** - información acerca del origen del documento: persona(s) o autoría, organización, procedimiento, etc.;
- **Estructura** - Información acerca de cómo debe ser la apariencia del documento, cómo se integran intelectualmente las partes de documentos complejos, hipervínculos, etc.;
- **Almacenamiento** - información acerca de características particulares del almacenamiento actual del documento, o cómo debe almacenarse (dispositivos especiales, algoritmos de compresión, etc.);
- **Accesibilidad** - medida en que el documento puede ser accedido, reproducido, copiado, modificado, transmitido, etc.; gratuidad o economía del mismo, etc.;

Características de la metainformación para preservación (metadatos)(2):

- **Preservación** - información acerca de cómo se debe preservar o se ha preservado ese documento; periodos de conservación, eliminación, etc.;
- **Autenticidad** - información acerca de los atributos que el documento tiene y debe conservar a lo largo del tiempo para ser considerado auténtico u original (autoría, identidad, integridad, estabilidad, custodia, firmas, DOI, ISBN, etc.).
- **Modificabilidad** - información acerca del grado en que el documento puede ser modificado, en qué condiciones y por quién (wikis, folksonomías, páginas web dinámicas, etc.);
- **Tecnología** - datos acerca del entorno tecnológico de producción y uso del documento: sistemas, equipos, y programas informáticos; formatos, estándares, escalas, procesos técnicos efectuados, envolturas, firmas o sellos digitales, sumas verificadoras, “hashing”, cifrados, certificados, etc.;
- **Modularidad** - medida en que el documento puede ser dividido en varias partes para su uso o interpretación.
- **Interoperatividad** - medida del grado en que el documento puede ser utilizado a lo largo de múltiples plataformas y estándares con pocas o nulas restricciones;
- **Dinamismo** - medida en que el documento se actualiza y cambia con las experiencias del usuario (bases de datos, páginas web dinámicas, etc.)

Características de la metainformación para preservación (metadatos)(3):

- Más allá del contenido, muchos de estos atributos se inscriben en forma inherente dentro del documento –como el nombre o identificador del archivo informático que lo contiene o como el formato principal del documento: doc, pdf, jpg, wav, mp3, etc.– y muchos otros se agregan en forma de metadatos específicos.
- Para cada tipo de documento se contemplan y/o recomiendan metadatos específicos con mayor o menor énfasis en algunos dependiendo de la importancia que se les otorga dentro de esos documentos: Así, se distinguen metadatos específicos descriptivos, documentales, de contexto, estructurales, funcionales, jurídicos, administrativos, tecnológicos, de ubicación, de autenticidad, de preservación, de privilegios o restricciones al acceso y uso, de procedimiento, etc.
- Dependiendo de las características y volumen, los metadatos pueden embeberse dentro del documento en sí o crearse en documentos aparte con sus correspondientes vínculos.
- “*PREMIS Data Dictionary for Preservation Metadata, Version 3.0*”, Nov. 2016. Library of Congress. Disponible en: <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>

La preservación de documentos digitales:

- **El gran problema de la preservación de recursos documentales digitales a largo plazo no reside mayormente en factores de tipo técnico –espacio de almacenamiento, formatos, servidores, respaldos, etc.–, sino en elementos de asignación y delegación de responsabilidades –quiénes deben hacerlo, cómo y cuándo–; además, en elementos de barreras legales derivadas de las legislaciones autorales, así como en problemas de jurisdicciones legales.**
- **La preservación de documentos digitales NO puede empezar a diseñarse y construirse tiempo después de que los repositorios están operando: debe diseñarse e implementarse al momento de construir el repositorio.**
- **La preservación de documentos digitales obedece a un plan específico al efecto: no es solo seguridad informática.**

Gracias

Juan Voutssás M.
voutssas@unam.mx

**IIBI - Instituto de Investigaciones Bibliotecológicas
y de la Información,
Universidad Nacional Autónoma de México**