



UNLP



## **Bibliotecas y Repositorios Digitales**

Tecnología y Aplicaciones

[sedici.unlp.edu.ar](http://sedici.unlp.edu.ar)

# Bibliotecas y repositorios digitales



## Capítulo 7:

- Crear conciencia en los profesionales, usuarios, funcionarios de bibliotecas y archivos, políticos, investigadores, etc., sobre los riesgos que conlleva mantener en el tiempo los objetos digitales y dar accesibilidad permanente a los mismos.
- Analizar formatos y estándares para cumplir con el objetivo de preservación.





# Preservación del contenido de los RIs



# Introducción



- En la actualidad, los recursos que se generan como resultado de los conocimientos de las personas y de sus expresiones “nacen”, cada vez más, en formas digitales, sean de carácter cultural, educativo, o engloben información de diferentes áreas del saber, ya sean de naturaleza técnica, artística o administrativa. Los productos de origen digital pueden no contar con un respaldo físico, por ejemplo en papel.
- Muchos de estos recursos son valiosos y constituyen un verdadero patrimonio a conservar a futuro para la sociedad.

Además del acceso abierto al material de investigación la preservación digital es una motivación importante para crear RIs y para asegurar que los materiales de investigación digitales estén disponibles y sean accesibles a largo plazo.





# La preservación de los contenidos

Criterios tradicionales para documentos en papel: en los documentos tradicionales en papel se habla de “negligencia benigna”: el olvido de un manuscrito en un arcón, puede que lo preserve.

En los documentos digitales, la negligencia benigna no sirve: un disco olvidado 5 años probablemente al encontrarlo no se puede recuperar su información.

No a la negligencia benigna.

No a la preservación basada en las condiciones ambientales.

No se conserva para cualquier usuario futuro sino para una comunidad designada: el conjunto de los consumidores que tienen que entender la información almacenada.

No necesariamente se conserva la integridad externa del documento sino las propiedades significativas.

Se debe asegurar la integridad y autenticidad del recurso





# Problemas en la preservación de OD

1. La propia naturaleza de los objetos digitales los hace efímeros.
2. La obsolescencia de los medios informáticos: dado que los OD siempre están mediados por la tecnología que cambia constantemente; una inadecuada vigilancia o falta de transformaciones puede dejarlos inaccesibles. La incompatibilidad entre sistemas nuevos y antiguos sumado a que los formatos, medios de soporte, software y hardware quedan obsoletos en poco tiempo.





# Preservación digital

La preservación digital supone, en relación con la conservación de los documentos en papel, un importante reto tecnológico, pero también de otros tipos:

legal, permisos de los autores para realizar las transformaciones necesarias

económico, ¿quién financia el personal y las acciones para la preservación?,

organizativo ¿de quién es la responsabilidad de cada acción?  
¿cómo se asegura la continuidad de las decisiones?)

**(Keefer; Gallart, 2007).**





# Etapas en la preservación

1. Archivar los documentos digitales  
gestión documental
2. Preservar el *bitstream*
3. Garantizar el acceso a largo plazo





# La preservación supone que:

1. Los datos se mantendrán en el repositorio sin sufrir daños, sin perderse o sin ser alterados de forma malintencionada/o no.
2. Los datos podrán ser localizados y entregados al usuario.
3. Los datos podrán ser interpretados y comprendidos por el usuario.
4. Las metas 1, 2 y 3 serán realizables a largo plazo.



# Preservación digital



La preservación digital se define como el conjunto de prácticas de naturaleza política, estratégica y acciones concretas, destinadas a asegurar la preservación, el acceso y la legibilidad de los objetos digitales a largo plazo.



# Noción de preservación de UNESCO



“La preservación digital puede definirse como el conjunto de los procesos destinados a garantizar la continuidad de los elementos del patrimonio digital durante todo el tiempo que se consideren necesarios”.

“La mayor amenaza para la continuidad digital es la desaparición de los medios de acceso. No puede decirse que se han conservado los objetos digitales si, al haber dejado de existir los medios de acceso a ellos, resulta imposible utilizarlos. El objetivo de la preservación de los objetos digitales es mantener su accesibilidad, es decir, la capacidad de tener acceso a su mensaje o propósito esencial y auténtico”. (UNESCO, 2003: p. 37)

# Objeto digital

## Acciones en su ciclo de vida para mantener el acceso

### OD Y METADATOS DE PRESERVACIÓN

Debe mantenerse en el repositorio de manera **segura**

Deben guardarse las relaciones que vinculen al objeto con otros

El repositorio debe tener los derechos suficientes para sostener el **acceso** al objeto

Si hay un cambio debe saberse **quién** lo efectuó



<input checked="" type="checkbox"/>	Sum sit amet, consectetur sectetur
<input type="checkbox"/>	Dolor sit amet, consectetur adipis
<input type="checkbox"/>	Lorem ipsum dolor sit amet, con sectetur adipiscing

#### Autenticidad

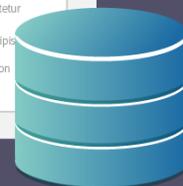
Mediante la documentación de su procedencia

Debe conocerse su **creador**

Debe poder ser **localizado** y **entregado** al usuario

Su soporte deber ser **compatible** con los sistemas actuales

Las estrategias de **emulación** y **migración** requieren datos sobre los objetos originales y sus entornos



“UNA METODOLOGÍA DE EVALUACIÓN DE REPOSITORIOS DIGITALES PARA ASEGURAR LA PRESERVACIÓN EN EL TIEMPO Y EL ACCESO A LOS CONTENIDOS”

Autora: Ing. Marisa R. De Giusti  
Directora: Dra. Silvia Gordillo

# Preservación de los contenidos de un RI



Criterios nuevos para los recursos digitales:

- que la institución tenga pleno derecho a manipular los datos para asegurar su acceso en entornos informáticos del futuro;
- que el recurso sea de un formato legible actualmente y previsiblemente en el futuro;
- que el recurso esté en un soporte gestionable para su transferencia y/o almacenamiento;
- que el recurso disponga de documentación, incluyendo los metadatos.

# Metadatos y metadatos de preservación



Los objetos digitales cambian, y dichos cambios deben registrarse y validarse para asegurar la autenticidad del objeto, por lo que también es preciso incorporar metadatos de procedencia y autenticidad. Dado que cualquier actividad de preservación está limitada por los derechos de propiedad intelectual, se hace necesario incluir metadatos para la gestión de los mismos.





# Preservación del contenido de los RIs

¿Qué materiales hay en los RIs?

resultados de la investigación (tesis, e-prints);  
objetos de docencia y aprendizaje;  
datos no elaborados;  
fondos digitalizados;  
material administrativo.

¿Qué materiales se tienen que preservar a largo plazo?

Criterios tradicionales para tomar la decisión sobre la preservación a largo plazo, principalmente los factores de

- valor
- pertinencia
- uso
- Otros condicionantes: misión, disponibilidad de recursos humanos, económicos, materiales, obligaciones legales o contractuales.





# Preservación del contenido de los RIs

## Selección de recursos para su preservación

¿Qué formatos? ¿qué versiones? ¿qué material adicional incluir?

¿Qué atributos se quieren preservar?

datos, funcionalidad

apariencia, esencia

La decisión dependerá de la misión institucional, las necesidades de la comunidad de usuarios, la capacidad técnica/ tecnológica institucional y los recursos disponibles.





# Un avance: estándares

El estándar 14721 (OAIS), los metadatos PREMIS y las directrices para la preservación, en conjunto con el esquema METS, constituyen el marco ideal para la gestión de un repositorio, para asegurar su interoperabilidad y dar preservación a sus contenidos.



# Problemas en la preservación: software



- Muchos problemas en lo relativo a la preservación derivan de una configuración deficiente del software que soporta el repositorio. Es necesario revisar las facilidades del software que soporta el repositorio en comparación con el modelo de preservación OAIS y realizar las personalizaciones necesarias para cumplir con algunos requerimientos del plan de preservación no brindados de forma nativa. Lo mismo con PREMIS.



# Preservación de contenido



Hay una muy importante necesidad de preservar el contenido digital en el tiempo, con el objetivo de conservarlo accesible frente a riesgos como

Incendios, Inundaciones, etc

Robos

Problemas de hardware (rotura de discos, etc.)

Cambios tecnológicos constantes

*Es un proceso continuo*

Además de lo técnico, los esfuerzos de preservación incluyen retos legales, económicos e institucionales.



# Obsolescencia



Es el resultado de la evolución de las tecnologías: a medida que surgen nuevas tecnologías, las viejas van quedando en desuso y se vuelven obsoletas.



Mantener tecnologías obsoletas en funcionamiento puede ser justificado en casos particulares, pero no en la mayoría.

Cornell University Library creó la "Cámara de los horrores"  
<http://dpworkshop.org/dpm-eng/oldmedia/chamber.html>



# Preservación de contenido

## Digital obsolescence



Mantener tecnologías obsoletas requiere conservar

- Hardware
- Software (aplicaciones, librerías, sistema operativo, etc)
- Documentación (manuales, instructivos, etc)
- Personal con la capacitación y las habilidades necesarias para trabajar en ese entorno obsoleto

Suelen ser opciones muy difíciles de mantener y muy costosas.

**En general no suele ser la mejor opción**



# Preservación de contenido

## Estrategias



Las formas de atacar los problemas de preservación, y en particular los problemas de obsolescencia, son:

- Migración continua
- Adhesión a estándares internacionales
- Emulación
- Encapsulamiento
- Metadatos de preservación
- Políticas de backup



# Preservación de contenido

## Migración continua



Migrar la información de una tecnología a la siguiente de forma continua, evitando así la obsolescencia.

- Es una de las opciones de mayor uso
- Asegura el acceso en todo momento (los datos son siempre accesibles mediante una tecnología actual)
- Requiere transformación de los datos originales
- Decisiones sobre qué se desea preservar



# Preservación de contenido

## Adhesión a estándares internacionales



Es una estrategia que busca apoyarse en la afirmación de que los estándares internacionales son relativamente estables en el tiempo.

- En la actualidad, los estándares evolucionan casi tan rápido como las tecnologías
- Es una estrategia que debería usarse en combinación con otras
- Según la National Initiative for Networked Cultural Heritage, los formatos que no serán declarados obsoletos (al menos en un futuro cercano) son: TIFF y PDF sin compresión, y ASCII y RTF sin compresión, para imágenes y texto respectivamente.



# Preservación de contenido

## Emulación



Se trata de imitar las características y capacidades de un software y/o hardware, de modo que los procesos "crean" que están funcionando en la plataforma original.

- No hay necesidad de modificar los datos originales (como en la migración), manteniendo la integridad de la información.
- Una vez que se archivaron los datos, solo hay que asegurarse que el soporte físico utilizado siga siendo accesible.
- Se puede usar un mismo emulador para múltiples objetos del mismo tipo.



# Preservación de contenido

## Encapsulamiento



Se basa en agrupar cada objeto a preservar junto con todos los elementos (incluso software) necesarios para asegurar su acceso en el tiempo.

Como elementos a encapsular podemos tener:

- Especificaciones del formato de archivo.
- Instructivos relacionados a la emulación necesaria.
- Información de configuración de alguna herramienta en particular.
- Software de emulación.
- Especificaciones de hardware.



# Preservación de contenido

## Metadatos de preservación



Generalmente considerados como metadatos administrativos

Buscan registrar información relativa a la evolución de los recursos en el tiempo según las acciones de preservación aplicadas, incluyendo información sobre formatos, usos, actividades de preservación realizadas, responsables de dichas actividades en el tiempo, etc.

Varias iniciativas:

- PREMIS: PREservation Metadata: Implementation Strategies
- OAIS: Open Archival Information System
- NEDLIB: Networked European Deposit Library



# Preservación de contenido

## Políticas de backup



Los riesgos de pérdida de datos por eventos desafortunados siempre son posibles:

- Incendios
- Inundaciones
- Robos
- Fallas de hardware

Para disminuir esos riesgos es necesario contar con un sistema de backups (datos, configuración, documentación, etc)

- Incremental
- Espejo



# Factores que gravitan para elegir formatos



1. **Divulgación.** Grado en que existen las especificaciones completas y herramientas para validar la integridad técnica y son accesibles para aquellos que crean y mantienen los contenidos digitales. Hay diferentes niveles de divulgación, lo más importante no es la aprobación por un organismo de normalización reconocido, pero sí la existencia de la documentación completa.
2. **Aprobación.** Grado en que el formato ya es utilizada por los principales creadores, difusores, o usuarios de los recursos de información.
3. **Transparencia.** Grado en que la representación digital está abierta al análisis directo con herramientas básicas.
4. **Auto-documentación.** Que permita agregar metadatos administrativos descriptivos, técnicos y otros porque esto es útil para la preservación.
5. **Dependencias externas.** Grado en que un formato en particular depende de hardware en particular, el sistema operativo o el software para la representación.
6. **Impacto de Patentes.** Grado en que la capacidad de las instituciones archivísticas para sostener contenido en un formato que puede ser inhibido por patentes.
7. **Mecanismos de protección técnica.** Implementación de mecanismos como el cifrado que impiden la conservación de los contenidos.

# Selección de formatos: generalidades



Para asegurar la preservación de la información, el formato elegido debe ser legible por una aplicación durante el mayor tiempo posible. Esto implica evitar formatos propietarios cerrados, como documentos en Word (en cualquiera de las versiones). Lo recomendable es utilizar formatos propietarios pero abiertos, como el formato TIFF, o mejor aún, formatos no propietarios como el ASCII. La razón es evidente: los documentos Word, por ejemplo, es necesario migrarlos a otros formatos Word más modernos no pudiéndose garantizar que estas migraciones no ocasionen pérdidas de información o cambios en la presentación de los documentos. Además existe otra razón importante para evitar el uso de formatos propietarios ¿Quién puede asegurar la supervivencia en un futuro de empresas por importantes que éstas sean en la actualidad?



# Selección de formatos: generalidades



La utilización de un formato de codificación simple y universal como XML permite perpetuar los documentos electrónicos. XML es el formato ideal ya que además de ser un formato no propietario, y por tanto ofrecer garantía de preservación de la información (ASCII), permite estructurar la información y el intercambio de información a todos los medios.



# Selección de formatos : generalidades



Para asegurar la integridad de los documentos que contienen objetos electrónicos (imágenes, sonidos, modelos, fórmulas, hiperenlaces..) se debe emplear la misma filosofía que con la información textual. Los formatos imagen considerados mejores para la conservación son el TIFF (Tagged Image File Format) que su compresión no experimenta ninguna pérdida de calidad, el PNG (Portable Network Graphics), cuya compresión experimenta apenas pérdidas en la resolución y además es muy ligero y el JPEG.



# Selección de formatos : generalidades



En cuanto a los Formatos mixtos los mejores son el Postscript, que puede ser enviado a cualquier periférico que soporte este lenguaje, sin tener en cuenta su resolución, produciendo un resultado adaptado a cada tipo de periférico y el PDF (Portable Document Format), basado en el Postscript, propietario pero abierto de la casa Adobe y que facilita un programa gratuito para poder leer este tipo de documentos. Para la preservación, se recomienda especialmente el PDF/A



# Sobre PDF/A



PDF/A es un estándar para codificar documentos en un formato “impreso”, que es portable entre sistemas y ampliamente usado para distribución y archivado de documentos. Sin embargo, la pertinencia de un archivo PDF para preservación depende de las opciones elegidas cuando el PDF fue creado: en particular, si se embebieron las fuentes necesarias para renderizar el documento, si se usa encriptación y si se preserva información adicional del documento original, más allá de lo que se precisa para imprimirlo.



# Sobre PDF/A



El estándar PDF/A no define una estrategia de archivado o los objetivos de un sistema de archivado. Sí identifica un “perfil” para documentos electrónicos que asegura que los documentos pueden ser reproducidos exactamente de la misma manera durante años. Un elemento clave para esta reproductibilidad es que los documentos PDF/A deben ser 100% auto-contenidos: esto significa que toda la información necesaria para mostrar el documento de la misma manera cada vez, debe embeberse dentro del archivo. Esto incluye (pero no se limita a) todo el contenido (texto, imágenes rasterizadas, gráficos vectorizados), fuentes, información de color, etc. Un documento PDF/A no puede jamás depender de información de fuentes externas (por ejemplo, programas fuente o *streams* de datos), aunque se permite que tengan anotaciones (como hipertextos) que enlacen a documentos externos.



# Otros elementos de la compatibilidad con PDF/A



El contenido de audio y video está prohibido.

Java script y enlaces a archivos ejecutables están prohibidos.

Todas las fuentes deben estar embebidas, y también deben ser legalmente embebibles para renderización ilimitada y universal. Esto significa para un usuario poder abrir el documento y que los caracteres se muestren de manera correcta (de aquí a X años) aunque no tenga esa tipografía en su computadora.

Los espacios de colores deben ser especificados de una manera independiente del dispositivo.

Se prohíbe la encriptación.

El uso de metadatos basados en estándares se mantiene.



# Otros elementos de la compatibilidad con PDF/A



Las referencias a contenidos externos están prohibidas.

La compresión de imágenes LZW y JPEG2000 están prohibidas en PDF/A1, pero JPEG 2000 se permite en PDF/A2.

Capas y objetos transparentes están prohibidos en PDF/A1 pero no en PDF/A2.

Firmas digitales provisionales se permiten en PDF/A2.

Los archivos embebidos están prohibidos en PDF/A1, pero PDF/A2 permite embeber archivos PDF/A.

PDF/A3 permite embeber cualquier formato como XML, CSV, CAD, archivos de Word, planilla de cálculo, otros PDF/A, etc. como objetos archivados completos.





# Niveles de cumplimiento

PDF/A1 posee dos niveles de cumplimiento:

**PDF/A-1a** aplica corrección semántica y estructura.

Cada carácter debe tener su equivalente Unicode. La estructura se expresa por medio de etiquetas.

**PDF/A-1b** aplica integridad visual.



# National Archives of Australia



Your story, our history

Explore:

Search the collection	Go >
Search the website	Go >

- Home
- The collection ▾
- Records management ▾
- Visit us ▾
- About us ▾

[Home](#) > [About us](#) > [Our organisation](#) > [Accountability and reporting](#) > [Operations and Preservation policies](#)

Our organisation

Advisory Council

Leadership team

Accountability and reporting

Access to information under the Freedom of Information Act

Annual Reports

Budget

Contracts

Corporate Plan

Departmental indexed file list

Grants

Information Publication Scheme

## Digital Preservation Policy

### Preserving Archival Digital Records Transferred from Commonwealth Agencies

July 2009

#### Version

1,2, July 2011

#### Contents

- [1. Policy statement](#)
- [2. What is digital preservation?](#)
- [3. Objectives of digital preservation at the National Archives](#)
- [4. Challenges to the preservation of digital records](#)
- [5. Scope](#)
- [6. Digital preservation principles](#)
- [7. Digital preservation approach](#)
- [8. Preservation formats](#)
- [9. Records which cannot be converted to an open format](#)
- [10. Essential characteristics of digital records](#)
- [11. Cooperation / collaboration](#)



# Preservation file formats

## Australia



The digital preservation file formats we have selected are listed in the following table.

The preferred preservation file formats are the targets for file format conversions. Both the preferred and acceptable preservation file formats are based on open standards and any digital records received in these formats are preserved as they are.





Format Category	Preferred Preservation open file format <sup>1</sup> (preserved as they are)	Files in these formats are converted to our preferred preservation open file format	Acceptable open file format <sup>2</sup> (Not converted, preserved as they are)
	<p>An index of the contents is created as XML. The content of the archive is converted according to the appropriate preservation file format. Free Lossless Audio Codec (flac).</p> <p>Not yet decided.</p> <p>XML and XSL files are created for each email. Any attachments are converted according to the appropriate preservation file format.</p>	<ul style="list-style-type: none"><li>• Compressed archives (gzip, bzip2, war, zip)</li><li>• Uncompressed archives (jar, tar, zip)</li><li>• Audio Interchange File Format (aiff)</li><li>• Broadcast Wave File (bwf)</li><li>• MPEG-2 audio layer 3 (mp3)</li><li>• Speex (spx)</li><li>• Vorbis (ogg, oga)</li><li>• Wave Audio File (wav)</li><li>• Drawing (dws, dwt, dwg)</li><li>• Design Web Format (dwf)</li><li>• Mailbox (mbx, mbox)</li><li>• Outlook Mail Message (msg)</li><li>• Outlook Personal Information Store (pst)</li></ul>	<ul style="list-style-type: none"><li>• Drawing Exchange Format (dxf)</li><li>• Email (eml)</li></ul>

1 Preferred preservation file formats are based on open standards and are the targets for file format conversions.

2 Acceptable preservation file formats are those based on open standards. Any digital records received in these formats are preserved as they are.



<b>Format Category</b>	<b>Preferred Preservation open file format<sup>1</sup> (preserved as they are)</b>	<b>Files in these formats are converted to our preferred preservation open file format</b>	<b>Acceptable open file format<sup>2</sup> (Not converted, preserved as they are)</b>
Geospatial data	Not yet decided.		
Image - raster	Portable Network Graphics (png).	<ul style="list-style-type: none"><li>• Spatial Data File (sdf)</li><li>• Bitmap (bmp, gif, pcx, pnm, ras, xbm)</li><li>• Photoshop (psd)</li><li>• Tagged Image File Format (tiff)</li><li>• Windows Cursor (cur)</li></ul>	<ul style="list-style-type: none"><li>• Geography Markup Language (gml)</li><li>• Open Document Graphics (odg)</li><li>• Joint Photographic Experts Group (jpeg)</li><li>• Portable Document Format (pdf)</li></ul>
Image - vector	Not yet implemented.	<ul style="list-style-type: none"><li>• Adobe Illustrator (ai)</li><li>• Encapsulated PostScript (eps)</li></ul>	<ul style="list-style-type: none"><li>• Scalable Vector Graphics (svg)</li></ul>
Office documents	Open Document Format (odf).	<ul style="list-style-type: none"><li>• Excel (xls, xlsx, xlt)</li><li>• PowerPoint (pot, pps, ppt, pptx)</li><li>• Rich Text Format (rtf)</li><li>• Symbolic Link (slk)</li><li>• StarOffice (sdd, sdc, sdw, sxc, sxi, sxw)</li><li>• Word (doc, docx, dot)</li><li>• Word Perfect (wpd)</li></ul>	<ul style="list-style-type: none"><li>• Open Document XML (fodt)</li><li>• OpenOffice.org XML (stw, stc, std, sti, sxg, sxm)</li></ul>
Project	XML	<ul style="list-style-type: none"><li>• Project (mpp)</li></ul>	







Format Category	Preferred Preservation open file format <sup>1</sup> (preserved as they are)	Files in these formats are converted to our preferred preservation open file format	Acceptable open file format <sup>2</sup> (Not converted, preserved as they are)
	Not yet decided.  Free Lossless Audio Codec (flac).  XHTML  Web ARChive (warc).	<ul style="list-style-type: none"><li>• Audio Video Interleave (avi)</li><li>• Advanced Systems Format (asf)</li><li>• Flash Video File (flv)</li><li>• MPEG video (mpeg-2, mpeg-4)</li><li>• QuickTime Movie (mov)</li><li>• RealMedia (rm)</li><li>• Advanced Audio Coding (aac)</li><li>• MPEG-2 audio layer 3 (mp3)</li><li>• RealMedia audio (ra, ram)</li><li>• HTML (htm, html)</li><li>• Active Server Page (asp, aspx)</li><li>• MIME HTML (mht)</li></ul>	<ul style="list-style-type: none"><li>• Ogg (ogv)</li><li>• Matroska (mkv)</li></ul>





## Preservation Formats for Electronic Records Smithsonian

In accordance with best practices, SI Archives prefers to preserve transferred electronic records in the formats described in the table below. This table lists the original/creating application by its native format(s), the corresponding primary preservation format (preferred) and the secondary format. The secondary format will be used only when a record of sufficient quality cannot be created in the primary preservation format. A glossary of acronyms used is appended to the end of this document.

Where contributing SI units have prepared their electronic records in a preservation format, SIA will preserve those records in the received format unless extenuating circumstances apply. Records originally transferred to SIA in the secondary preservation format may be later converted to the appropriate primary preservation format by SIA staff as part of the long term maintenance of the record.

This document does not address related considerations and procedures required in the conversion from original formats to preservation formats. It is essential that individuals responsible for these activities refer to forthcoming format-specific Electronic Record publications as appropriate.





<b>Original/ Creating Application</b>	<b>Primary Preservation Format</b>	<b>Secondary Preservation Format</b>	<b>TRM References</b>
<i>(various)</i> .txt	ASCII (keep original extension)		
<i>Corel WordPerfect</i> .wpd, .wpx, .doc, .rtf	PDF 4.0 or higher	RTF (text)	Pg 4-202 (PDF) Pg 4-93 (RTF)
<i>Microsoft Word</i> .doc, .dot, .rtf	PDF 4.0 or higher	RTF (text)	Pg 4-202 (PDF) Pg 4-93 (RTF)
<i>Lotus 1-2-3</i>	<i>Same as original</i>	CSV	No reference
<i>Microsoft Excel</i> *.xls	PDF	CSV	Pg 4-202 (PDF)



<b>Original/ Creating Application</b>	<b>Primary Preservation Format</b>	<b>Secondary Preservation Format</b>	<b>TRM References</b>
<i>Microsoft Powerpoint</i> *.ppt	PDF		Pg 4-202 (PDF)
<b>Graphics Applications</b>			
<i>(various graphics applications)</i> *.bmp, *.tif, *.jpg, *.gif	TIFF (24-bit RGB) (minimum 3000 pixels along long dimension or 400 dpi, whichever is greater)	JPEG (uncompressed) (for JPG and GIF formats only.)	Pg 4-98 (TIFF) Pg 4-101 (JPEG)
<i>Adobe Photoshop, Illustrator</i> .psd, .eps	TIFF (24-bit RGB) (minimum 3000 pixels along long dimension or 400 dpi, whichever is greater)		Pg 4-98 (TIFF)
<i>Macromedia Fireworks/ Freehand</i> .png	TIFF (24-bit RGB) (minimum 3000 pixels along long dimension or 400 dpi, whichever is greater)	TIFF (8-bit, grayscale or b/w for black and white images only) (minimum 3000 pixels along long dimension or 400 dpi, whichever is greater)	Pg 4-98 (TIFF)





<b>Original/ Creating Application</b>	<b>Primary Preservation Format</b>	<b>Secondary Preservation Format</b>	<b>TRM References</b>
<i>Microsoft Image Composer .mic</i>	TIFF (24-bit RGB) (minimum 3000 pixels along long dimension or 400 dpi, whichever is greater)	TIFF (8-bit, grayscale or b/w for black and white images only) (minimum 3000 pixels along long dimension or 400 dpi, whichever is greater)	Pg 4-98 (TIFF)
<b><i>Database Management Systems (DBMS)</i></b>			
<i>Dbase III+ .dbf, .dbt</i>	<i>Same as original</i>	CSV	CSV not listed Dbase not listed
<i>Lotus Notes (database)</i>	<i>Same as original</i>	CSV	CSV not listed Lotus Notes not listed
<i>Microsoft Access (database) *.mdb</i>	<i>Same as original</i>	CSV	Pg 4-42, 4-74, 10-7
<i>Microsoft SQL Server *.bak</i>	<i>Same as original, full backup made with MS SQL</i>	CSV	Pg 10-10 (MS SQL Server)
<i>Sybase SQL Server, Adaptive Enterprise Server, SQL Anywhere</i>	<i>Same as original, full backup made with Sybase dbms. Also requires full backup of Sybase master database.</i>	CSV	Pg 4-74 (Sybase)





<b>Original/ Creating Application</b>	<b>Primary Preservation Format</b>	<b>Secondary Preservation Format</b>	<b>TRM References</b>
<b><i>WebPage Developers and HTML Editors</i></b>			
<i>Adobe PageMill, Microsoft Frontpage, Macromedia Dreamweaver (other various)</i> .htm, .html, .shtml	XHTML 1.0	PDF (Isolated, single webpage only)	Pg 4-86 (XHTML) Pg 4-89
<i>Active Server Page(various)</i> .asp	XHTML 1.0	ASP (depending on parameter complexity)	Pg 4-86 (XHTML) Pg 4-25 (ASP)
<i>Macromedia Cold Fusion</i> .cfm	XHTML 1.0		Pg 4-86 (XHTML) Pg 4-33 (CFM)
<i>Macromedia Flash/Shockwave</i> .swf	SWF (No change)	<i>Analyze carefully for embedded URLs or other file dependencies.</i>	Pg 4-27, 4-30, 4-105, 10-15, 10-23
<i>Real Media</i> .rm, .ram	<i>Same as original</i>		Pg 4-27, 10-6
<b><i>Architectural Design/Engineering Graphics</i></b>			
<i>AutoCAD</i> *.cad	PDF (v6.0 with layer retention)		Pg 4-202 (PDF) Pg 10-15 (CAD)
<i>Microsoft Visio</i> *.vsd	PDF (v6.0 with layer retention)		Pg 4-202 (PDF) Pg 10-16 (Visio)





### Recommended File Formats

#### Recommended File Formats for Preservation

The UT Digital Repository accepts work in any digital format; however, submission in a file format listed below is strongly encouraged. The formats listed below are considered relatively stable and therefore facilitate long-term preservation efforts. These formats exhibit all or many of the following characteristics: open documentation; support across a range of software platforms; wide adoption; no compression (or lossless compression); no embedded files or embedded programs/scripts; and non-proprietary format. For additional information on the UT Digital Repository's preservation strategy, please see the [Preservation Policy](#).

#### Textual Formats

Acrobat PDF/A  
Comma-Separated Values  
Open Office Formats  
Plain Text (US-ASCII, UTF-8)  
XML

#### File Extensions

.pdf  
.csv  
.odt, .ods, .odp  
.txt  
.xml

#### Image/Graphic Formats

JPEG  
JPEG2000  
PNG  
SVG 1.1 (no Java binding)  
TIFF

#### File Extensions

.jpg  
.jp2  
.png  
.svg  
.tif, .tiff

#### Audio Formats

AIFF  
WAVE

#### File Extensions

.aif, .aiff  
.wav

[https://repositories.lib.utexas.edu/recommended\\_file\\_formats](https://repositories.lib.utexas.edu/recommended_file_formats)



# Algunas recomendaciones en digitalización



Como guía en procesos de digitalización pueden usarse:

"Technical Guidelines for Digitizing Cultural Heritage Materials" generado en 2010 por la Federal Agencies Digitization Guidelines Initiative (FADGI).

"Directrices para proyectos de digitalización de colecciones y fondos de dominio público" , IFLA (2002).

"Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files Raster Images" , NARA (2004).

"Recomendaciones para la digitalización de los documentos en archivos". Junta de Castilla y León (2011).



# Algunas recomendaciones en digitalización



De cada obra digitalizada es recomendable tener al menos 3 versiones: a) imagen maestra resultante del escaneo, b) documento maestro con imagen derivada de la anterior y el OCR generado sobre ella y c) otra(s) derivada(s) a los fines de uso y distribución en la web.



# Enlaces de interés

<http://www.digitizationguidelines.gov/>

Directrices IFLA para digitalización

Recomendaciones para la digitalización de archivos. Junta de Castilla y León



# Formatos. ¿Cómo conocer lo que tiene un RI?



- 1. Perfilamiento automatizado de los objetos del repositorio:** esto involucra al objeto de contenido (CDO) con sus propiedades significativas y a la información de representación de ese objeto (RI). Realizar el perfil con DROID que contrasta con el registro PRONOM y brinda un reporte.

El punto 1 es una de las 3 partes que se consideran importantes a la luz de cumplir con la ISO 14721 y realizar una evaluación del repositorio en los aspectos de preservación y accesibilidad



# ¿Qué acciones se proponen?

Nombre	Descripción	Formato	Ver	Orden
<b>Bloque: TEXT</b>				
<input type="checkbox"/> Tesina de Licen ... mazan Maria Belen.pdf.txt	Extracted text	Text	[Ver]	1 (Anterior:1)
<input type="checkbox"/> presentación.xps).pdf.txt	Extracted text	Text	[Ver]	2 (Anterior:2)
<b>Bloque: ORIGINAL</b>				
<input type="checkbox"/> Tesina de Licenciatura - Almazan Maria Belen.pdf (principal)	Documento completo	Adobe PDF	[Ver]	1 (Anterior:1)
<input type="checkbox"/> ...	Presentación	Adobe	[Ver]	2



Información descriptiva (DI)

De Giusti, Mrarisa R. (2014). Tesis doctoral: “UNA METODOLOGÍA DE EVALUACIÓN DE REPOSITARIOS DIGITALES PARA ASEGURAR LA PRESERVACIÓN EN EL TIEMPO Y EL ACCESO A LOS CONTENIDOS”. Disponible en: <http://hdl.handle.net/10915/43157>



# Resumen PD en RIs

Regulación de todos los procedimientos.

Regulación de los derechos de preservación digital sobre los documentos.

Regulación de los formatos admisibles.

Control de formatos en la ingestión.

Formatos de visualización y de preservación

Almacenaje de metadatos técnicos.

Copias sistemáticas externas.

Creación de procedimientos de contingencia ante desastres.

Auditoría interna/externa de seguridad.

Plan de preservación...

