



# Objetivos

- 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 estándares e implementaciones para cumplir con el objetivo de preservación.
- La digitalización sus dificultades y la generación de nuevos materiales para preservar.



# 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

En los documentos en papel se habla de “negligencia benigna”: el olvido de un manuscrito en un arcón, puede que lo preserve. En los digitales, la negligencia benigna no sirve: un disco olvidado 5 años... no sirve.



- 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:

- ❖ Los datos se mantendrán en el repositorio sin sufrir daños, sin perderse o sin ser alterados de forma malintencionada/o no.
- ❖ Los datos podrán ser localizados y entregados al usuario.
- ❖ Los datos podrán ser interpretados y comprendidos por el usuario.
- ❖ 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



#### 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, robos, problemas de hardware (rotura de discos, etc.) y 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. “Obsolescencia digital”

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.



# Preservación de contenido. Estrategias

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

- Migración
- 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

# Preservación de contenido

- Los riesgos de pérdida de datos por eventos desafortunados siempre son posibles.
- Para disminuir esos riesgos es necesario contar con un sistema de backups (datos, configuración, documentación, etc) como se hablara en clases precedentes.
- También es necesario elegir los formatos de acuerdo a los criterios que se mencionaron: uso de una gran comunidad, apertura, licenciamiento libre...

# 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.



# 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 (contenedores) 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.

## 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/A posee dos niveles de cumplimiento:

**PDF/a** aplica corrección semántica y estructura. Cada carácter debe tener su equivalente Unicode. La estructura se expresa por medio de etiquetas.

**PDF/b** aplica integridad visual.



# National Archives of Australia



Australian Government



NATIONAL ARCHIVES OF AUSTRALIA

Explore:

Search the collection	Go >
Search the website	Go >

Your story, our history

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

Format Category	Preferred Preservation open file format <sup>1</sup> Files in these formats are converted to our preferred preservation open file format (preserved as they are)	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.</p> <p>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.



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)
Geospatial data	Not yet decided.	<ul style="list-style-type: none"> <li>• Spatial Data File (sdf)</li> </ul>	<ul style="list-style-type: none"> <li>• Geography Markup Language (gml)</li> </ul>
Image - raster	Portable Network Graphics (png).	<ul style="list-style-type: none"> <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>• Open Document Graphics (odg)</li> <li>• Joint Photographic Experts Group (jpeg)</li> </ul>
Image - vector	Not yet implemented.	<ul style="list-style-type: none"> <li>• Adobe Illustrator (ai)</li> </ul>	<ul style="list-style-type: none"> <li>• Portable Document Format (pdf)</li> <li>• Scalable Vector Graphics (svg)</li> </ul>
Office documents	Open Document Format (odf).	<ul style="list-style-type: none"> <li>• Encapsulated PostScript (eps)</li> <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> Files in these formats are converted to our preferred preservation open file format (preserved as they are)	Acceptable open file format <sup>2</sup> (Not converted, preserved as they are)
	Plain text in Unicode or ASCII.	<ul style="list-style-type: none"> <li>•</li> <li>•</li> <li>•</li> </ul>
	<p>Not yet decided.</p> <p>Video files are currently preserved as is.</p>	<ul style="list-style-type: none"> <li>•</li> <li>•</li> <li>•</li> <li>•</li> </ul>

- Flash Movie file (swf)
- Motion JPEG/JPEG 2000
- RealVideo (rv, rmvb)
- Windows Media Video (wmv)



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>Not yet decided.</p> <p>Free Lossless Audio Codec (flac).</p> <p>XHTML</p>		<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> </ul>	<ul style="list-style-type: none"> <li>• Ogg (ogv)</li> <li>• Matroska (mkv)</li> </ul>
Web ARChive (warc).	• MIME HTML (mht)		•



## 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</i> .mic	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+</i> .dbf, .dbt	<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)</i> *.mdb	<i>Same as original</i>	CSV	Pg 4-42, 4-74, 10-7
<i>Microsoft SQL Server</i> *.bak	<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)



Original/ Creating Application	Primary Preservation Format	Secondary Preservation Format	TRM References
<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)



# University of Texas



Digital Repository

## 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

#### Image/Graphic Formats

JPEG  
JPEG2000  
PNG  
SVG 1.1 (no Java binding)  
TIFF

#### Audio Formats

AIFF  
WAVE

#### File Extensions

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

#### File Extensions

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

#### 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?

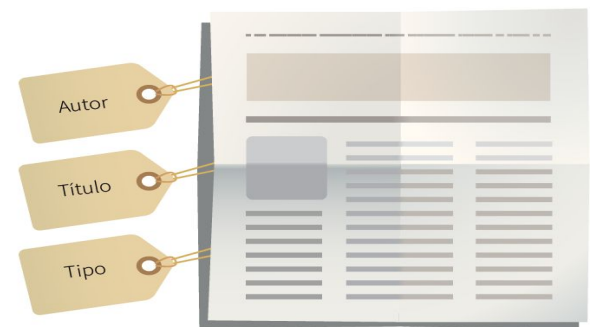
**Perfilamiento automatizado de los objetos del repositorio:** esto involucra al objeto de contenido (CDO) con sus propiedades significativas y a la información de representación de ese objeto (RI). Realizar el perfil con DROID que contrasta con el registro PRONOM y brinda un reporte.

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>				
 Tesina de Licen ... mazan Maria Belen.pdf.txt	Extracted text	Text	[Ver]	1 (Anterior:1)
 presentación.xps).pdf.txt	Extracted text	Text	[Ver]	2 (Anterior:2)
<b>Bloque: ORIGINAL</b>				
 Tesina de Licenciatura - Almazan Maria Belen.pdf (principal)	Documento completo	Adobe PDF	[Ver]	1 (Anterior:1)
	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...



# Bibliotecas y repositorios digitales

## Capítulo 8:

La preservación en el repositorio institucional: el Modelo OAIS ISO 14721. Comparación con las facilidades que ofrece DSPACE.



# Estándares

El estándar 14721 (OAIS), los metadatos PREMIS y las directrices para la preservación, 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.



# El Modelo OAIS

Modelo de Referencia  
para un Sistema Abierto de  
Archivo de Información.

ISO 14721: 2012

ISO Reference Model  
of an Open Archival  
Information System (OAIS).



# El Modelo OAIS

- Archivo que comprende una organización de personas y sistemas que han asumido el compromiso de preservar a largo plazo y hacer disponible un determinado corpus de información (cualquier tipo de conocimiento a intercambiar) para una comunidad designada.
- Se refiere a la información analógica y a la digital, pero el foco está en esta última.
- Open (abierto): se usa para indicar que esta recomendación ha sido realizada en foros abiertos. No significa que el archivo es de acceso gratuito o irrestricto. Puede ser cualquiera.



# El modelo de Referencia OAIS

1. Introducción: propósitos, alcance, campo de aplicación, razones, conformidad, estándares relacionados y definiciones.
2. Conceptos: Medioambiente, información e interacciones externas de alto nivel.
3. Responsabilidades: obligatorias y deslindes.
4. Modelo: funcional, de información, transformaciones.
5. Preservación: de la información y del acceso a la información.
6. Interoperabilidad.



# Sección 1



# Justificación del Modelo de referencia

- Ninguna discusión sobre la conservación de repositorios y flujos de trabajo estaría completa sin al menos una breve introducción al modelo de referencia OAIS.
- Una introducción a este modelo sirve para mostrar cómo implementa muchos de los procesos de flujos de trabajo y cómo se relaciona con la conservación digital.
- Se recomienda como la mejor práctica actual.



# Antecedentes

- El Comité Consultivo para los Sistemas de Datos Espaciales (CCSDS, por sus siglas en inglés), un foro para agencias nacionales espaciales interesadas en desarrollar acuerdos de cooperación sobre normas de gestión de datos en la investigación espacial, llevó a cabo el desarrollo inicial de esta norma para permitir el almacenamiento de datos digitales a largo plazo, generados a partir de las misiones espaciales.
- En colaboración con la Organización Internacional para la Normalización ISO, el modelo de referencia fue aprobado como norma ISO en 2002 (ISO-14721).



# Funciones del Modelo de referencia

- Las dos funciones principales del modelo son **conservar** la información y **garantizar el acceso** a la misma.
- El modelo funcional OAIS, que se propone lograr estos objetivos amplios, en cierta medida, define la arquitectura aproximada de cualquier tipo de sistema de software diseñado para cumplir con esta norma y con todo tipo de flujos de trabajo asociados con el repositorio.

# Propósito y campo de Aplicación

- Es aplicable para cualquier archivo, pero especialmente está enfocada en organizaciones con responsabilidad de hacer que la información esté disponible a largo plazo para una **comunidad designada**.
- Es de interés para aquellos que crean información que puede necesitar preservación a largo plazo.
- No especifica un diseño o una implementación. Cada implementación dará lugar a una funcionalidad distinta.
- El foco primario es la información inherentemente digital.
- El modelo se acomoda para información que no es inherentemente digital pero el modelo y la preservación de esa información no está descripto en detalle.



# Propósito y campo de Aplicación

- Estandariza las relaciones y los componentes de un sistema de archivos. Es un framework que sirve para entender mejor de qué se habla.
- Establece un vocabulario común.
- Ofrece un marco consensuado internacional para la definición de entidades, procesos y funciones de los archivos de datos.
- Facilita comprender y aplicar conceptos necesarios para la preservación de información digital a largo plazo.

# Sección 2



# Conceptos en OAIS

El propósito de esta sección es motivar y describir varios conceptos clave, de alto nivel del OAIS. Un punto de vista más completo y una modelización formal de estos conceptos, se da en la sección 4.

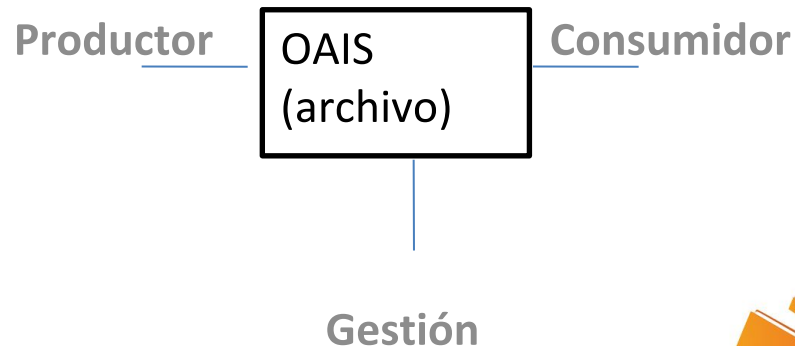


# Medioambiente OAIS

- Un productor que provee la información.
- Una política global de gestión (management), NO las operaciones diarias.
- Un consumidor que busca, encuentra y adquiere la información de su interés que ha sido preservada.
- La comunidad designada es el conjunto de los consumidores que son capaces de comprender la información preservada.

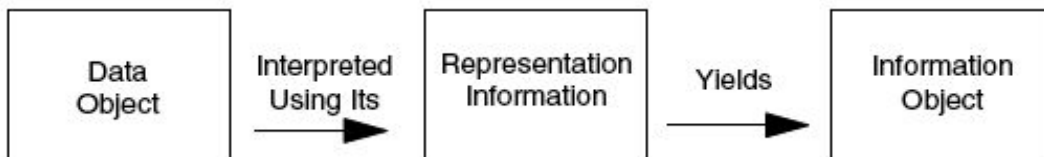
## Actores en el modelo

- **Producer-Consumer-Management**



# Conceptos en OAIS

Una definición clara de información es central para la capacidad del OAIS para preservar esa información. Una persona o un sistema, tienen una base común de conocimientos (KB) que le permite comprender la información. Se considera información en este campo a cualquier tipo de conocimiento que puede intercambiarse y que se expresa a través de algún tipo de datos: la información en un artículo periodístico, se expresa por caracteres (datos), los cuales bajo el paraguas de un lenguaje (KB), se convierten en información relevante. Si el receptor desconoce la lengua, entonces el artículo tendrá que ser acompañado por información extra, por ejemplo, un diccionario o una gramática.



## Conceptos en OAIS

- A fin de que este objeto de información se preserve con éxito, es fundamental para un OAIS identificar con claridad y comprender los objetos de datos y la representación de la información asociada.
  - Para la información digital, esto significa que el OAIS debe identificar claramente los bits y la representación de la información que se aplica a los bits.
- El OAIS debe entender la base de conocimientos de su comunidad determinada/designada para comprender la representación de la información mínima que debe mantenerse.

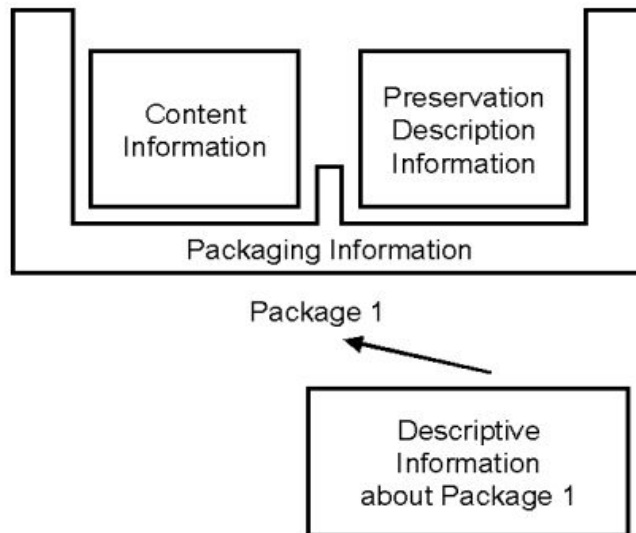


## Conceptos en OAIS

- La unidad de intercambio entre un OAIS y su medioambiente es el paquete de información –IP.
- Un IP contiene 2 tipos de información:
  - De contenido
  - De descripción de preservación (PDI)
- La información de contenido y la PDI pueden verse como encapsuladas e identificables por medio de la información de empaquetado.
- El paquete resultante es recuperable en virtud de la información descriptiva: DI.



# Conceptos en OAIS



**ISO 14721: Fig 2-3: Paquete de información: conceptos y relaciones**

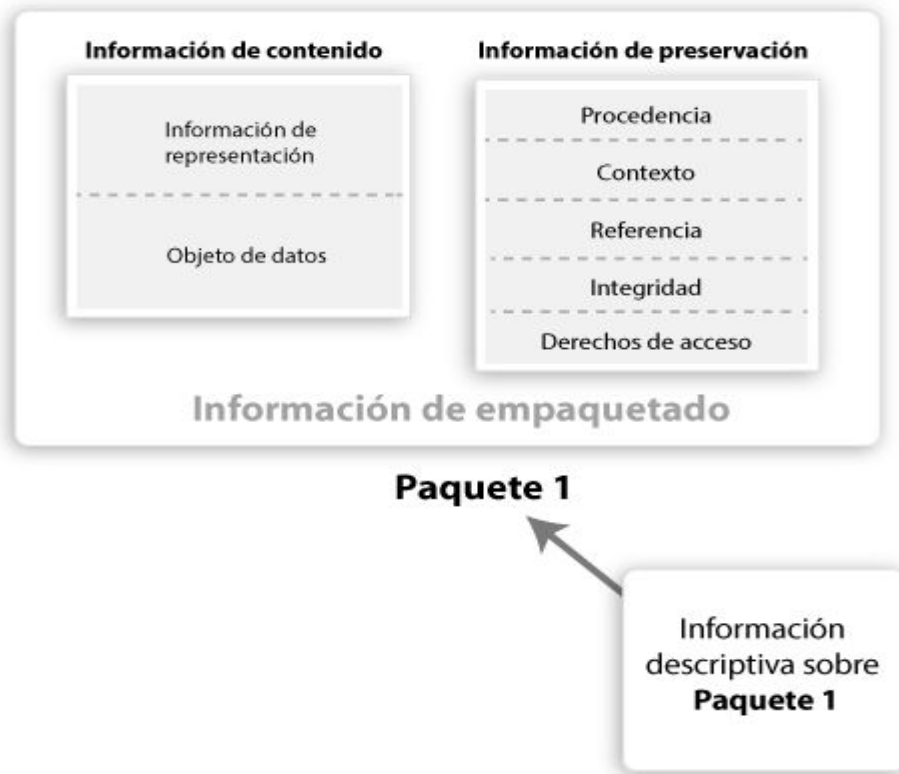


## Conceptos en OAIS

- La información de empaquetado es la información que, ya sea real o lógicamente, une, identifica y relaciona la información del contenido y la PDI.
- La información descriptiva es la información que se utiliza para descubrir qué paquete tiene la información de contenido de interés.



# Estructura del Paquete de Información



## El paquete de información -IP

La norma define el IP como un contenedor conceptual con dos tipos de información: de contenido y de preservación. La *información de contenido (CI)* es el objeto mismo que se desea mantener en el tiempo y la *información descriptiva de preservación (PDI)*, debe brindar datos suficientes sobre la **procedencia**, el **contexto**, la **referencia**, la **integridad** y los **derechos de acceso**.



## Elementos de la PDI

La **procedencia**, más allá de describir la fuente, incluye los procesos que se han realizado sobre la información: la historia del objeto, cambios, versiones y responsables. El **contexto** muestra las relaciones con otras fuentes de información o contenidos. La **referencia** provee una identificación única del contenido. La **integridad (o fijeza)** provee una protección para que la información no sea alterada de manera intencional /no. Los **derechos de acceso** proveen información sobre los términos de acceso incluyendo preservación, distribución y uso de la información de contenido.

# Conceptos en OAIS

- Variantes del paquete de información:
  - Submission Information Package (SIP)
  - Archival Information Package (AIP)
  - Dissemination Information Package (DIP)
- Los paquetes de información variarán dependiendo de su rol:
  - Por ejemplo master file y versiones derivadas (thumbnails, JPEG, PDFs...).



## Clases de IPs según su función

Submission Information Package (**SIP**): es el paquete que proviene del productor y se va a incorporar al OAIS. Suele contener menos información que el AIP.



## Clases de IPs según su función

Archival Information Package (AIP): contiene, como mínimo, suficiente información de un objeto como para garantizar la preservación a largo plazo. Busca mantener la mayor calidad posible de información descriptiva de preservación y de representación de los objetos representados o contenidos.



## Clases de IPs según su función

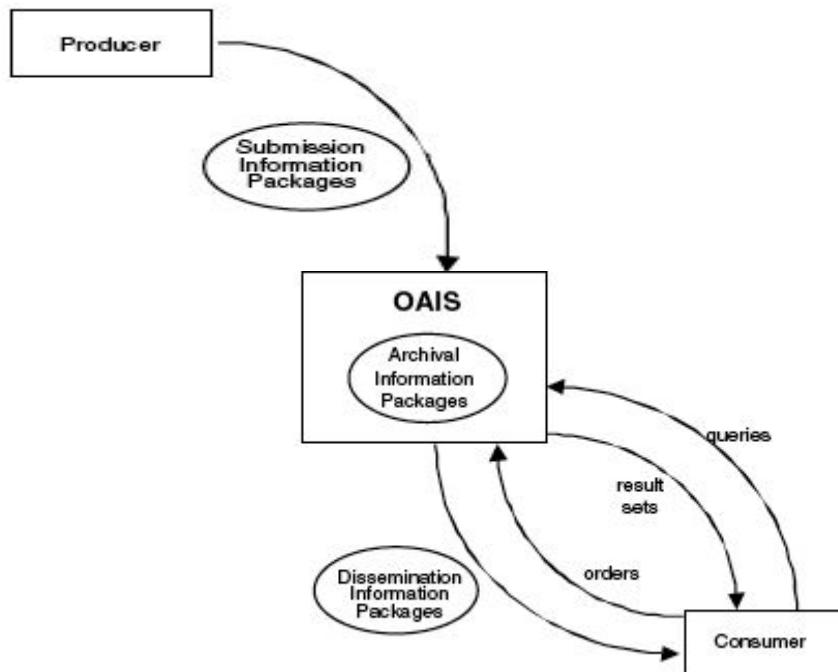
Dissemination Information Package (DIP): es el paquete que se entrega a un consumidor en respuesta a una solicitud. La información de empaquetado toma muchas formas dado que los usos de OAIS son diversos, puede ser tan completo como los AIP a partir de los cuales se construye o ser sólo una breve descripción del paquete.



# OAIS interacciones externas de alto nivel

La figura que sigue es un diagrama de flujo de datos que representa los flujos de información entre productores, consumidores y el OAIS y no incluye flujos que involucren al management.

# OAIS interacciones externas



# Visión de alto nivel de las interacciones en un entorno OAIS

- Interacción de la gestión
  - financiación, utilización de recursos, pagos, resolución de conflictos.
- Interacción del productor
  - los acuerdos de ingesta. Acuerdo por los SIPs que va a mandar, tiempo (acuerdo por data submission session)
- Interacción de los consumidores
  - Ayudas, descubrimiento de información, ordenamiento de la información. (Data dissemination session).



# Sección 3: Responsabilidades del OAIIS



# OAIS Responsabilidades obligatorias

- Negociar y aceptar la información adecuada de los productores de información.
- Obtener un control de la información proporcionada al nivel necesario para asegurar la conservación.
- Determinar, por sí o con otras partes, cuáles comunidades deben convertirse en la comunidad designada y, entender la información proporcionada.
- Asegurar que la información que se conserva es comprensible independientemente de la comunidad designada.
  - la comunidad debe comprender la información sin la ayuda de los expertos que han producido la información.



# OAIS Responsabilidades obligatorias

- Cumplir con las políticas y procedimientos documentados para:
  - garantizar que la información se conserva en contra de todas las contingencias razonables, y
  - permitir que la información sea difundida como copia autenticada de la original, o como **trazabilidad** a la original.
- Lograr que la información preservada esté disponible para la comunidad designada.
  - Secciones 3.1 y 3.2 de la norma.



# Mecanismos de deslinde de responsabilidades

1. Las negociaciones para y la información que acepta.
2. El control eficiente de la preservación.
  - Cuestiones de copyright, propiedad intelectual y restricciones legales para el uso.
  - Autoridad para modificar el modo de representación de la información.
  - Acuerdos con organizaciones externas.
3. Determinación de la comunidad designada de consumidores.



# Sección 4

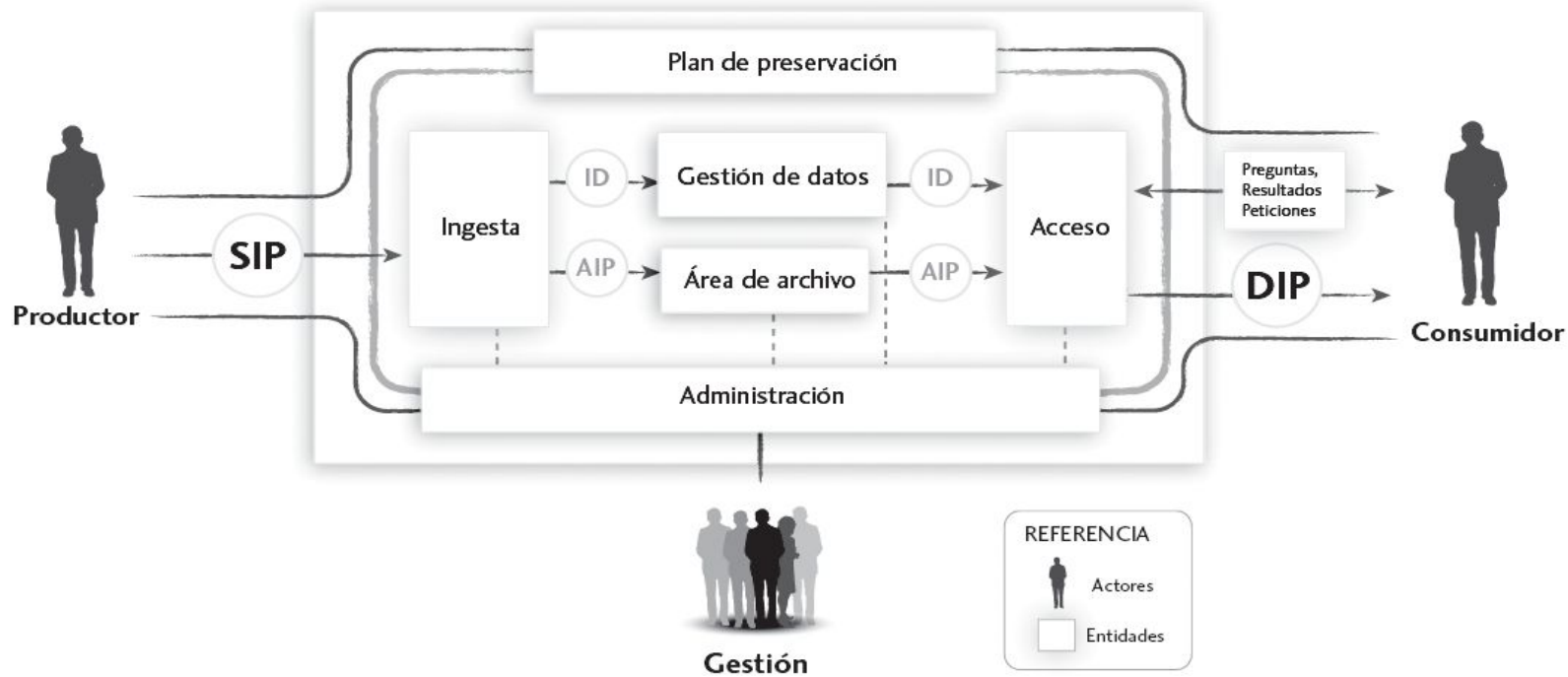


# OAIS

## Modelo Funcional

### Sección 4.1





# OAIS Modelo funcional

Seis entidades funcionales e interfaces relacionadas:

- Ingesta- Ingest
- Almacenamiento de archivos-Archival storage
- Gestión de datos-Data management
- Administración-Administration
- Planeamiento de la preservación-Preservation Planning
- Acceso- Access



## Modelo OAIS

El proceso puede iniciarse cuando el productor suministra el recurso (paquete de entrada) llamado SIP a través del *ingest*, que luego se convierte en AIP terminando en la entidad *archival storage*. El flujo puede continuar cuando el consumidor busca una información en el sistema, que es entregada como un DIP a través de la entidad *access*, ya que la información está preservada en el sistema previamente.

# Modelo OAIS

Los datos relacionados con los documentos y el repositorio mismo se mantienen organizados a través de la entidad *data management*. Luego hay una entidad *administration* dedicada a la administración adjunta a la gestión (administradores y responsable del repositorio) y esta entidad se relaciona con las secciones de ingesta, *gestión de datos*, *almacenamiento de archivos* y *planificación de la preservación*. Esto permite una gestión estructural y ayuda a mantener los AIP a lo largo del tiempo.



# Modelo OAIS

El módulo de *planificación de la preservación* desarrolla estrategias y normas de conservación, monitorea las últimas novedades y avances en el campo, y monitorea los cambios en la comunidad designada, para que toda la información nueva que se solicite, se pueda adjuntar a los AIP correspondientes.

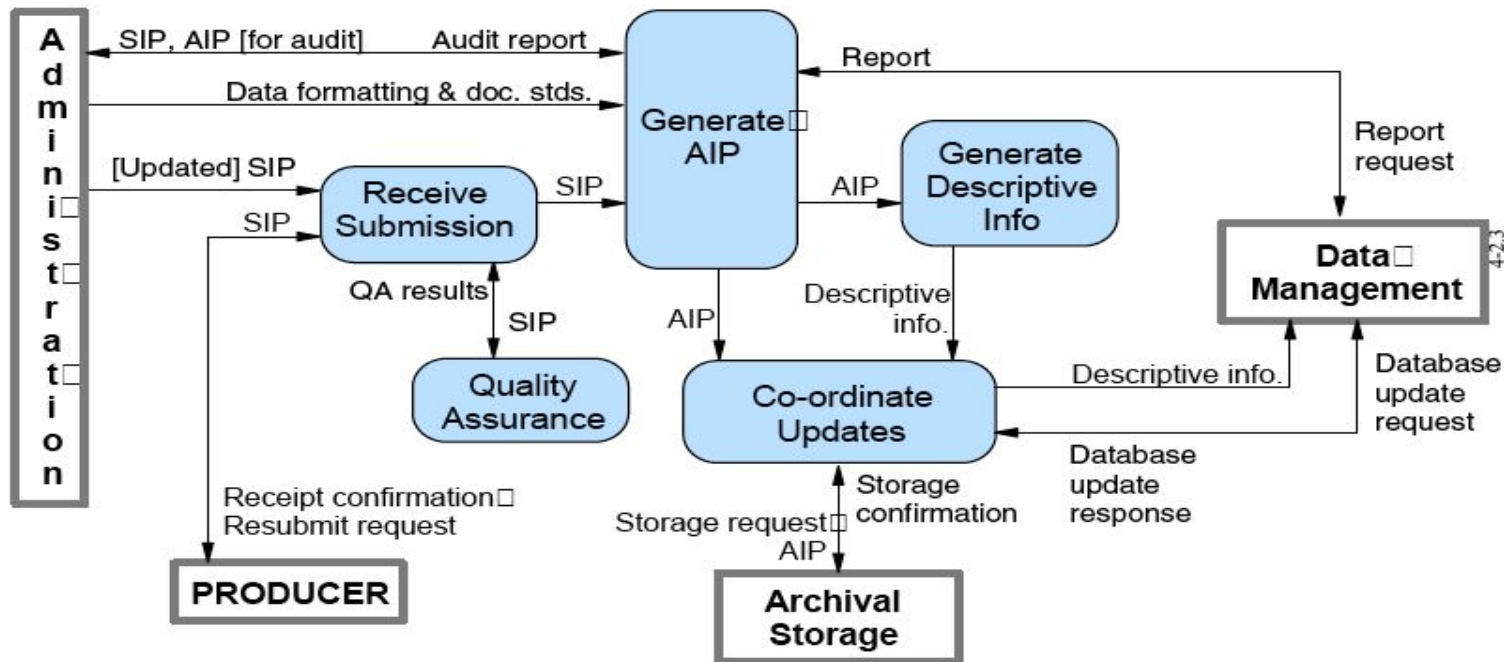


# Ingesta/Ingest/presentación

Provee los servicios y funciones para aceptar el paquete de información presentado (SIP) por parte de los Productores (o a partir de elementos internos bajo control de la administración) y preparar los contenidos para almacenaje y gestión dentro del archivo.



# Functions of Ingest

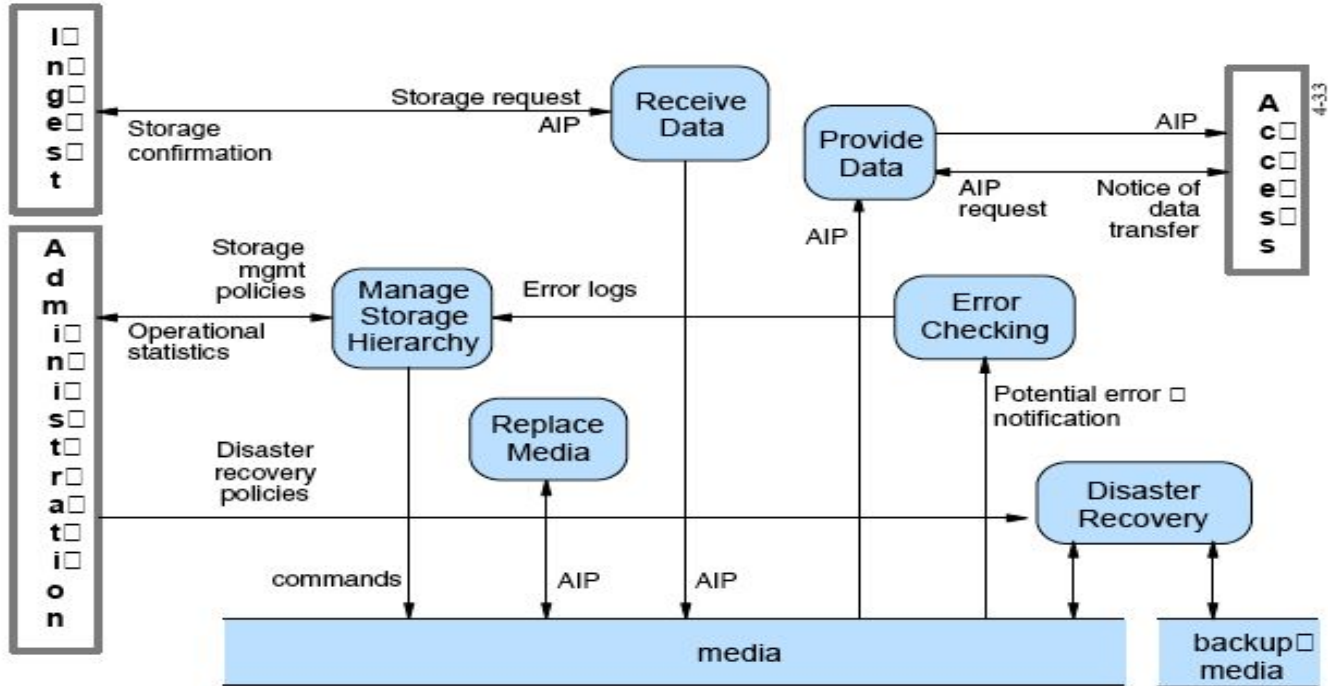


# Entidad OAIS Ingest

- **Descripción:** Provee los servicios y funciones para aceptar un SIP por parte de los Productores o bajo el control de la Administración.
- Prepara los contenidos para almacenamiento y gestión dentro del archivo.
- Realiza el aseguramiento de calidad/validación de los SIPs.
- Genera el AIP que cumple con los estándares de formato de datos y documentos.
- Extrae la información descriptiva y la envía al *data management*.
- Coordina las actualizaciones en el *archival storage* y en el *data management* de la base de datos.



# Functions of Archival Storage

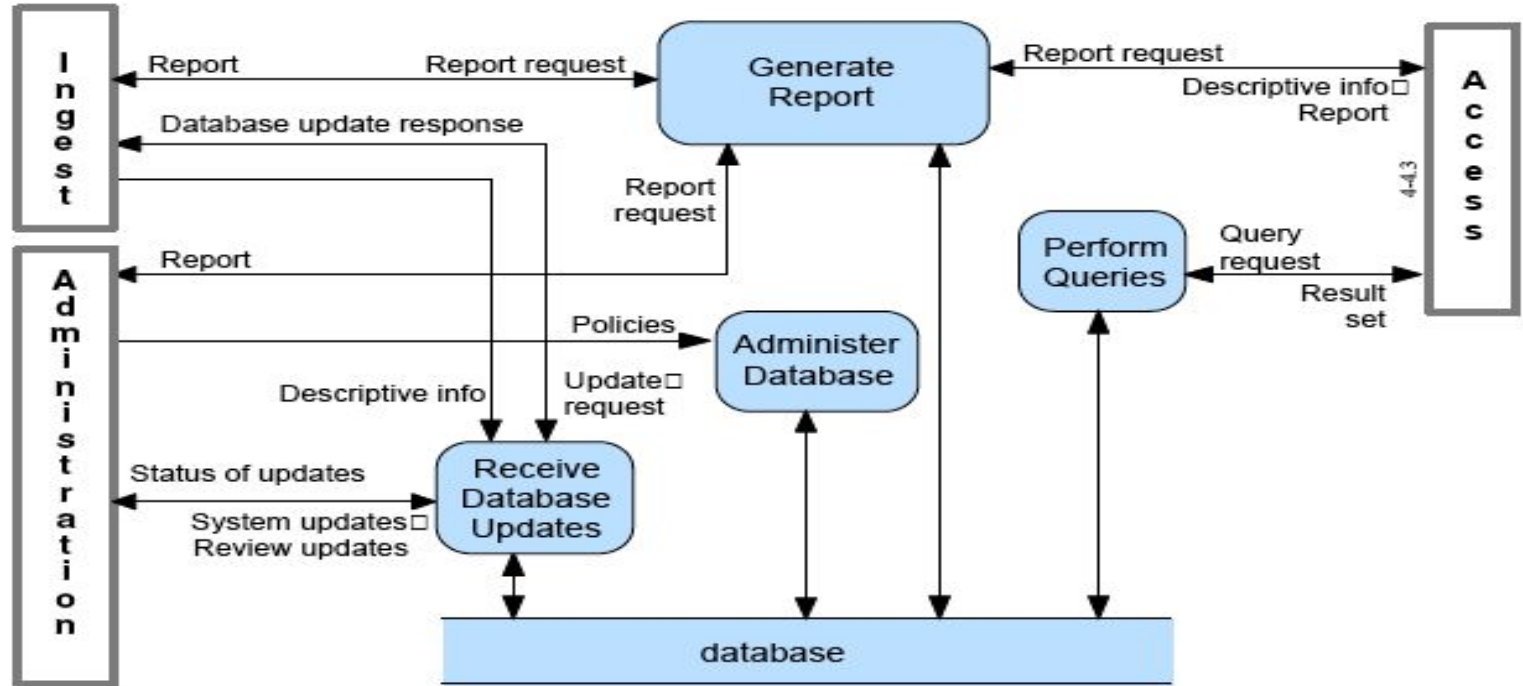


# Entidad OAIS Archival Storage

- **Descripción:** Provee los servicios y funciones para el almacenamiento, mantenimiento y recuperación de los AIPs.
- Recibe el AIP de la entidad ingest y lo almacena. Gestiona las jerarquías de almacenamiento. Configura niveles especiales de servicio, seguridad y protección (por ejemplo backups). Provee estadísticas de inventario, capacidad disponible, etc. Transforma los datos que constituyen la información de empaquetado para reproducir el AIP en el tiempo.
- Realiza una verificación de errores. Provee un mecanismo estándar para el seguimiento y verificación de la validez de los datos. Provee un mecanismo de duplicación de los contenidos en una lugar físico separado. Provee copia de los AIPs almacenados a la entidad *access*.



# Functions of Data Management

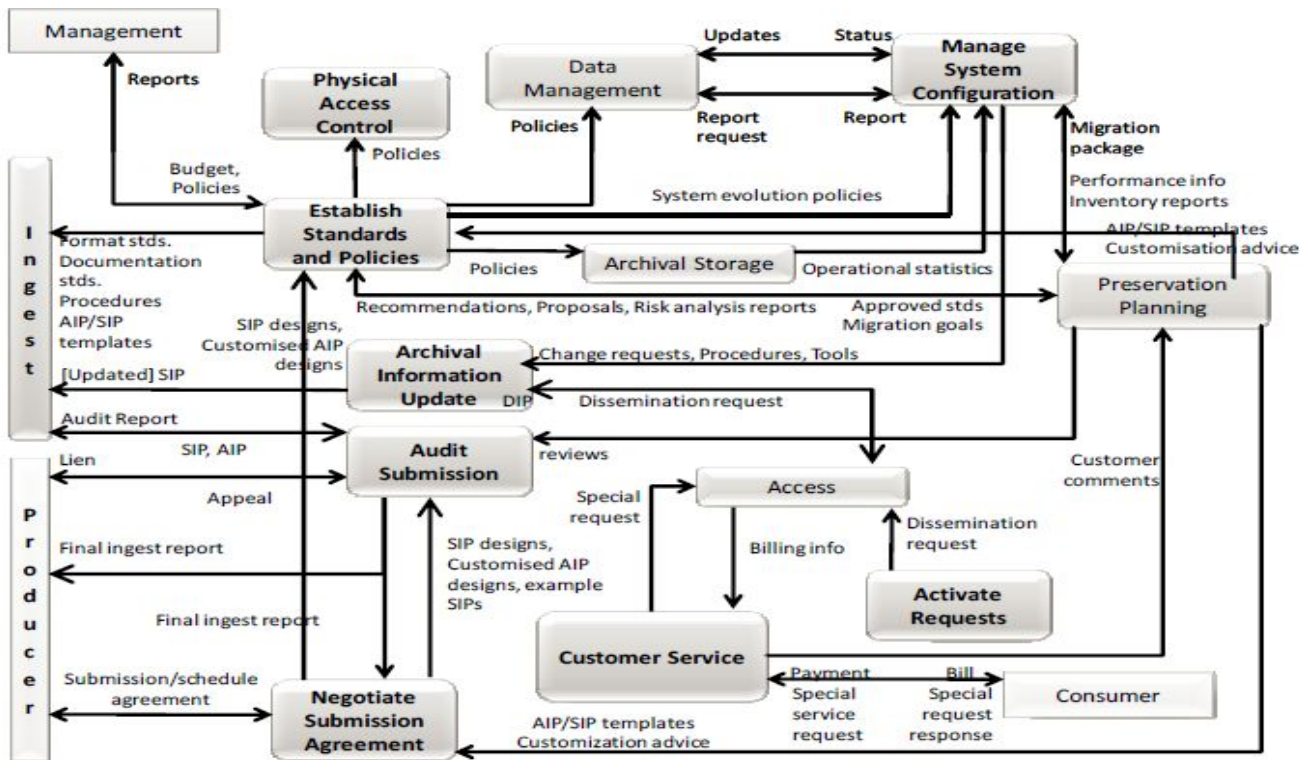


# Entidad OAIS Data Management

- **Descripción:** Provee los servicios y funciones para poblar, mantener y acceder a la información descriptiva que identifica y documenta el contenido del Archivo, y a los datos administrativos usados para gestionarlo.
- Es responsable de la administración de la base de datos.
- Recibe solicitudes de la entidad *access* y genera un conjunto de resultados.
- Recibe pedidos de las entidades *ingest*, *access* y *administration* y genera reportes.
- También recibe actualizaciones de *ingest* y *administration*.



# Functions of Administration



# Entidad OAIS Administration

**Descripción:** Provee los servicios y funciones para la operación global del sistema de archivos.

Solicita la información necesaria sobre los archivos y negocia los acuerdos con los Productores.

Monitorea la funcionalidad del sistema de archivos, controla los cambios de la configuración y mantiene su integridad y trazabilidad. Audita las operaciones del sistema, performance y uso. Envía reportes al *data management* y recibe reportes de esa entidad. Sumariza todos los reportes y provee información sobre performance del OAIS e inventario y envía esta info a *preservation planning* para establecer políticas y estándares. Recibe los paquetes de migración para *preservation planning*.



# Entidad OAIS administration

Recibe los pedidos de cambio, procedimientos y herramientas para la actualización del archivo.

Responsable de enviar un pedido de disseminación a *access*, actualizando los contenidos de los DIP y resuministrando los SIP a *ingest*.

Provee mecanismos para restringir/permitir acceso a los elementos del archivo.

Es responsable de enviar información para establecer estándares y políticas. Desarrolla políticas de gestión de archivo por jerarquías, incluyendo políticas de migración. Es responsable de la recuperación ante desastres.



# Entidad OAIS administration

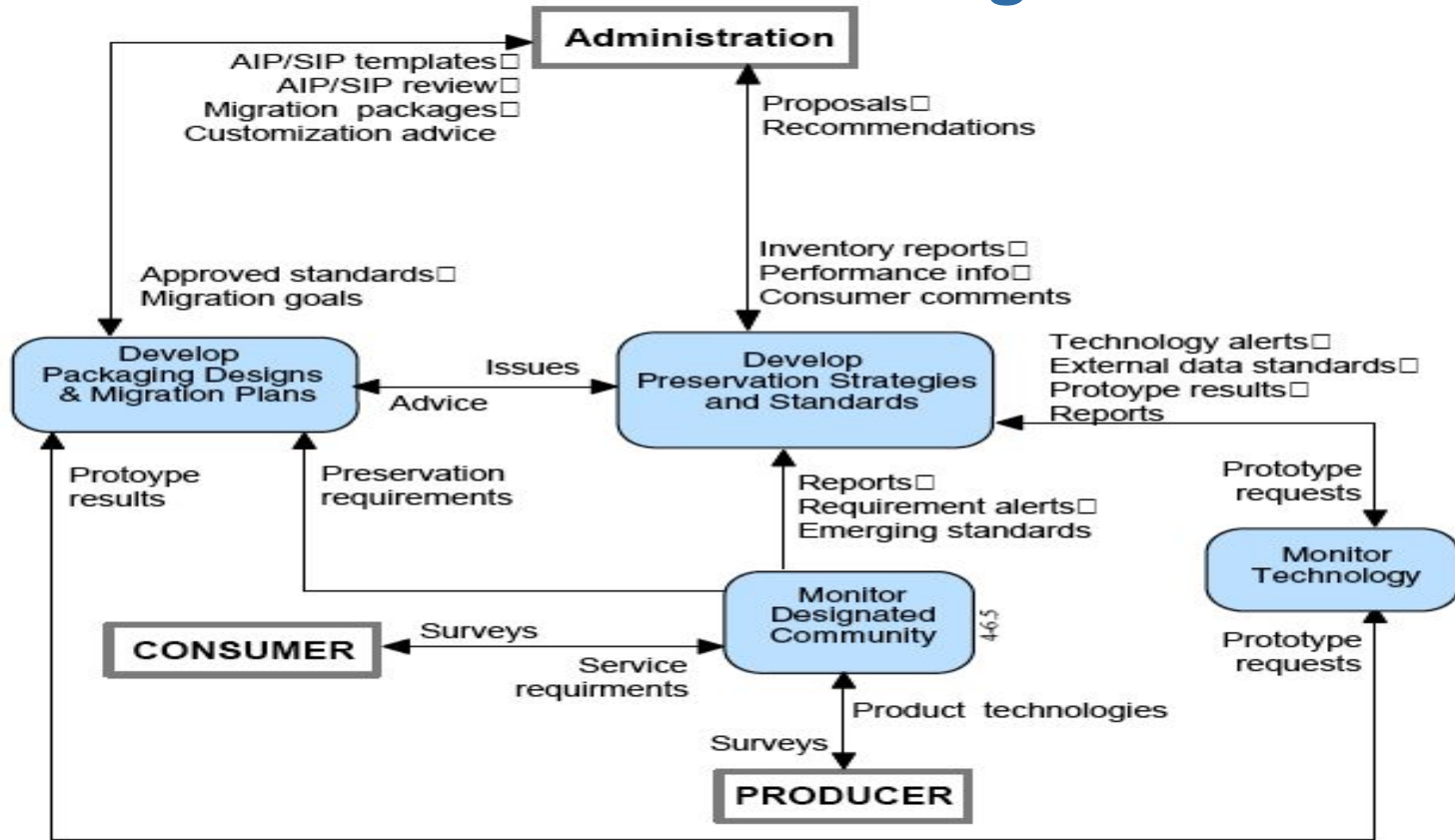
Verifica que los AIP y SIP suministrados sigan las especificaciones. En el caso de SIP y de AIP verifica la comprensión por parte de la comunidad designada. Verifica que la Información de representación y la PDI son adecuadas y comprensibles para la comunidad designada.

Mantiene un registro de de solicitudes y revisa periódicamente los contenidos del archivo para determinar si los datos están disponibles.

Crea/mantiene/borra las cuentas de acceso de los consumidores.



# Functions of Preservation Planning



# Entidad OAIS preservation planning

**Descripción:** Interactúa con los consumidores y productores de archivos. Proporciona reportes, alertas de requisitos y estándares independientes.

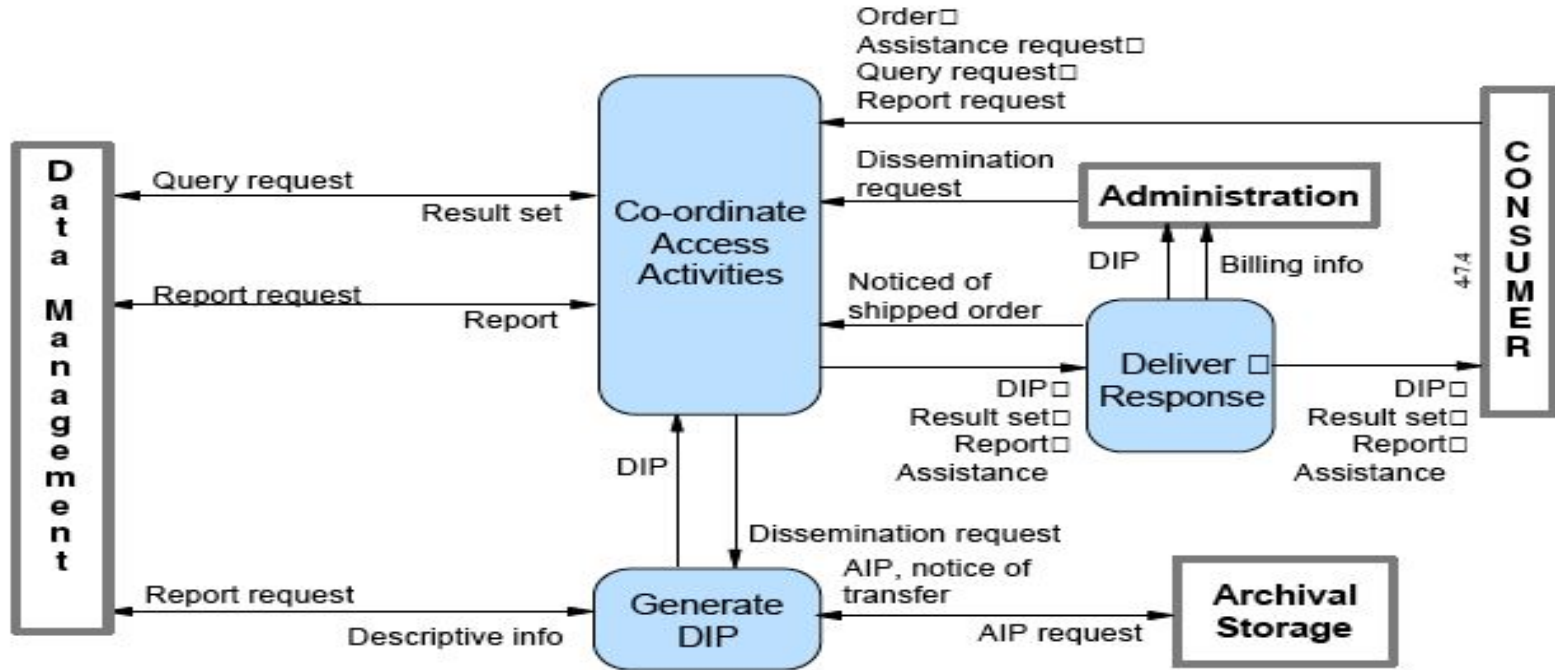
Identifica tecnologías que pueden causar obsolescencia.

Desarrolla y recomienda estrategias y estándares, que envía a *administration*.

Desarrolla nuevos IP y planes de migración y prototipos, para implementar políticas y directivas de administración de IPs.



# Functions of Access



# Entidad OAIS Access

**Descripción:** Proporciona una interfaz única de usuario para el acceso a la información de los archivos. Tiene 3 categorías, los *query requests*, los *result sets* y los *report requests*.

Acepta los requerimientos de los paquetes de diseminación recuperados de los AIP de la entidad *archival storage* y transmite un *report request* al *Data Management* generando un DIP.

Entrega las respuestas en línea y fuera de línea de los consumidores.

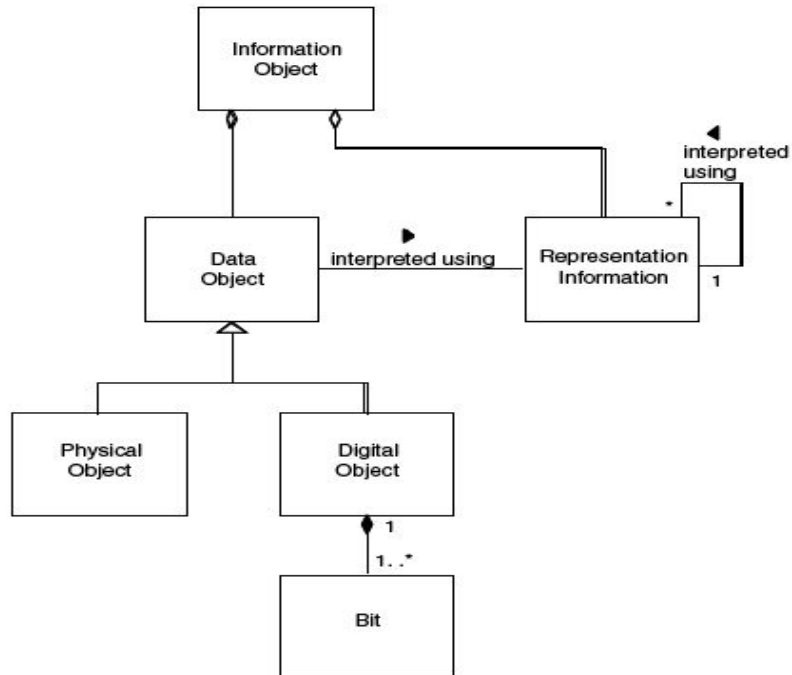
# OAIS

Modelo de Información

Sección 4.2



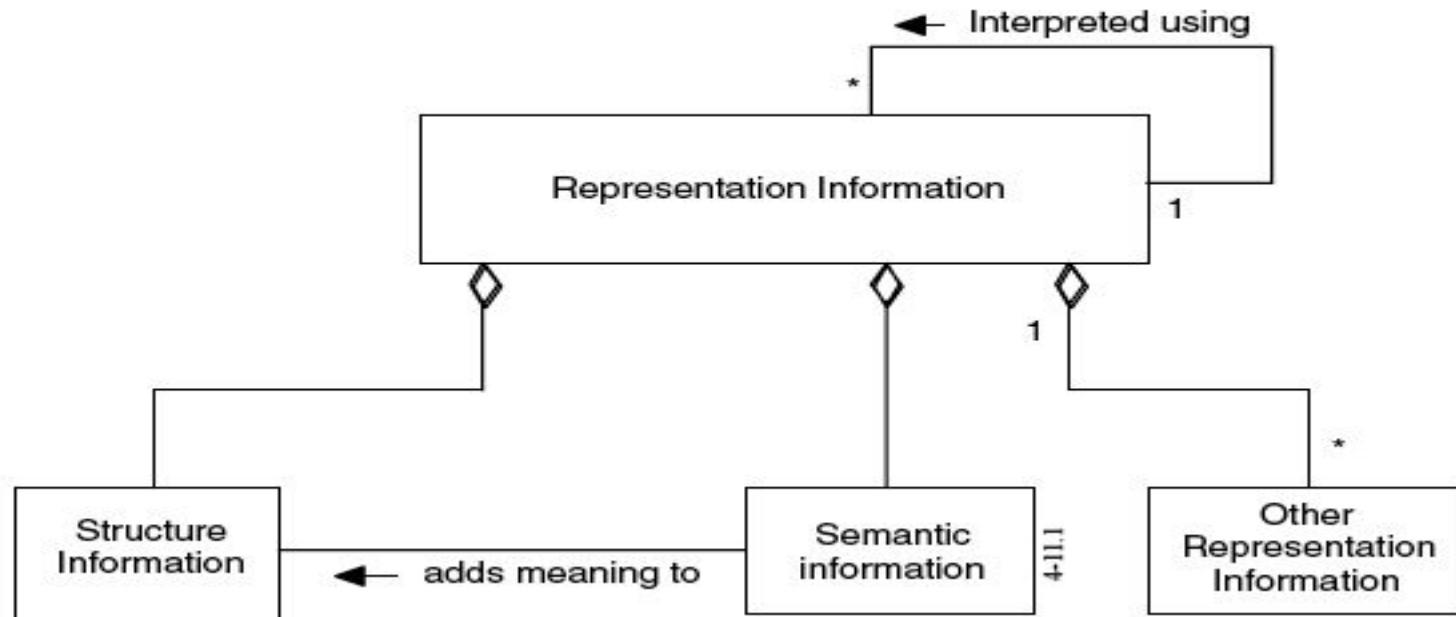
# OAIS Objeto de información



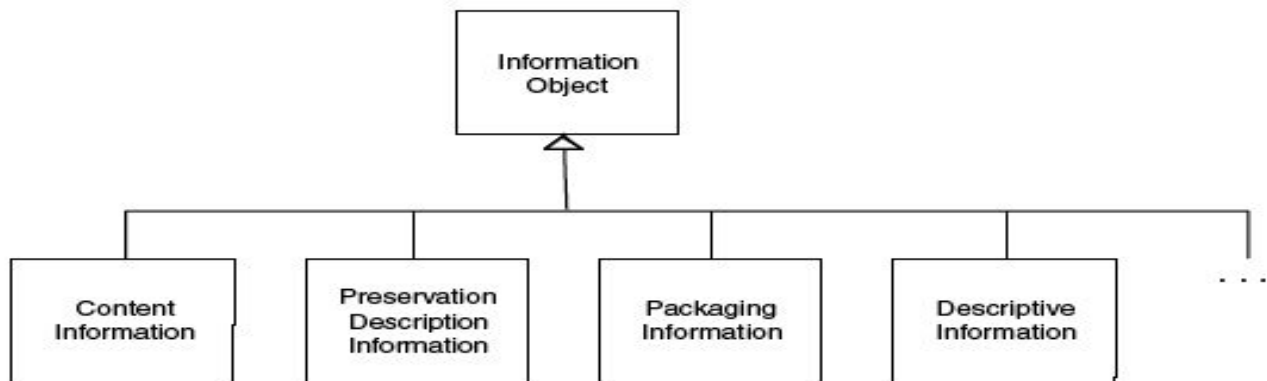
- El **Objeto de Información** está compuesto de un Objeto de Datos, que puede ser físico o digital, e Información de Representación que permite la interpretación completa de los datos.



# Representation Information Object



# Tipos de objetos de información



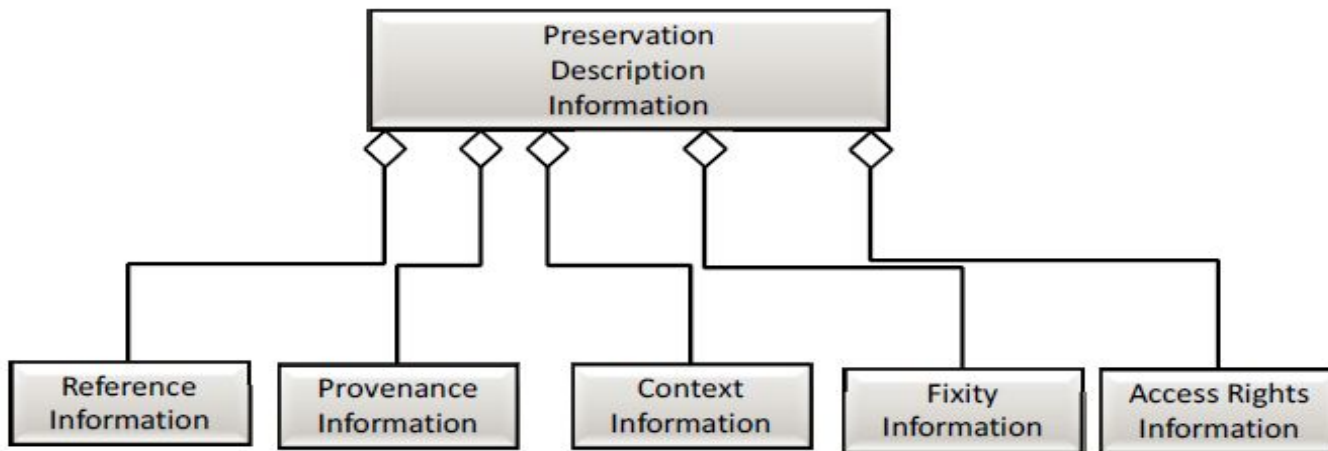
Los objetos de información se clasifican por su contenido y función como : objetos de información de contenido, de descripción de la preservación, de empaquetado y de información descriptiva.



# Información de contenido

- La información de contenido es el conjunto de información que es el objetivo original de la preservación de la OAIS.
- La información de contenido es el contenido de datos del objeto, junto con su representación de la información.
- Los objetos de datos contenidos en la información de contenido puede ser un objeto digital o un objeto físico (por ejemplo, una muestra física de microfilm,).  
Cualquier objeto de información puede servir como información de contenido.

# Información descriptiva de preservación (PDI)



**Figure 4-16: Preservation Description Information**



# Información descriptiva de preservación

**Información de referencia:** identificación y descripción de uno o más mecanismos para proporcionar los identificadores asignados para la información del contenido. También proporciona los identificadores.

**Información de contexto:** documenta las relaciones de la información de contenido con su entorno (¿por qué la información de contenido fue creada y cómo se relaciona con otra información de contenido).

# Información descriptiva de preservación

**Información de procedencia:** los documentos de la historia de la información de contenido (origen o fuente, los cambios y la custodia) de procedencia puede ser visto como un tipo especial de información de contexto.

**Información de la fijeza:** proporciona los controles de integridad de los datos o claves de validación usados para asegurar que la información de contenido no ha sido alterada.

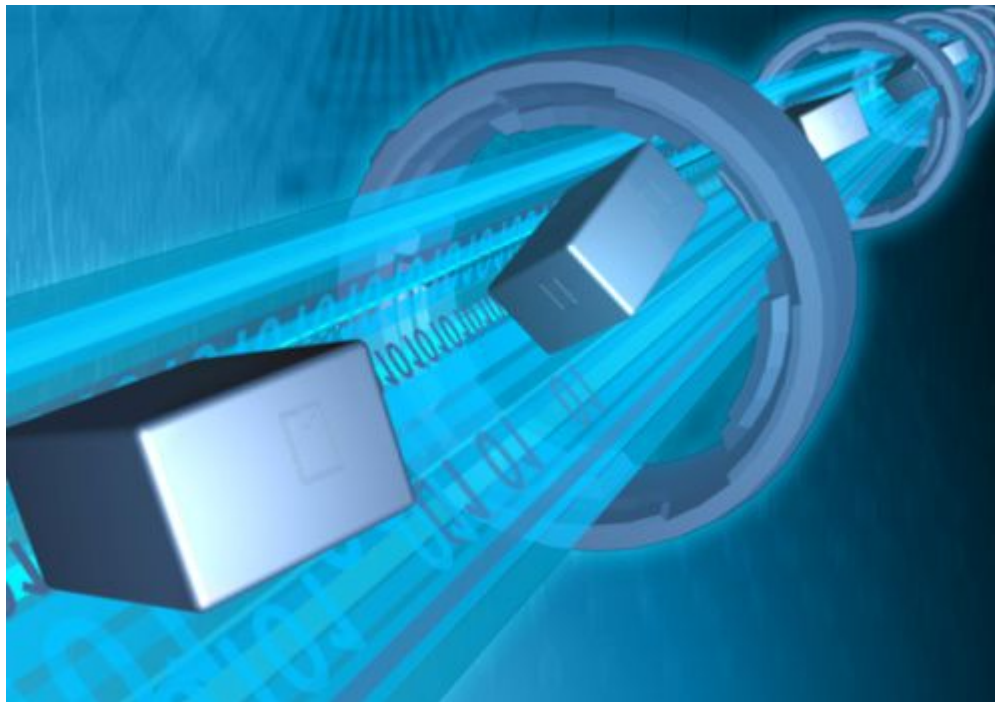
**Información de sobre derechos de acceso:** proporciona los permisos de uso de la información de contenido.

# Paquetes de información en OAIS

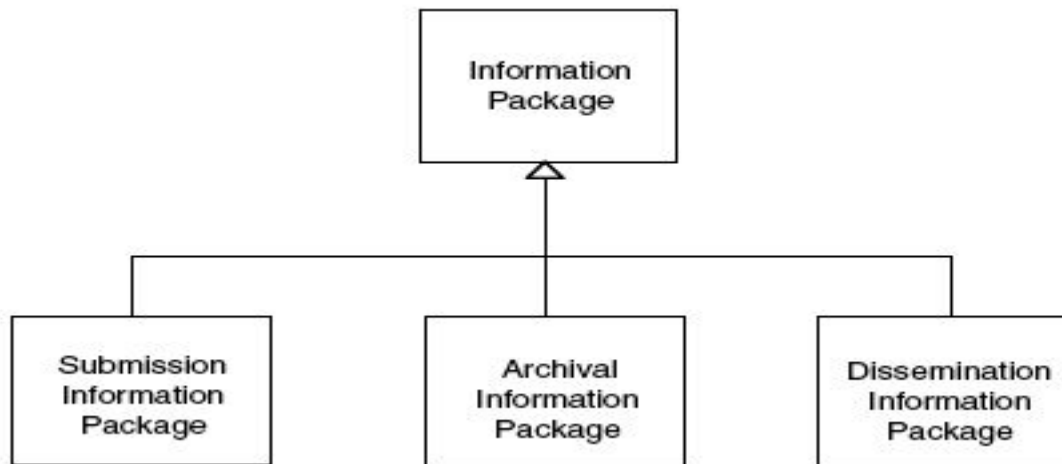
- Las estructuras de información conceptual necesarias para cumplir las funciones OAIS.
- Un paquete de información es un contenedor.
- Hay varios tipos de paquetes de información que se utilizan en el proceso de archivo. Estos paquetes de información pueden ser utilizados para:
  - estructurar y almacenar las participaciones OAIS (AIP);
  - para transportar la información desde el productor hasta el OAIS (SIP)



# Paquetes de información en OAIS



# Tipos de paquetes de información



# SIP

- La forma y el contenido detallado de un SIP típicamente se negocia entre el productor y el OAIS.
- La mayoría de los SIPs se tiene alguna información de contenido y algunas PDI, pero se puede requerir varios SIPs para proporcionar un conjunto completo de información de contenido y PDI asociados.
- Si hay varios SIPs que utilizan el mismo Repositorio de información, éste sólo se proveerá una **vez?**.
- Dentro de la OAIS, uno o más SIPs se transforman en uno o más AIPs para su conservación.

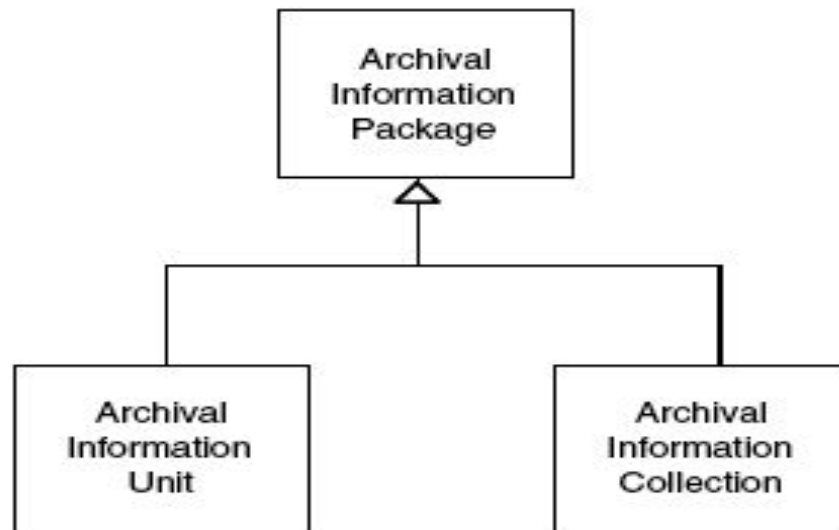


# AIP

Un Paquete de Información de Archivo es una especialización del Paquete de Información. El PIA se define para proporcionar una forma concisa de referirse a un conjunto de información que tiene, en un principio, todas las cualidades necesarias para una Conservación a Largo Plazo de un determinado Objeto de Información, de forma permanente o indefinida. El PIA es en sí mismo un Objeto de Información que contiene otros Objetos de Información.



# Tipos de AIPs



# DIP

- En respuesta a una petición, el OAIS ofrece la totalidad/parte de la AIP a un consumidor en la forma de un DIP.
- El DIP también puede incluir las colecciones de la AIP, según el acuerdo de difusión entre OAIS y Consumidores.
- La información de paquetes siempre estará presente para que el consumidor distinga claramente la información solicitada.
- El propósito de la información descriptiva de un DIP es dar al consumidor información suficiente para reconocer el DIP de entre los posibles paquetes similares.



# Participantes

- El productor es el autor o quien lo presenta, y suministra los artículos para el archivo a través de los procedimientos de entrada (ingest/ingesta) que constituiría el **flujo de trabajo de presentación**.
- El paquete de información presentada resultante (SIP, Submission Information Package) se convierte en el paquete de información archivada (AIP, Archival Information Package) a través del proceso del **flujo de trabajo de post-presentación** y por lo tanto pasa al almacenamiento de archivos.



# Participantes

- Sección especializada para la administración adjunta a la gestión: **administradores**.
- Se relaciona con la sección de gestión de datos y la de planificación de la conservación.
- Esto permite una gestión estructural y también ayuda a mantener los AIPs a lo largo del tiempo.



# Participantes

Para satisfacer los diversos requisitos detallados que exige este modelo de referencia, un sistema de repositorio debe captar todos los metadatos relevantes para convertir el SIP en un AIP con garantía de calidad y rastros de auditoría colocados al momento de la presentación, además de la información asociada como por ejemplo las normas del formato de archivo y otro tipo de metadatos técnicos.

# Participantes

El AIP debe ser colocado en el archivo de almacenamiento, y se deben mantener referencias actualizadas en el sistema de gestión de datos. El almacenamiento del archivo debe permitir el uso de técnicas de almacenamiento tradicionales y verificadas, por ejemplo copias de seguridad y la verificación del contenido a lo largo del tiempo y la migración a otros medios de almacenamiento.



# Participantes

- La **administración** del sistema requiere la creación de políticas y autorizaciones para permitir el acceso, y la gestión de la configuración del sistema.
- Relacionada con el proceso de ingesta, la auditoría de presentación se define dentro de su alcance y en última instancia pasa a formar parte del AIP, y también la negociación del acuerdo de presentación, que está muy asociado al tema de las licencias.
- OASIS recomienda que los administradores manejen los pedidos de diseminación y se encarguen de resolver los problemas de atención al cliente en caso de que surgieran o fueran relevantes al manejo del repositorio.

# Participantes

El **acceso** a los materiales se garantiza al consumidor, quien se define según el modelo como un miembro de la comunidad designada, este es un concepto que detalla quién debe comprender el material: si la búsqueda archivada está en el campo de la física, la comunidad designada se especificará como “físicos” y los metadatos y los documentos relacionados respecto del significado del contenido se omiten por la razón de que la comunidad designada podrá comprender el material sin recurrir a estos.



# Participantes

- La comunidad se asigna con el DIP, que puede contar con la mediación de los administradores o puede ser manejado exclusivamente por el sistema.
- El DIP se obtiene realizando una búsqueda en el módulo de gestión de datos, que a su vez ofrece referencias a los AIPs que deben convertirse y entregarse.
- El modelo recomienda mantener un registro de todas las solicitudes de contenido que se agregarán al rastro de auditoría del AIP.

# Participantes

El módulo de **planificación de la conservación** abarca todas estas secciones, y su trabajo es desarrollar estrategias y normas de conservación, monitorear las últimas novedades y avances en el campo, y monitorear los cambios en la comunidad designada, para que toda la información nueva que se solicite se pueda adjuntar a los AIP correspondientes.



# Participantes

Los resultados de este módulo servirán como pautas para que los administradores diseñen sus políticas, y en última instancia, guiarán las actividades de conservación de los materiales. Debe tenerse en cuenta que la migración y demás políticas de cambio de formatos, exigen la generación de nuevos AIP, y de ninguna manera deben modificarse los ya existentes.



# Sección 5: Perspectivas sobre preservación

- 5.1 Información para la preservación.
  - Motivadores para la migración.
  - Contexto.
  - Tipos de migración: refresco, replicación, reempaquetado, transformación.
  - Versiones de los AIP.
- 5.2 Preservación del acceso.

# Saliendo de la 14721



# Aproximaciones a la preservación

Existen numerosas estrategias para asegurar la preservación de la información:

- Guía UNESCO: “Directrices para la preservación del patrimonio cultural”.
- Servicio PRONOM
- Herramienta DROID
- Metadatos de Preservación
- El estándar PREMIS

# Preservación en el repositorio

Basado en el servicio de PRONOM provisto por The National Archives (TNA) y la herramienta DROID (Digital record object identification service) que usa los perfiles de formato de más de 200 repositorios del registro PRONOM. DROID permite clasificar y evaluar los riesgos de los distintos formatos que usa un repositorio y de este modo elaborar un **plan activo** de preservación que identifique el formato o sugiera el cambio.

# Metadatos

Los metadatos se clasifican en distintas categorías de acuerdo con las funciones que cumplen: los **descriptivos** ayudan a describir y recuperar los recursos; los **administrativos** gestionan un recurso: mantenimiento, almacenamiento y entrega, incluyendo datos técnicos sobre la creación, control de acceso y calidad, gestión de derechos, utilización y condiciones de preservación, migración, etcétera; y los **metadatos estructurales** refieren la estructura interna del recurso y los elementos que lo integran, indican cómo reunir objetos digitales complejos para que se puedan utilizar, por ejemplo: página, sección, capítulo, numeración, índices, tablas de contenidos, entre otros.

Los **metadatos de preservación** soportan los datos necesarios para cumplir con una serie de requerimientos de preservación con el objetivo de asegurar la utilización a largo plazo de un recurso digital. A continuación se incluyen algunos de estos requerimientos sobre cada objeto digital:

- Debe mantenerse en el repositorio de manera segura sin perderse ni ser modificado sin autorización.
- Se debe conocer su creador.
- Si cambia se debe conocer quién realizó el cambio.
- Debe poder localizarse y entregarse al usuario.
- Debe almacenarse en soportes que puedan leer los sistemas actuales de manera que el usuario pueda comprenderlos.



- Del mismo modo las estrategias de emulación y migración requieren metadatos sobre los formatos de los objetos originales y los entornos de hardware y software que los soportan.
- Soportar la autenticidad mediante la documentación de la *procedencia digital* a través de su cadena de custodia y el historial de cambios autorizados.
- El repositorio debe disponer de los derechos suficientes como para llevar adelante las transformaciones necesarias para mantener el acceso al objeto.
- Si el objeto está relacionado con otros del repositorio o de otros depósitos externos, estas relaciones deben guardarse.

# Metadatos de preservación

En resumen, los **metadatos de preservación** están destinados a almacenar los detalles técnicos sobre el formato, la estructura, el acceso y el uso de los contenidos digitales, la historia de todas las acciones realizadas en el recurso, incluyendo los cambios, la información de autenticidad, las características técnicas o la historia de la custodia y las responsabilidades y la información sobre los derechos con que se cuenta para realizar las acciones de preservación.



# PREMIS

PREMIS es un grupo de trabajo internacional patrocinado por Online Computer Library Center (**OCLC**) y Research Libraries Group (**RLG**) que, como su nombre lo indica, se enfoca en estrategias de implementación de metadatos de preservación en Archivos Digitales.

En 2008, este grupo elaboró el Diccionario de Datos PREMIS para Metadatos de Preservación, el cual define los metadatos de preservación como *“la información que utiliza un repositorio para dar soporte al proceso de preservación digital”*.

# Diccionario de datos PREMIS

El diccionario define un conjunto de *unidades semánticas*, propiedades, e información que la mayoría de los repositorios necesita conocer de sus entidades para asegurar la preservación.

PREMIS plantea la necesidad de representar las unidades semánticas de forma abstracta, aunque no regula su implementación ni representación.



# Modelo de Datos PREMIS

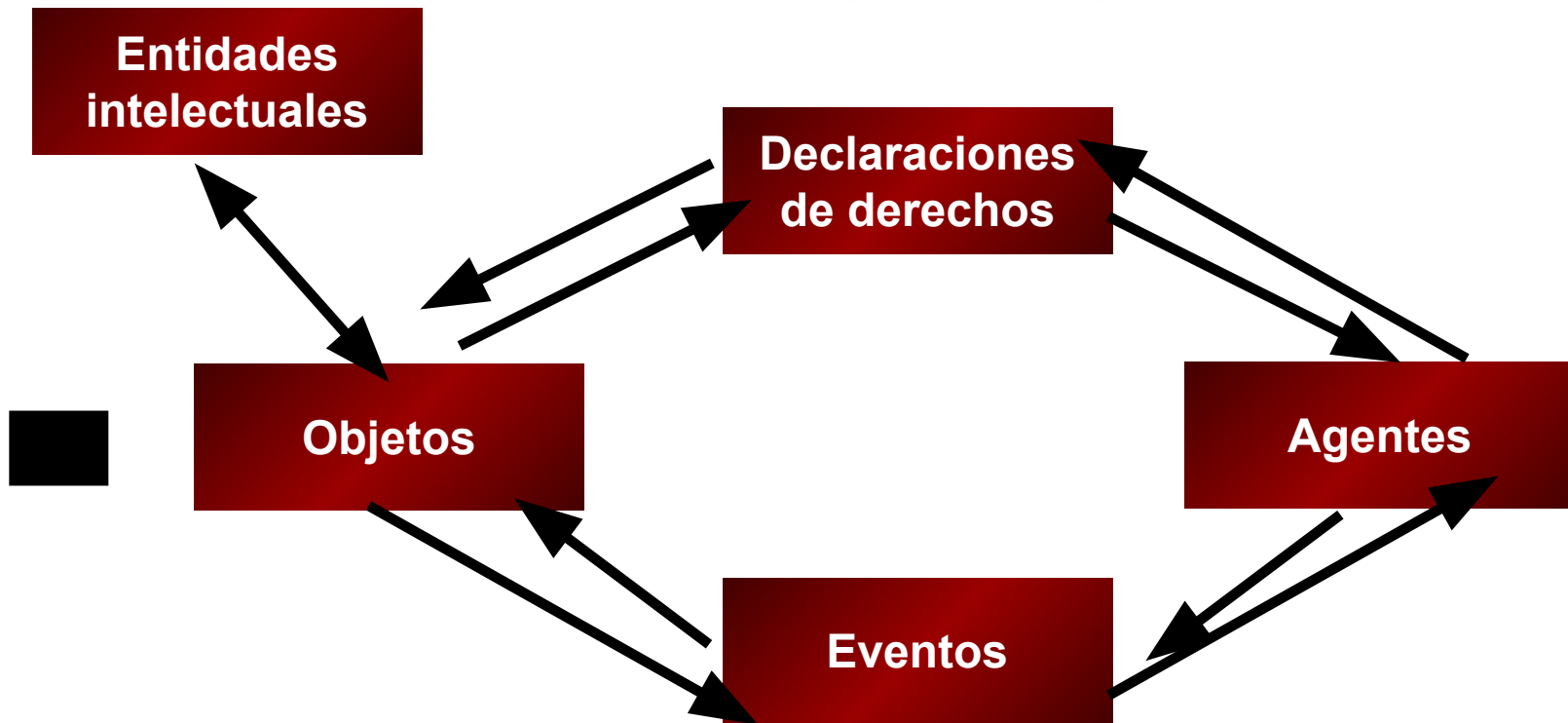
Las entidades que este modelo define se denominan:

- Entidades intelectuales
- Objetos
- Derechos
- Agentes
- Eventos



## Modelo de datos

PREMIS



# PREMIS

**Entidad intelectual:** conjunto coherente de contenido que se describe como una unidad: por ejemplo, un libro, un mapa, una fotografía, una publicación periódica,... etc. Una entidad intelectual puede incluir otras entidades intelectuales: por ejemplo, un sitio web, puede incluir una página web, una página web puede incluir una fotografía.

Una entidad intelectual puede tener una o más representaciones...

**Objeto** en PREMIS difiere de la definición de objeto digital normalmente utilizada en la comunidad de las bibliotecas digitales, que entiende el término “digital object” como una combinación de identificador+datos+metadatos. No es en absoluto un conflicto. La entidad objeto en el modelo de PREMIS es una abstracción definida sólo para agrupar atributos (unidades semánticas) y clarificar relaciones.



# PREMIS

**Evento:** Acción que incluye al menos un Objeto Digital y/o un agente conocido en el repositorio de preservación

**Agente:** Actor (humano, máquina o software) asociado con uno o más eventos asociados a un objeto digital

**Derechos:** Afirmación de uno o más derechos o permisos que pertenecen a un objeto digital y/o a un agente



# Entidad intelectual

Una ***entidad Intelectual*** es un conjunto de contenidos que se considera como una unidad intelectual individual al propósito de gestión y descripción. El diccionario de datos no determina los metadatos descriptivos a vincular a una entidad intelectual, sino que deja abierta la elección a cualquier formato deseado.



# Objetos

Los **Objetos** son unidades discretas de información en forma digital, que se clasifican en tres tipos: **archivo (file)**, **representación (representation)** y **cadena de bits (bitstream)**. El objeto *archivo* es tal cual entendemos normalmente, es decir un archivo PDF de un capítulo de un libro, un archivo JPEG, etc. El objeto *representación* es el conjunto de todos los archivos que se necesitan para representar la entidad **Intelectual** (un libro, una foto, un mapa, un sitio web), incluyendo los metadatos estructurales. Los objetos *cadena de bits* son subconjuntos de archivo con propiedades útiles a la preservación, en el ejemplo del archivo JPEG cada imagen puede tener sus propios identificadores y metadatos. La información que se puede registrar en los objetos incluye: un identificador, la integridad, el tamaño, información sobre la creación, sobre el entorno, el soporte y la relación con otros objetos y otros tipos de entidades.



# Eventos

La entidad **Eventos** agrega información sobre acciones que un agente, o varios, lleva adelante sobre los objetos de los repositorios, por ejemplo: el identificador del acontecimiento (no repetible), el tipo (creación, migración, etc), la fecha de ocurrencia del evento, la descripción y el resultado codificado del acontecimiento así como los agentes.



# Agentes

Los **Agentes** pueden ser personas, organizaciones o aplicaciones de software con actividades o responsabilidades en los eventos. El Diccionario de datos aconseja como información: un identificador único, el nombre del agente y su tipo (por ej. persona).



# Derechos

La entidad ***Derechos*** agrega información sobre los permisos y derechos sobre los objetos que le han sido otorgados al repositorio por parte su poseedor. Se debe incluir: identificador único, un agente que concede, datos sobre la licencia y las acciones permitidas.





## VIII Workshop

*Proyectos digitales*

<http://www.um.es/biblioteca/Workshop/introduccion.htm>

# Metadatos para la preservación digital: PREMIS

**Eva M<sup>a</sup> Méndez Rodríguez**

emendez@bib.uc3m.es

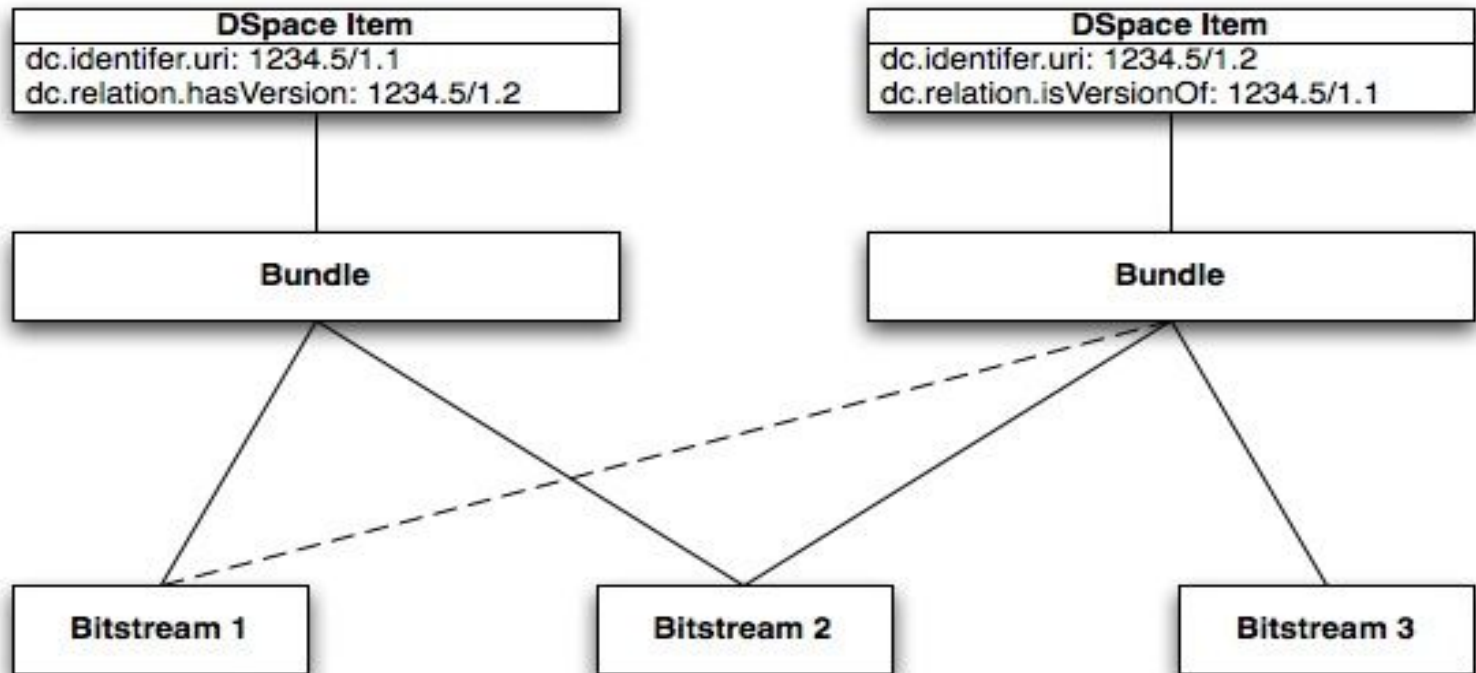
Dpto. Biblioteconomía y Documentación



Universidad Carlos III de Madrid



# DSPACE: Modelo de datos



# OAIS y DSPACE

**Dspace mantiene tres grupos lógicos de metadatos para el contenido archivado:**

**1) Metadatos descriptivos:** soporta múltiples formatos de metadatos no jerárquicos y permite mantener metadatos de distintos formatos para un mismo ítem. Para colecciones y comunidades, sólo permite gestionar una serie de campos fijos y básicos aunque probablemente en las próximas versiones de Dspace se permita configurar los metadatos como sucede con los items.

**2) Metadatos Administrativos:** incluye metadatos de preservación, información de procedencia y permisos de acceso y uso sobre cada item, colección y comunidad. Se registran campos adicionales sobre varios elementos: tamaño, suma de comprobación y tipo Mime de cada archivo.



# OAIS y DSPACE

**3) Metadatos estructurales:** mantiene sólo unos pocos metadatos estructurales sobre los archivos de cada ítem p.e. la relación de pertenencia entre paquetes de archivos (bundles) y orden. Dependiendo cómo se armen las comunidades y colecciones, pueden considerarse las relaciones de pertenencia a colección como un metadato estructural. Para los ítems, es posible almacenar información estructural en los bitstreams o metadatos, aunque inicialmente no se provee nada.



# OAIS y DSPACE

Dspace está influenciado por el modelo OAIS, en términos de metodología y funciones. Siempre que es posible, adopta el modelo y vocabulario OAIS para articular su propio diseño.

El modelo de datos, los metadatos y el módulo de autorizaciones cumplen con el modelo de referencia. En términos de la preservación digital, el sistema proporciona los metadatos de preservación como en el modelo OAIS y el identificador persistente Handle.



# OAIS en DSPACE

La arquitectura de Dspace no coincide con el modelo funcional de OAIS; por ejemplo, no hay disponible un módulo separado de planeamiento de la preservación (*Preservation Planning*).



# OAIS en DSPACE

El AIP, surge en Dspace como un objeto lógico que se conforma de datos dispersos en varias tablas dentro de una base de datos relacional y en archivos en el sistema de archivos.

La llamada unidad de información OAIS en Dspace es el *item* y se gestiona como AIP en formato METS. Para cumplir con los requerimientos mínimos de preservación sobre los archivos.



# OAIS en DSPACE

Dspace permite exportar e importar paquetes DIP y SIP basados en METS o, si se desarrolla un packager plugin ad-hoc, en cualquier otro formato. Adicionalmente, es posible importar y exportar AIPs completos de forma muy simple, generando paquetes totalmente autocontenidos para items, colecciones, comunidades e incluso para todo el repositorio. A diferencia de los SIP y DIP, estos AIP contienen todos los datos sobre el recurso en el repositorio.

# OAIS en DSPACE

Dspace brinda mecanismos para comprobar el estado de los elementos que se encuentran en un repositorio. Algunos de estos mecanismos, pueden utilizarse para corroborar, por ejemplo, la información requerida por la norma OAIS.

Uno de estos mecanismos es la posibilidad de realizar **tareas de curación** automáticas



# OAIS en DSPACE

## Tareas de Curación en Dspace

### Definición

Una **tarea de curación** es una actividad mediante la cual se comprueba y/o resuelve el estado de un aspecto de un elemento, con el objetivo de asegurar la preservación del mismo en el tiempo.

Una tarea de curación iterativa o repetitiva, puede ser automatizada.

### Tareas de curación en Dspace

Dspace proporciona herramientas para implementar distintas tareas de curación automáticas.



# OAIS en DSPACE

## Tareas de Curación en Dspace

### ¿Qué se puede realizar?

- Comprobación de la existencia del elemento
- Comprobación de los metadatos y de la calidad de los mismos
- Arreglos específicos en propiedades o en metadatos de un elemento
- Evaluaciones y recopilación de estadísticas de un elemento
- **Aplicación** de una rutina **repetitiva** a varios elementos distintos.



# OAIS en DSPACE

## Tareas de Curación en Dspace

### ¿Cómo se pueden realizar?

#### Dspace provee:

- Algunas tareas de curación predefinidas
- Una interfaz de tareas de curación, que permite implementar de manera relativamente sencilla, nuevas tareas de curación
- Archivos de configuración para las tareas de curación, lo que permite que las mismas guarden y utilicen configuraciones personalizadas
- Distintos lugares desde donde pueden ser utilizadas: la consola, la xmlui.



# OAIS en DSPACE

## Tareas de Curación en Dspace

### ¿Cómo se pueden realizar?

#### Se pueden implementar:

- El cuerpo de la tarea de curación, en código JAVA
- El archivo de configuración (de ser necesario) con las propiedades a almacenar en texto plano
- La vista adecuada si va a utilizarse desde alguna interfaz de usuario que no sea la consola
- El llamado y el análisis de las estadísticas obtenidas ( se puede implementar como **cronJob**)



# OAIS en DSPACE

## Tareas de Curación en Dspace

### Ejemplos:

- Control de enlaces válidos a los elementos del repositorio
- Copia automática de metadatos entre dos elementos relacionados
- Arreglo de “metadatos controlados” (requeridos) por el repositorio en los elementos que no los posean
- Normalización de metadatos controlados en un elemento
- Evaluación automática del estado de los elementos según la directiva OAIS



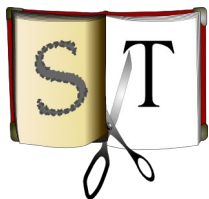


# Herramientas



# Herramientas

Una vez que se recibe material en el repositorio es necesario hacer un correcto relevamiento de los metadatos a partir de la obra, así como analizar los típicos problemas que pueden presentarse a partir de la variedad de materiales y formatos. A continuación se presentarán brevemente algunos casos y posibles soluciones que permiten llevar a cabo de un modo más simple la tarea de descripción documental.



**Video Downloader GetThemAll**

ofrecido por FVD Inc



Adobe Acrobat DC



sedici.unlp.edu.ar  
prebi.unlp.edu.ar

# Tratamiento y mejoras de archivos PDF

Suele ocurrir que el repositorio recibe archivos PDF de revistas o eventos que deben divididos para que cada artículo sea cargado por separado. Una de las opciones disponibles es el Acrobat DC.

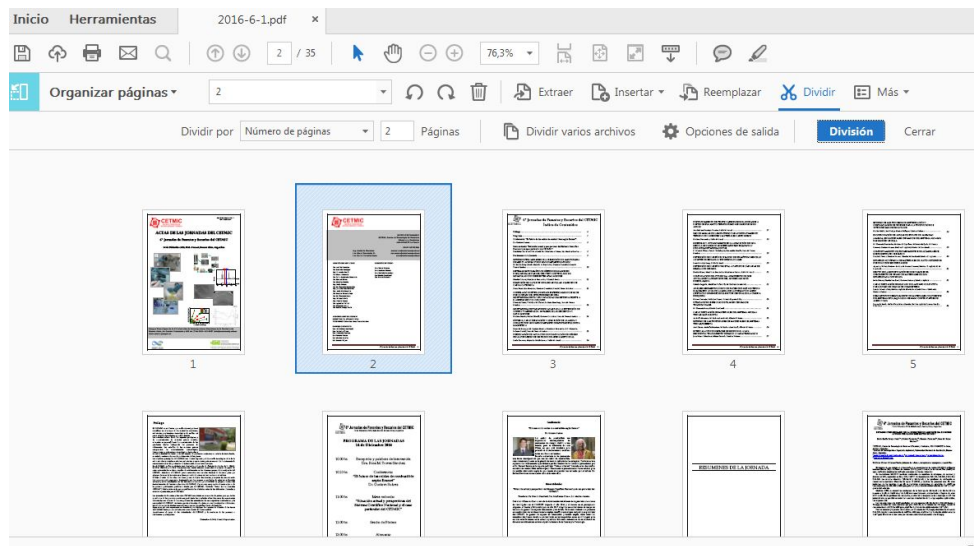


Adobe Acrobat DC



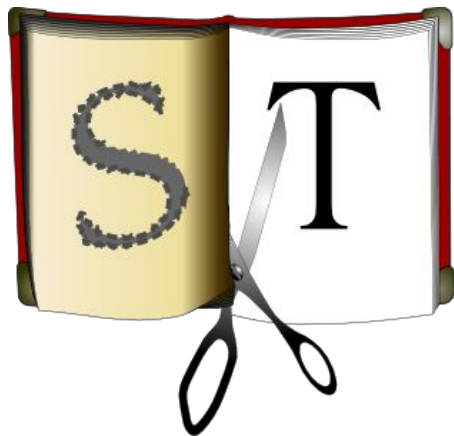
# Acrobat DC - División de Archivos

Este programa permite dividir por cantidad de páginas o por marcadores. En el caso de que el archivo no cuente con marcadores es posible agregarlos manualmente de manera fácil para poder dividir el archivo correctamente.



# Scan Tailor

Otra ocurrencia común es que lleguen pdfs con hojas escaneadas en modo apaisado o que los editores de una revista o un libro envíen la versión para imprimir. En este caso debemos contar con un programa que como el Scan Tailor permita dividir al medio las hojas y organizar el pdf de manera correcta.



# Scan Tailor

Se puede elegir manualmente el área correspondiente a cada hoja o el programa puede hacerlo de manera automática.

Archivos Herramientas Ayuda

- 1 Corregir Orientación
- 2 Dividir Páginas
- 3 Alineación
- 4 Selección de Contenido
- 5 Margenes
- 6 Salida

Debricon de Pagina

Debricon Automática

Cambiar...

Línea Divisoria

Auto Manual

**1º Jornada de Pasantes y Docentes de Escuela Media y Grado del CETMIC.**  
16 de Noviembre de 2011, M.B. Correo, Buenos Aires, Argentina

**CETMIC**  
CARACTERIZACIÓN POR DRYS Y CARÁCTERÍSTICAS PERIFÉRICAS DE SISTEMAS AG-Cu-Si-MONTMORILLONITA

Javier Lantaras\*, Mariela Fernandez† y Rosa Torres Sánchez†

†CETMIC, Centro de Tecnología de Recursos Minerales y Carbonácos, CIC-CONICET La Plata, Corrientes, Argentina  
‡Departamento de Química, Facultad de Ciencias Exactas UNLP

**Palabras Clave:** agua y Pb, Montmorillonita, Plata, Cobre, Uso como bioadsorbente.

Estudios previos, realizados en el CETMIC, demostraron la aplicabilidad de la Ag adsorbida en montmorillonita (Ag-MMT) como sistema de floculación/adsorción de fango activo para evitar la contaminación de cuerpos de agua. En este trabajo se estudia la especificidad de sistemas Cu-MMT y Ag-Cu-MMT para su uso en la actividad generadora del primer, la de adsorción y fangos que brinda el Cu y Ag-Cu-MMT para su uso en la adsorción de los metales Ag-MMT, Cu-MMT, Ag-Cu-MMT y Ag-Cu-MMT.

El objetivo de este trabajo es evaluar la inversión de los metales en la síntesis de la MMT y los reagentes adsorbidos especificando un protocolo para evitar su contaminación. Los análisis de DRX, permitieron determinar la relación de los átomos de cobre de oxidación de ambos sistemas. Los resultados fueron interpretados en términos de la adsorción de los metales en los distintos sistemas.

El momento del pico de reflexión (01) (Fig. 1), respecto de la muestra inicial MMT (13.42 Å), evidenció la incorporación de los cationes en la estructura de la MMT.

**Figura 1.** Espectro de DRX del pico de reflexión (01).

**Figura 2.** Curvas de potencial zeta vs pH.

Las curvas de potencial zeta, en el rango de pH estudiado (pH= 3-7), muestran valores más negativos, para las muestras caracterizadas con los metales Ag o Cu, potencial Zeta = -47 mV con los óxidos para MMT = 30 mV, mientras que las muestras con ambos metales (Ag-Cu-MMT y Ag-Cu-MMT) MMT) tienen valores intermedios entre las anteriores, la muestra MMT evidenciando la incorporación de ambos metales en su superficie externa. Los espectros de XPS, evidenciaron la incorporación de Ag y Cu en la superficie de la MMT, con presencia simultánea de Ag<sup>+</sup> y Cu<sup>2+</sup> tanto a su superficie externa, en relación cercana a 1:5 y un menor desplazamiento de la Ag que el aumento de concentración de Cu en las muestras nuevas. El análisis de difracción de Ag-Cu-MMT, indica una menor liberación de Ag que Cu en todas las muestras. La presencia de cationes en el agua de los ácidos, permite generar una mayor liberación de Ag y Cu de la MMT.

**Prólogo**

El CETMIC es un Centro que realiza investigaciones científicas en el campo de los materiales nanoestructurados, sintéticos y aplicaciones tecnológicas de estos, en áreas relacionadas con la industria. Los estudios sobre estos los aspectos tecnológicos y de caracterización de materiales nuevos (textiles, metales, etc.) son parte de la capacitación de los profesores (tesis, maestrías, etc.) y de la colaboración. Los estudios de los materiales nuevos comprenden la caracterización morfológica, electroquímica y aplicaciones tecnológicas destinadas a su uso industrial. Referente a la elaboración de productos nuevos, se estudia la síntesis, caracterización, control de calidad y distribución de los productos.

Los objetivos principales del CETMIC son la investigación y el desarrollo tecnológico de todo lo referido a los materiales nanoestructurados, sintéticos y sus nuevas formas de uso. La transferencia de tecnología al medio productivo y la formación de recursos humanos.

El CETMIC, se lleva adelante desde diciembre y de grado de docentes formados de la UNLP y la UBA. Actualmente, varios docentes argentinos de grado de estas universidades realizan estudios de colaboración en los diversos países de investigación del CETMIC. Asimismo el CETMIC, posee convenios con escuelas medias de su zona, por estar quienes alumnos de los mismos tienen presencia de entrenamiento en temas científicos.

Con el Objetivo de comunicar internamente, y horizontalmente los avances y resultados de estas tareas llevadas adelante por los docentes en que el CETMIC se propone realizar la Primera Jornada de Pasantes y Docentes de Escuela Media y Grado del CETMIC. Como parte comprendiendo el primer momento de la presente publicación periódica editada por el CETMIC, titulada "Actas de las Jornadas CETMIC", Publicación en la que se publicarán los distintos actos de los diferentes Jornadas o eventos organizados por el CETMIC.

Los **Jornadas de Pasantes y Docentes de Escuela Media y Grado del CETMIC** consisten en una reunión de grado que reúne a cada uno de los pasantes que completaron sus actividades durante sus respectivos cursos de grado con el resto de sus compañeros. Los actividades de sus respectivos institutos, y la comunidad del CETMIC, incluyen previa a la sesión de grado. Investigadores del CETMIC, trabajan un par de horas antes del sistema Científico Argentino en Corrientes, del Centro de Tecnología de Recursos Minerales (CETMIC) en particular, con el objetivo de orientar a los estudiantes las posibilidades de inserción laboral que presenta el sistema.

Esperamos que esta primera experiencia sea fructífera y multiplique los espacios de diálogo de las tareas del CETMIC, para pasar adelante con el resto de la comunidad.

Agradecemos el apoyo de los autoridades del CETMIC, y la conformación de las instituciones participantes.

Noviembre de 2011, Corrientes Organizador.

**COORDINADORES**

**DRYSS TIVA**  
Dr. Alberto Sola

**CAROLINA DRYSS TIVA**  
Dr. Rosalva Lombardi

**ROSALVA LOMBARDI**  
Dr. Patricia Alvarez

**DR. PATRICIA ALVAREZ**  
Dr. Fabiana García

**DR. FABIANA GARCÍA**

**COMITÉ ORGANIZADOR**  
Dr. Pablo Sosa  
Dr. Carlos Lanzetta  
Dr. Nicolás Lanzetta  
Dr. Carolina Belmont

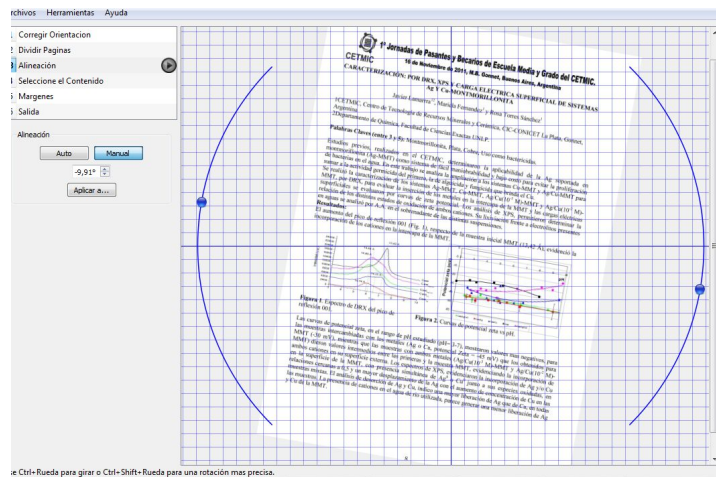
**CETMIC**

3

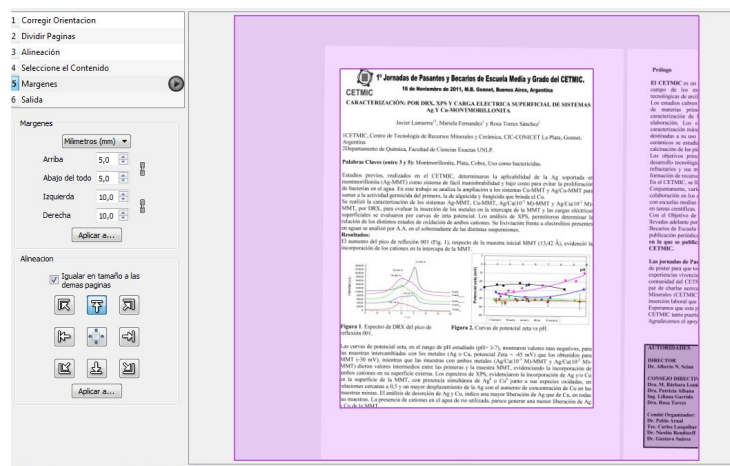
Use the mouse wheel or +/- to zoom. When zoomed, dragging is possible.

# Scan Tailor

También es posible corregir la orientación de las páginas, seleccionar los márgenes y la sección del contenido para realizar OCR y por último ajustar el color y contraste de las páginas.



Ctrl+Rueda para girar o Ctrl+Shift+Rueda para una rotación más precisa.



# Descarga masiva de archivos PDF

Muchas veces ocurre que debemos descargar archivos PDF de manera masiva, por ejemplo, cuando vamos a cargar en el repositorio un volumen de una revista que se encuentra en una página web. Existen numerosas herramientas, en SEDICI utilizamos una extensión de Chrome llamada GetThemAll

Video Downloader GetThemAll  
otredato por PVD Inc.  
★★★★☆ (8610) | Productividad | 1.439.271 usuarios

DESCRIPCIÓN GENERAL OPINIONES AYUDA RELACIONADOS

Universal video and file downloader!

GetThemAll (GTA) is a whole new dimension in download management software. Tired of searching for millions of files on websites and downloading them one by one? You will no longer have to! GetThemAll will find all files available on the page you view by itself - all you need to do is choose exactly what you want to download.

Can't download videos from YouTube.

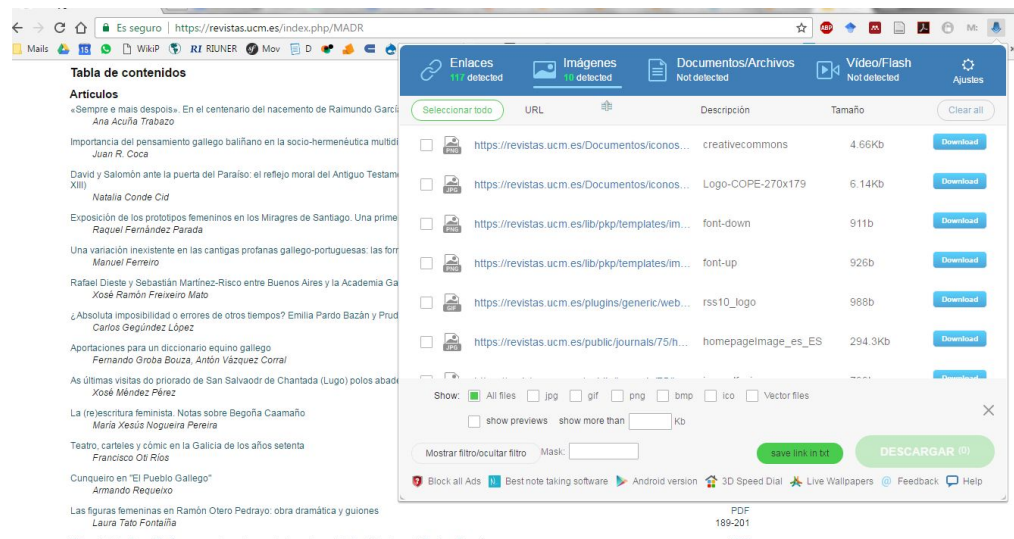
[Privacy policy]  
This extension collects information through Google Analytics.

Notificar uso inadecuado

**Información adicional**  
Versión: 25.1.1  
Última actualización: 13 de febrero de 2017  
Tamaño: 1.22MiB  
Idiomas: Ver todos (11)

# Descarga masiva de archivos PDF

La extensión GetThemAll detecta automáticamente los archivos pdf de una página web y permite la descarga en breves pasos. Ejemplo: [Journal of Computer Science and Technology](#).



The screenshot shows a web browser window with the URL <https://revistas.ucm.es/index.php/MADR>. The page displays a table of contents with various articles. Overlaid on the page is the GetThemAll extension interface, which has detected 177 links, 10 images, and 0 documents/archives. The interface includes a table with columns for URL, Description, and Tamaño (Size), and a 'DESCARGAR (0)' button.

Seleccionar todo	URL	Descripción	Tamaño	Clear all
<input type="checkbox"/>	<a href="https://revistas.ucm.es/Documentos/iconos...">https://revistas.ucm.es/Documentos/iconos...</a>	creativecommons	4.66Kb	Download
<input type="checkbox"/>	<a href="https://revistas.ucm.es/Documentos/iconos...">https://revistas.ucm.es/Documentos/iconos...</a>	Logo-COPE-270x179	6.14Kb	Download
<input type="checkbox"/>	<a href="https://revistas.ucm.es/lib/pkp/templates/im...">https://revistas.ucm.es/lib/pkp/templates/im...</a>	font-down	911b	Download
<input type="checkbox"/>	<a href="https://revistas.ucm.es/lib/pkp/templates/im...">https://revistas.ucm.es/lib/pkp/templates/im...</a>	font-up	926b	Download
<input type="checkbox"/>	<a href="https://revistas.ucm.es/plugins/generic/web...">https://revistas.ucm.es/plugins/generic/web...</a>	rss10_logo	988b	Download
<input type="checkbox"/>	<a href="https://revistas.ucm.es/public/journals/75/n...">https://revistas.ucm.es/public/journals/75/n...</a>	homepagelimage_es_ES	294.3Kb	Download

# Normalización de Texto

Es muy común que el título de un artículo, su resumen y otros metadatos sean copiados directamente desde el archivo PDF. Esta acción suele presentar problemas de diferente tipo que alteran el texto a pegar, por ejemplo: codificación inválida, exceso de saltos de línea y tabulaciones, textos en Mayúsculas, caracteres especiales que no se veían en el PDF, entre otros.

⇒ Normalizador de Texto. <http://sedici.unlp.edu.ar/norm.php>

Es una aplicación PHP que a partir de una mínima configuración revisa y corrige el texto ingresado. Por ejemplo:

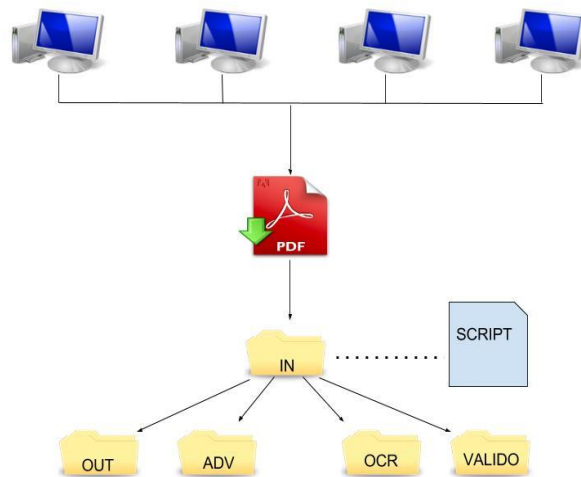
- multiplicidad indeseada de espacios y saltos de línea
- detección de saltos de línea indebidos
- Normalización de caracteres (UNICODE NORM)



## 3-height

Para trabajar con Documentos PDF se generó un script que utiliza un programa estilo **servidor/cliente**, encargado de:

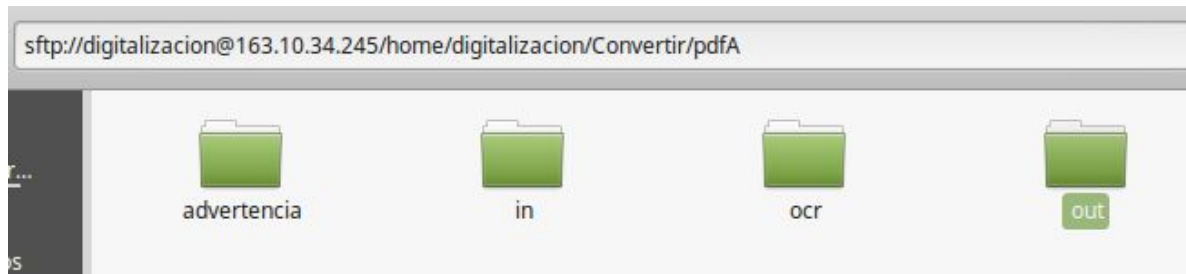
- Detección de archivos
- Análisis
- Conversión
- Verificación



## 3-height

Simplemente tenemos una carpeta compartida con el nombre PDFA que consta de 4 directorios donde los administradores podrán transformar los archivos PDF en PDFA. Los directorios son:

- Una carpeta “in” para ingresar los archivos a procesar
- Una Carpeta “out” donde se depositarán los archivos resultantes.
- Y dos carpetas destinadas a diferentes tipos de errores llamadas “advertencia” y “ocr”



# Generación de PDF/A - 3-Height

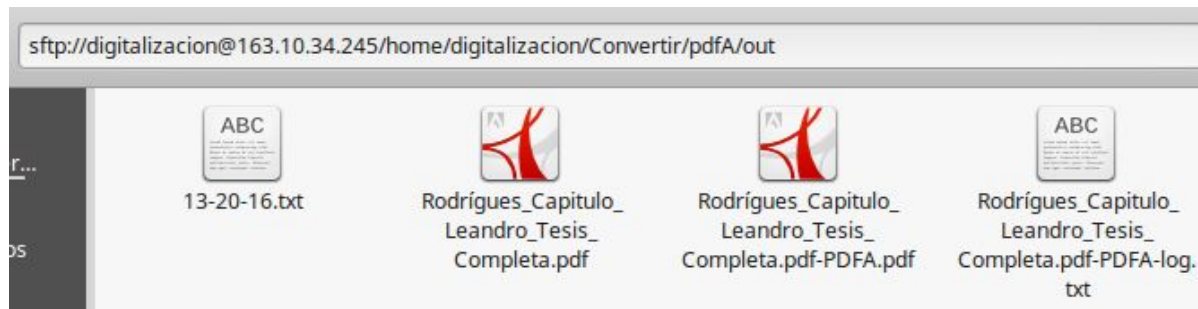
El 3 height analiza el pdf y elige en qué versión va a convertirlo. Si la conversión sale bien, en la carpeta out tendremos los siguientes archivos:

El archivo con la fecha 13-20-16.txt presenta el log de la ejecución del script..

El archivo pdf original.

El archivo convertido con la terminación: -PDFA.pdf

El último archivo txt da más detalles de la conversión del archivo original

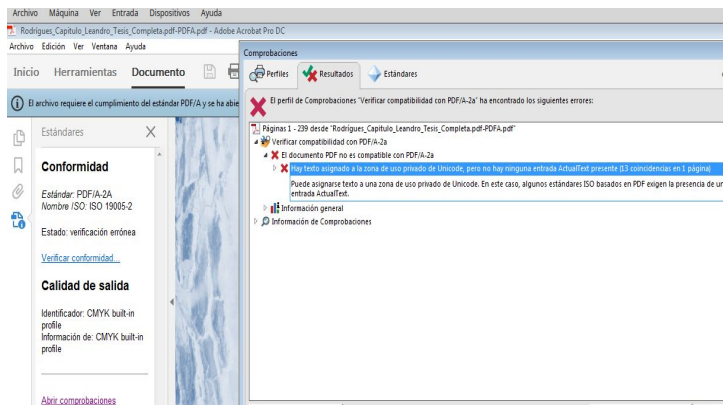


# Validación de PDF/A - Acrobat DC

Una vez obtenido el pdf/A de 3-Height es necesario validarlo también en Acrobat DC. Si la verificación es errónea dependiendo el caso de error podemos arreglarlo desde el mismo Acrobat. Por ejemplo: cuando un archivo no pasa la verificación porque el texto no es unicode en todo el pdf. Generalmente este problema se soluciona transformando el archivo en la versión de pdfA llamada pdf/A-2u.

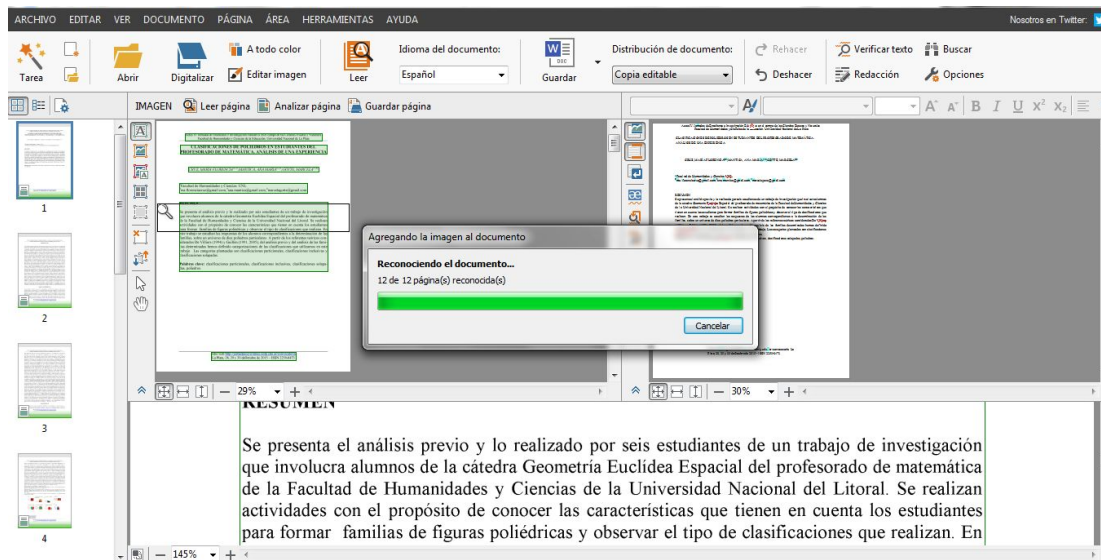


Adobe Acrobat DC



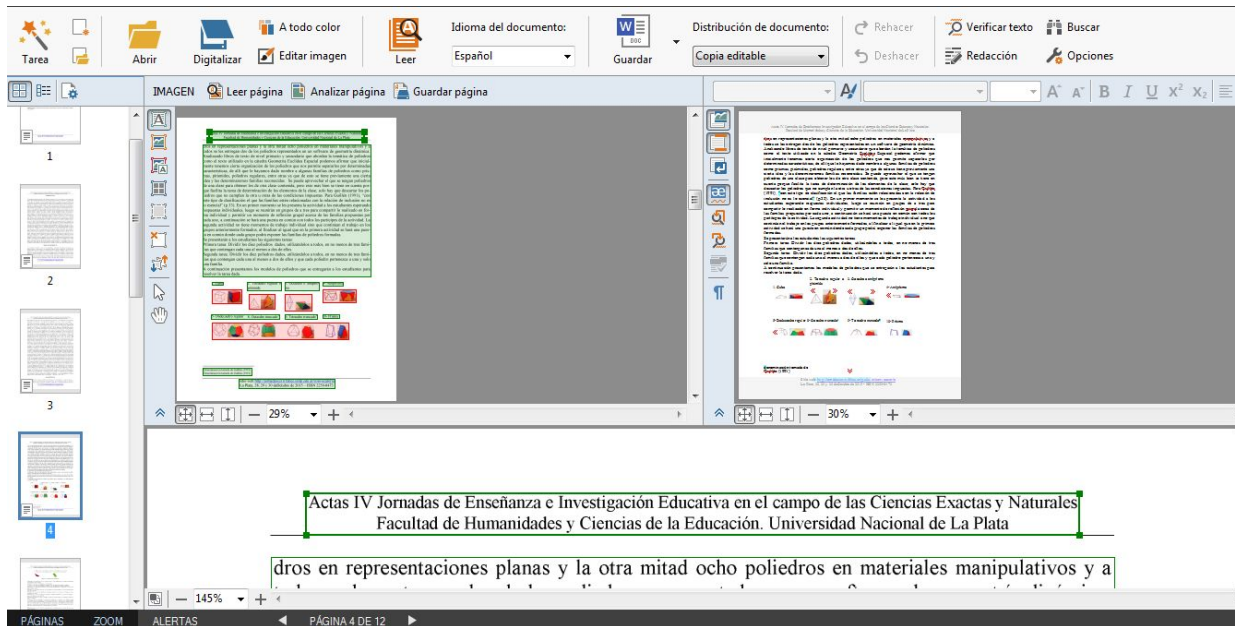
# Generación de OCR con ABBYY Fine Reader

ABBYY Fine Reader es un software de OCR que permite trabajar y editar pdfs de manera rápida y confiable.



# ABBYY Fine Reader

El software permite seleccionar imágenes e indicar qué parte de la página debe ser reconocido y qué no. Es de los softwares con motor de reconocimiento óptico de caracteres más precisos del mercado.



# Digitalización

*Digitalización y preservación*

*Digitalización de documentos*

*Escáner DAL*

*Proceso de imágenes*

*Digitalización para dispositivos móviles*

*Formatos*

*PDF/A*



# Introducción a la Digitalización

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.

Una estrategia de preservación es la de adoptar estándares internacionales, es decir, apoyarse en la afirmación de que los estándares internacionales son relativamente estables en el tiempo.

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.



# Generación de PDF/A

Un formato de preservación para documentos de texto es el estándar PDF/A, descrito en las normas ISO 19005-(1-2-3).

Este formato está basado en el estándar PDF 1.4, al que le incorpora algunos requerimientos adicionales, por ejemplo:

- Especificaciones sobre los metadatos y la estructura del archivo.
- La paleta de colores (incluyendo escala de grises y blanco/negro) no deben ser representados en un espacio de color de dispositivo (DeviceRGB, DeviceCMYK, DeviceGray).
- Las fuentes usadas en texto visibles deben estar embebidas (incluidas dentro del archivo).

Uno de los propósitos de los requerimientos del estándar PDF/A es de proveer soporte para personas con capacidades diferentes, por ejemplo, incorporando la información requerida y necesaria para aplicaciones que hagan el pasaje de texto a voz.



# Digitalizaciones en SEDICI

- 1) Documentos de papel físicos (libros, revistas, tesis, informes, etc.)
- 2) Documentos .doc, .docx, .odt, etc.
- 3) Documentos PDF



# 1) Digitalización de documentos

- Escáner HP scanjet 7500 / Fujitsu fi-7160

Softwares de escaneos: HP Smart Document Scan Software 3 - PaperStream

- Escáner DAL

Captura de imágenes

Rotación de la imagen

Renombrado

Combinación

Proceso de imágenes mediante ScanTailor/Photoshop

Impresión en PDF, OCR en Abbyy FineReader 12



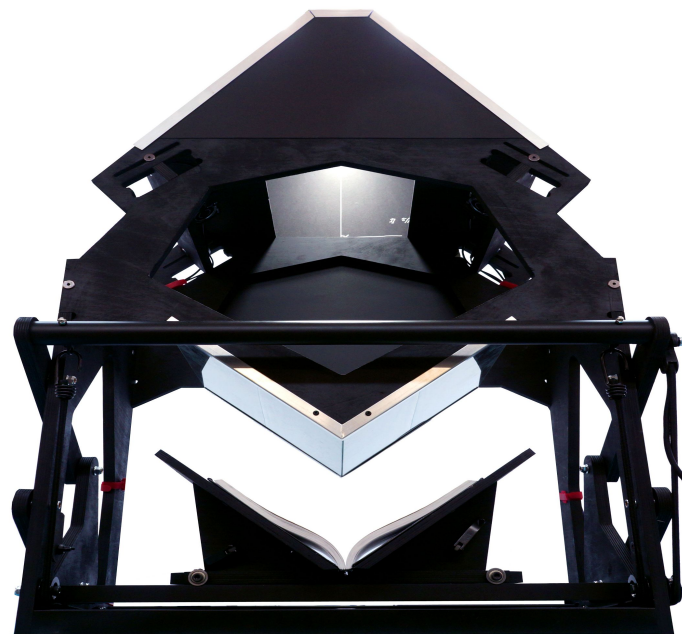
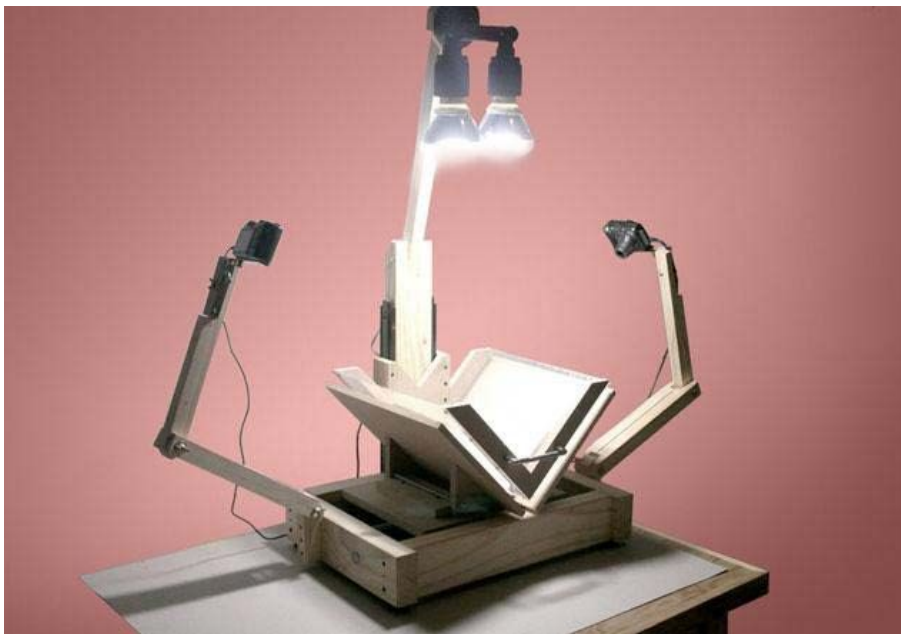
# DAL

Derecho a Leer (DaL) (Facultad de Filosofía y Letras - UBA). Es un espacio de difusión de las políticas a favor de la lectura. Inspirados en la experiencia de *Do It Yourself* desarrollaron un escáner para libros de forma artesanal y muy económica, utilizando madera y cámaras digitales estándar. El objetivo es tomar una foto con la mayor precisión posible de cada hoja del libro.

Ejemplo de digitalización realizada en SEDICI:  
Academia Nacional de Agronomía y Veterinaria (ANAV).  
Tomo I 1932- 1934



# Escáneres DAL



# Captura de imágenes

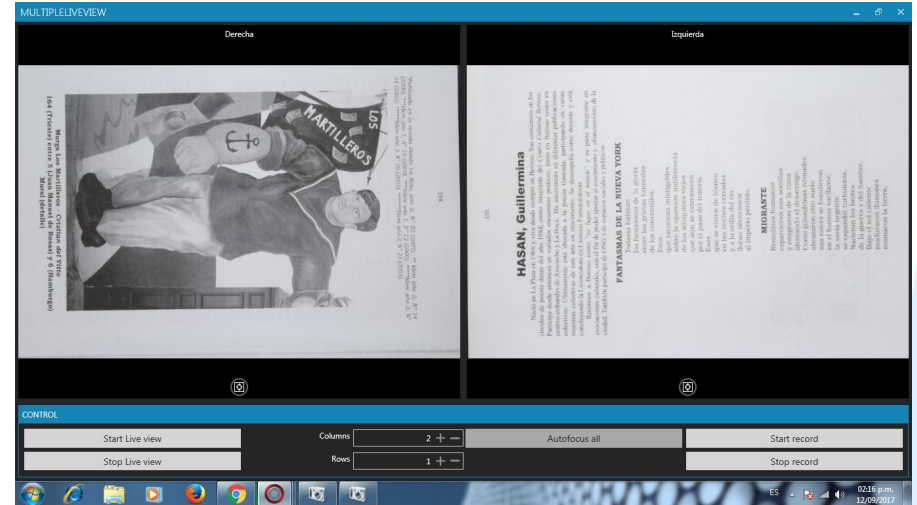
Maquinas Nikon D5300

Software *digiCamControl*

Mecanismo de disparador manual

Configuración a través de software de las cámaras

Previsualización en tiempo real



# Pre-procesamiento de imágenes

Luego de la descarga de las fotos al PC, se corre un script que rota, renombra y combina las imágenes sacadas de las cámaras.

```
prebi@Digitalización ~ $ ./editar  
MiNuevoProyecto
```

Procesamiento:

Se crea un proyecto en Scan Tailor donde a cada imagen se le hace un proceso de:

- alineación
- selección de contenido
- centralización e igualar tamaño
- blanquear fondo y borrar manchas.

ScanTailor crea imágenes en formato .TIFF Luego de generados los TIFF se crea un nuevo proyecto de AbbyyFineReader, donde se selecciona el texto e imágenes para crear un PDF/A-1a.

El PDF es verificado y optimizado para la vista rápida en web con Acrobat DC.



# HP 7500/ Fujitsu fi-7160

Se necesitan las hojas del documento sueltas, sin bordes salientes

Se pueden escanear a una cara o dos

Escaneos en blanco y negro: 600ppp

Escaneos en color/ escala de grises:

400ppp

TIFF con compresión sin pérdida (LZW).

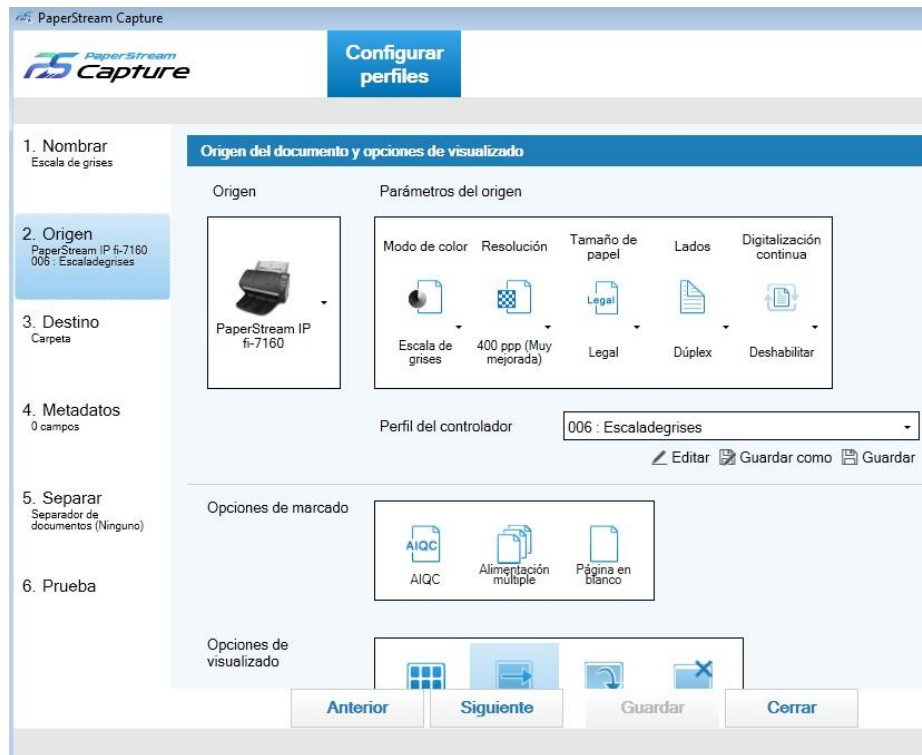
Máximo de hojas por paquete: 100

Software de escaneo: HP Smart

Document Scan Software 3 /

PaperStream

Formato de salida del escaneo: .TIFF



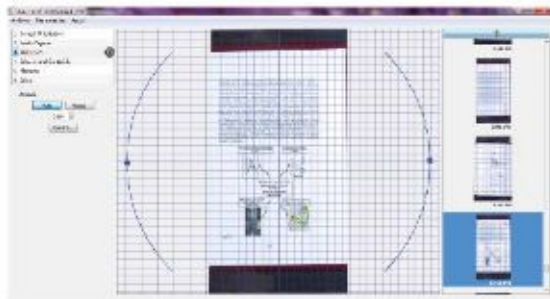
# Clasificación del material según relevancia

Dependiendo de la utilidad y el interés del material, los procesos de edición de imagen tienen mayor o menor automatización y revisión posterior. El material de alta relevancia (copias únicas por ejemplo) requieren un proceso de revisión de la edición de imagen y del OCR página por página. Materiales de relevancia media requieren una revisión detallada de la tapa y los índices y una revisión general del resto. En cambio, los materiales de digitalización rápida requieren una revisión general y un proceso casi totalmente automatizado.

# Programas utilizados

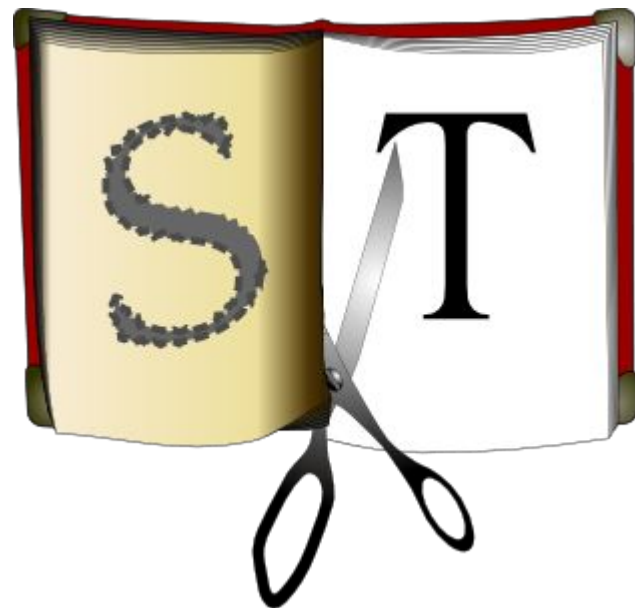


Scantailor



## Edición de imagen


Luego de obtener las imágenes escaneadas, se editan en Photoshop o ScanTailor. De esta manera se puede corregir la orientación, dividir y alinear las páginas, seleccionar el contenido, limpiar los márgenes, eliminar las manchas y modificar el contraste.



Archivo Herramientas Ayuda

- 1 Corregir orientación
- 2 Dividir páginas
- 3 Alineación
- 4 Seleccione contenido
- 5 Márgenes


**Rotar**



Restablecer

**Extensión**


Aplicar a...



Archivo Herramientas Ayuda

- 1 Corregir orientación
- 2 Dividir páginas
- 3 Alineación
- 4 Seleccione contenido
- 5 Márgenes


**Rotar**



Restablecer

**Extensión**

Aplicar a...




Archivo Herramientas Ayuda

- 1 Corregir orientación
- 2 **Dividir páginas**
- 3 Alineación
- 4 Seleccione contenido
- 5 Márgenes
- 6 Salida

### Diseño de página



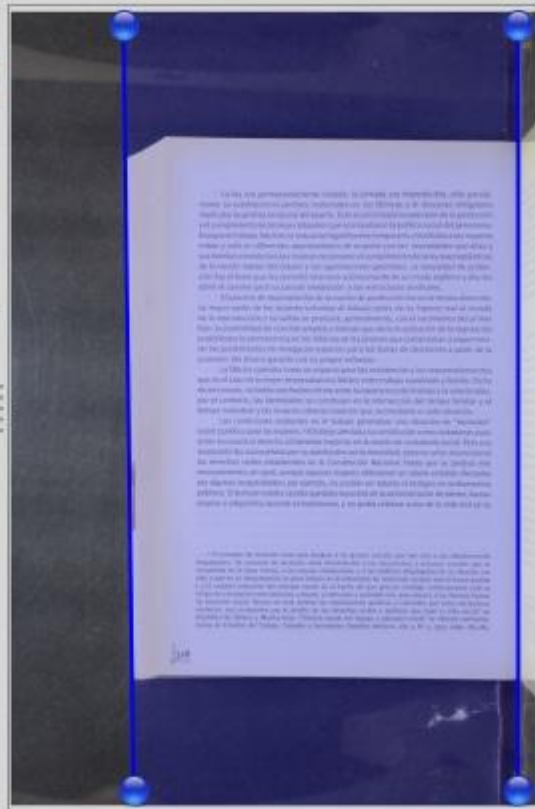
Configurar manualmente

Cambiar...

### Línea divisoria

Auto

Manual



316.jpg

317.jpg

Orden natural



Archivo Herramientas Ayuda

1 Corregir orientación

2 Dividir páginas

3 Alineación

4 Seleccione contenido

5 Márgenes

6 Salida

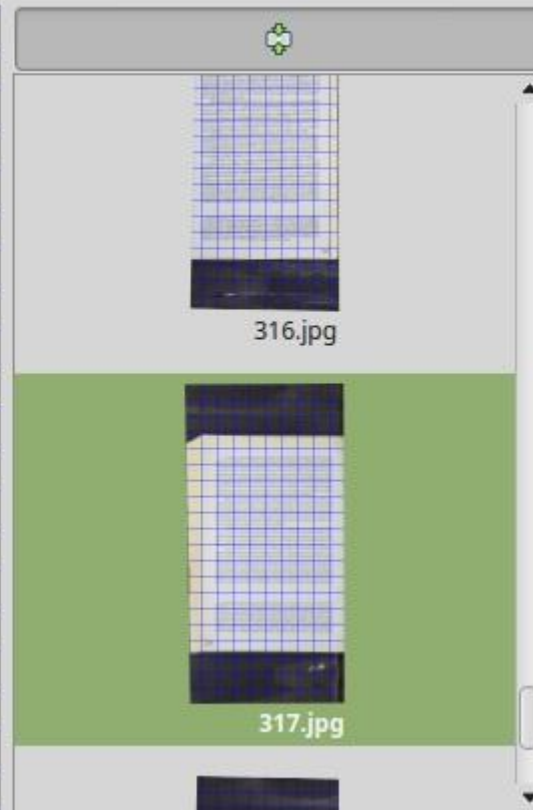
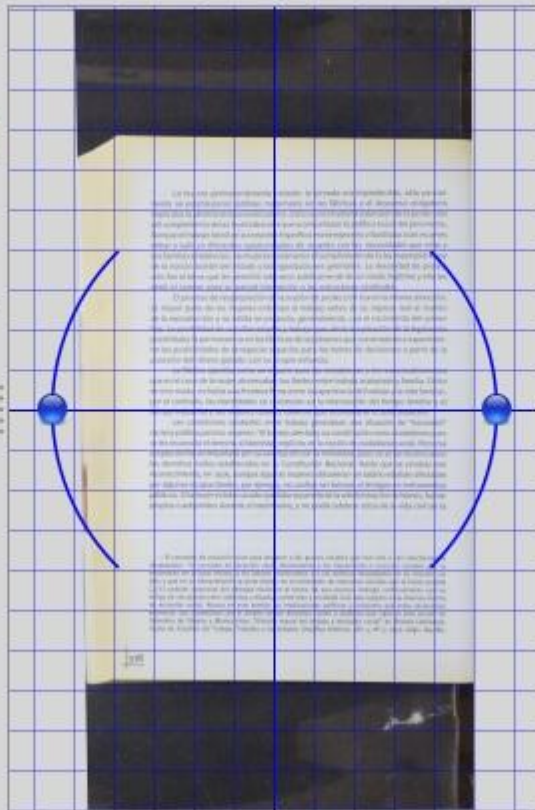
## Alineación

Auto

Manual

0,19°

Aplicar a...





Archivo Herramientas Ayuda

3 Alineación  
4 Seleccione contenido  
5 Márgenes

**Márgenes**

Milímetros (mm)

Arriba 10,0

Abajo 5,0

Izquierda 10,0

Derecha 10,0

Aplicar a...

**Alineación**

Igualar en tamaño con las demás páginas

Aplicar a...

315.jpg

316.jpg

317.jpg

Orden natural



Archivo Herramientas Ayuda

4 Seleccione contenido  
5 Márgenes  
6 Salida

**Resolución de salida (DPI)**  
400  
Cambiar...

**Modo**  
Combinado

0  
Más fino Más grueso

Aplicar a...

**Antideformación**  
Apagado  
Cambiar...

**Eliminar manchas**  
Aplicar a...

La ley era permanentemente violada: la jornada era impredecible, sólo parcialmente se establecieron jardines maternales en las fábricas y el descanso obligatorio implicaba la pérdida temporal del salario. Esto ocurrió hasta la extensión de la protección y el cumplimiento de las leyes laborales que acompañaron la política social del peronismo. Aunque el trabajo fabril en la industria frigorífica era temporario y facilitaba a las mujeres entrar y salir en diferentes oportunidades de acuerdo con las *necesidades* que ellas y sus familias establecían, las mujeres reclamaron el cumplimiento de la ley reapropiándose de la noción tutelar del Estado y las organizaciones gremiales. La necesidad de protección fue el tema que les permitió intervenir públicamente de un modo legítimo y ello les abrió el camino para su parcial integración a las estructuras sindicales.

El proceso de reapropiación de la noción de *protección* fue en la misma dirección. La mayor parte de las mujeres entraban al trabajo antes de su ingreso real al mundo de la reproducción y su salida se producía, generalmente, con el nacimiento del primer hijo. La posibilidad de conciliar empleo y trabajo que abrió la aplicación de la legislación posibilitaba la permanencia en las fábricas de las jóvenes que comenzaban a experimentar las posibilidades de renegociar espacios para las tomas de decisiones a partir de la posesión del dinero ganado con su propio esfuerzo.

La fábrica operaba como un espacio para las resistencias y los reacomodamientos que en el caso de la mujer atavesaban los límites entre trabajo asalariado y familia. Dicho de otro modo, no había una frontera firme entre la experiencia del trabajo y la vida familiar, por el contrario, las identidades se construían en la intersección del tiempo familiar y el tiempo industrial y las mujeres obreras tuvieron que acomodarse a cada situación.

Las condiciones existentes en el trabajo generaban una situación de "exclusión" social y política para las mujeres.<sup>4</sup> El trabajo afectaba su constitución como ciudadanas pues se les reconocía el derecho al bienestar implícito en la noción de ciudadanía social. Pero esa aceptación iba acompañada por su asimilación con la minoridad, pues no se les reconocieron los derechos civiles establecidos en la Constitución Nacional. Hasta que se produjo ese reconocimiento, en 1926, aunque algunas mujeres obtuvieran un salario estaban afectadas por algunas incapacidades, por ejemplo, no podían ser tutoras ni testigos en instrumentos públicos. Si la mujer estaba casada quedaba separada de la administración de bienes, fueran propios o adquiridos durante el matrimonio, y no podía celebrar actos de la vida civil sin la

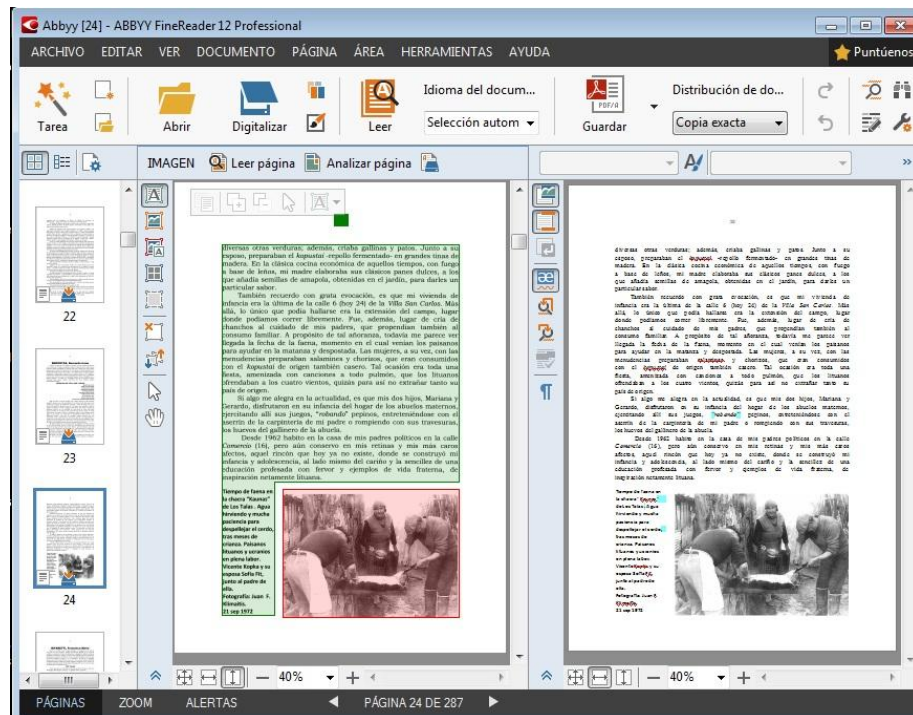
<sup>4</sup> El concepto de exclusión sirve para designar a los grupos sociales que han sido o son selectivamente desplazados. "El concepto de exclusión alude directamente a los mecanismos o procesos sociales que se encuentran en la base misma, a los actores involucrados y a las políticas desplegadas en su relación con ella; y que en su interpretación se pone énfasis en el entramado de relaciones sociales que la hacen posible [...] El carácter relacional del enfoque reside en el hecho de que procura inteligir continuamente cuál es el tipo de vinculación entre individuo y estado, y entre éste y sociedad civil, que subyace a las diversas formas de exclusión social. Abarca en este sentido las implicaciones políticas y culturales que estas exclusiones encierran, sus conexiones con el ámbito de los derechos civiles y políticos que rigen la vida social" en Oriandina de Oliveira y Marina Ariza: "División sexual del trabajo y exclusión social" en *Revista Latinoamericana de Estudios del Trabajo, Trabajo e Sociedades: Desafíos teóricos*, año 3, N° 5, 1997, págs. 184-185.

318



# Reconocimiento óptico de caracteres

Luego de editar las imágenes se realiza el OCR con el Abbyy FineReader. En esta etapa del proceso se selecciona el contenido según sea texto, imagen o cuadro. Luego se revisa la corrección del OCR y se generan los archivos PDF/A.



# Compresión de PDF

The image shows two overlapping dialog boxes from Adobe Acrobat. The background dialog is 'Opciones' (Options) with the 'Avanzadas' (Advanced) tab selected. The foreground dialog is 'Configuración personalizada' (Customized Configuration) for PDF/A export.

**Opciones (Advanced Tab):**

- Formato: PDF/A
- Tamaño de papel predeterminado: Automático
- Modo de guardado: Texto debajo de imagen de página
- Conservar texto y colores de fondo:
- Conservar encabezados y pies de página:
- Comprimir las imágenes utilizando MRC:
- Aplicar ABBYY Precise Scan para suavizar los caracteres de las imágenes:
- Configuración de imágenes: Personalizado...  Conservar imágenes

**Configuración personalizada:**

- Resolución:  Reducir la resolución de imagen en función de la salida de destino
  - Impresión, pantallas de alta calidad (300 ppp)
  - Lectura en pantalla, web (150 ppp)
  - Compartir (96 ppp)
- Control del color: Automático
- Calidad: Pérdida de calidad permitida
- Imágenes en escala de grises y en color: JPEG, ZIP, LZW, J2K
- Imágenes en blanco y negro: CCITT, JBIG2
- Calidad (45%): [Slider]
- Buttons: Restablecer, Aceptar, Cancelar



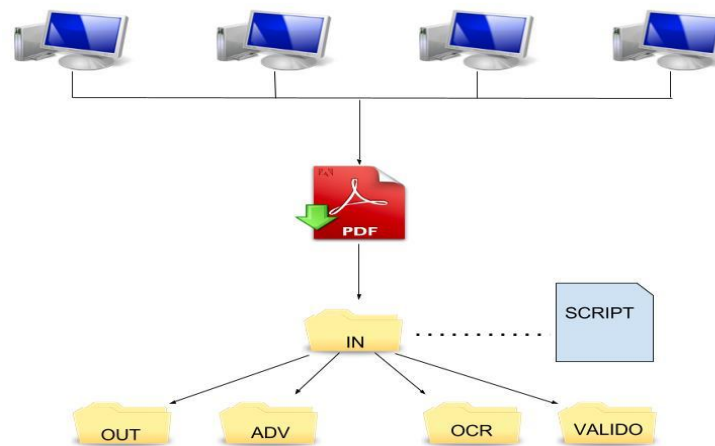
# Conversión a PDF/A

2) Conversión de formatos de texto (.doc, odt, etc).

Se guarda el documento como “PDF/A” desde el mismo word/libreOffice, etc.

3) Documentos PDF. Se utiliza un programa estilo servidor/cliente, encargado de:

- Detección de archivos
- Análisis
- Conversión
- Verificación

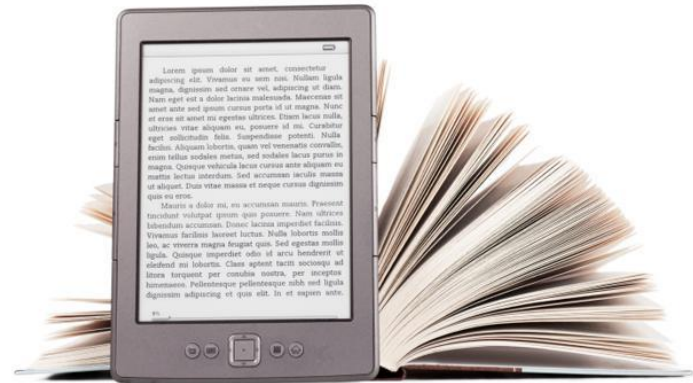


# Digitalización para dispositivos móviles

## MOBI

Es un formato para archivos de libro electrónico (e-book) creado por Mobipocket S. A. En el formato de libro digital MOBI se marca el contenido, pero no se delimita su formato, que se adapta a los diferentes tamaños de las pantallas de los múltiples lectores de libros electrónicos del mercado.

El formato MOBI está basado en las especificaciones Open eBook.

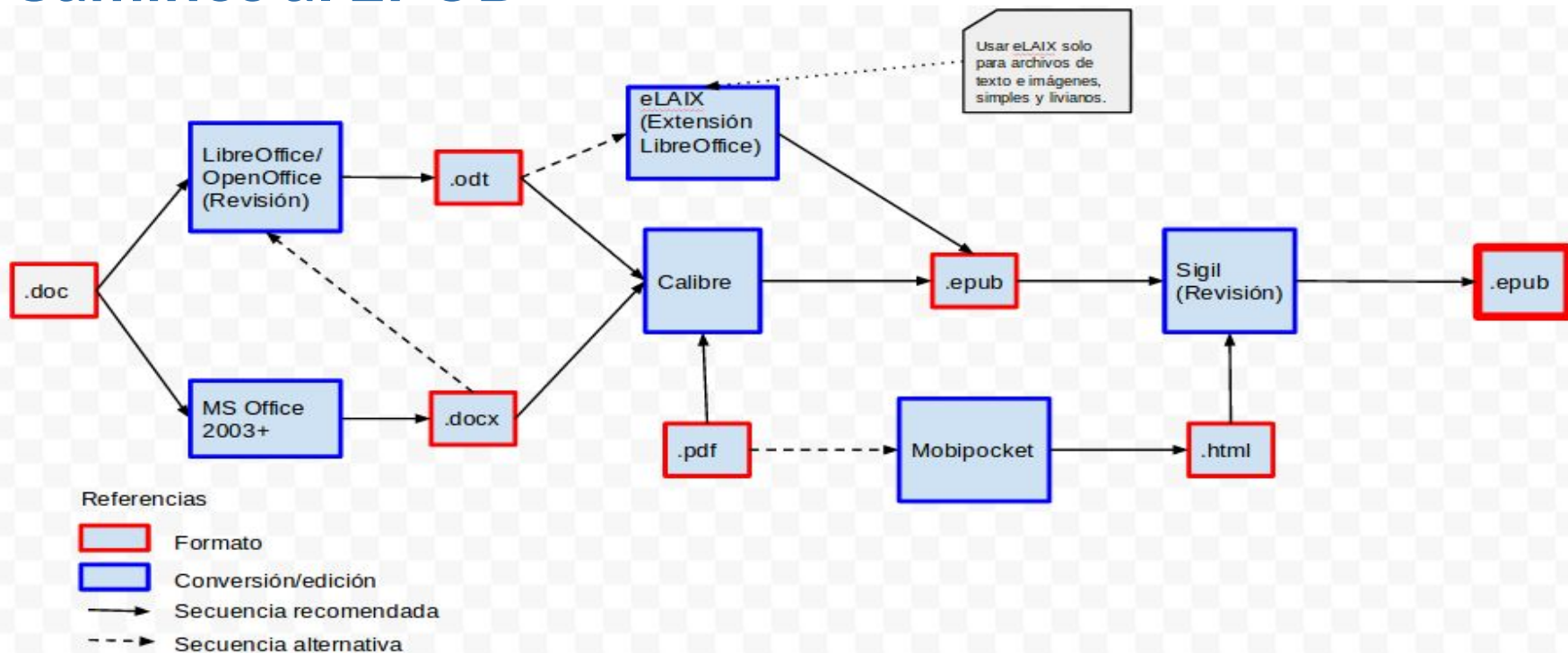


# Formatos EPUB

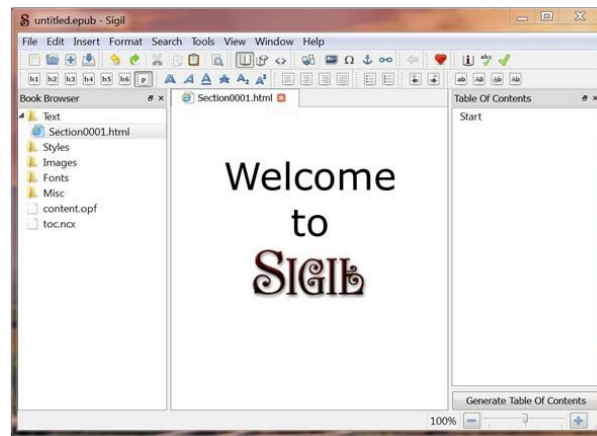
**EPUB:** El formato se diseñó para adaptarse a distintos tamaños de letra y pantalla. También se puede cambiar el tipo de letra. En el formato de texto para libro digital ePub, se marca el contenido, pero no se delimita su formato, que se adapta a las diferentes pantallas y tamaños de pantalla de los múltiples dispositivos. El formato fue creado por International Digital Publishing Forum (IDPF), como un formato específico para visualizar libros.

**EPUB3:** El formato, además de las funcionalidades habituales del EPUB ofrece salida de audio en la funcionalidad de texto-a-audio, características de voz, normas de pronunciación, especificación de fallbacks (conversión de un vídeo en texto, para discapacitados).

# Caminos al EPUB



# Programas utilizados en el proceso



# Audiolibros

Se trata de grabaciones habladas (o leídas automáticamente por programas informáticos) de libros.

Software usado para convertir texto a voz o archivos de audio: Natural reader

Formato de salida: .MP3 o .WAV



NaturalReader is an easy-to-use software that converts written text into natural sounding spoken words giving educators, professionals and students a powerful aid and learning tool. NaturalReader can assist ESL students (English as a Second Language) and people with learning difficulties such as dyslexia.

