

ESCUELA COMPLUTENSE LATINOAMERICANA

*preservación
digital de
documentos*
archivos, bibliotecas y museos

Título: Preservación Digital

subtítulo: metadatos y preservación

Dra. Marisa R. De Giusti

Universidad Nacional de La Plata

marisa.degiusti@sedici.unlp.edu.ar

Objetivos específicos de esta clase

Dar una visión general de las tipologías de metadatos para llegar a abordar los metadatos de preservación y el estándar PREMIS de metadatos de preservación, aunque no lo suficiente para implementar PREMIS, sí dar una idea de todo lo que incluye PREMIS. Para muchos será suficiente. Para quienes necesiten dominar el *Diccionario de datos PREMIS de metadatos de preservación*, esta clase puede servir como una sencilla introducción. Temas: Metadatos, Preservación, Metadatos de preservación, PREMIS, PREMIS DD, PREMIS XML cruces con METS.

CONTENIDOS de esta clase

1. Metadatos. Tipologías. Esquemas.
2. PREMIS: entidades, diccionario de datos, PREMIS XML, METS...
3. Herramientas para automatizar los procesos de preservación: perfilado, migración, ...Plan de preservación...

Sección 1



Metadatos

Son la información que se aplica sobre un recurso para facilitar su organización, son datos que describen otros datos, información estructurada para describir, explicar, localizar o facilitar la obtención, uso o administración de un recurso de información.

La norma ISO 15489-1 del 2001 define los metadatos, en el contexto de la gestión de documentos, como: *"datos que describen el contexto, contenido y estructura de los documentos, así como su gestión a lo largo del tiempo (...) Como tales, los metadatos son información estructurada o semiestructurada que posibilita la creación, registro, clasificación, acceso, conservación y disposición de los documentos a lo largo del tiempo y dentro de un mismo dominio o dominios diferentes."*

Papel de los metadatos

- Recuperación de la información
- Administración de documentos
- Gestión de derechos, autoría y propiedad intelectual
- Estado de archivo
- Control y descripción de procesos
- Seguridad y autenticación
- Valoración de contenidos
- Preservación y conservación
- Visibilidad de la información
- Actualización de la información

Tipo de metadatos

- Administrativos
- Estructurales
- Descriptivos
- Técnicos
- De uso
- De preservación

Tipo	Uso	Ejemplos
Administrativo	Usados en la identificación, gestión y administración de recursos de información	Adquisición de información Derechos y reproducción Requerimientos legales para el acceso Localización de información Criterios de selección para la digitalización Control de la versión
Descriptivo	Utilizados para representar recursos de información	Registros catalográficos Proporcionar ayuda en la búsqueda Índices especializados Hiperenlazar relaciones entre recursos Anotaciones de los usuarios
Preservación	Para salvaguardar los recursos de información	Informar sobre las condiciones de uso de los recursos físicos Informar sobre las acciones llevadas a cabo para preservar versiones físicas y digitales de recursos
Técnico	Relativos a cómo funcionan los sistemas o el comportamiento de los metadatos	Documentación de hardware y software Digitalización de la información (formato, ratio de compresión...) Autenticación y datos de seguridad (encriptación, passwords, etc.) Control de tiempo de respuesta de sistemas
Uso	Relativos al nivel y tipo de uso que se hace con los recursos informativos	Información sobre versiones Reutilización del contenido del recurso

Metadatos – Esquemas

- [Dublin Core Metadata Initiative \(DC\)](#)
- [Networked Digital Library of Theses and Dissertations \(ETD\)](#)
- [Metadata Objects Description Schema \(MODS\)](#)
- [LUCIS \(UNCUYO\)](#)
- Categories for the Description of Works of Art (CDWA)
- Cataloguing Cultural Objects (CCO)

Metadatos – Dublin Core

DC es un modelo de metadatos elaborado por la [DCMI](#) (Dublin Core Metadata Initiative). Actualmente es el esquema de datos más utilizado. Consta de 15 elementos básicos.

Contenido	Propiedad	Instanciación
	Intelectual	
Title	Creator	Date
Subject	Publisher	Type
Descripción	Contributor	Format
Source	Rights	Identifier
Lenguaje		
Relation		
Coverage		

Metadatos – Dublin Core cualificado

El **Dublin Core cualificado** es una extensión del Dublin Core donde algunos de sus elementos son acompañados de un cualificador que los hace más restrictivos .

- Title

- dc.title
- dc.title.alternative

- Relation

- dc.isVersionOf
- dc.isPartOf

- Date

- dc.date.created
- dc.date.available

Networked Digital Library of Theses and Dissertations (ETD)

- Universidad de Utah State
- Versiones electrónicas de tesis o disertaciones
- Metadatos utilizados en SEDICI
- thesis.degree.grantor (Institución garante)
- thesis.degree.name (Grado alcanzado)

Metadata Objects Description Schema (MODS)

- Desarrollado por la Oficina de Desarrollo de Redes y Normas MARC de la Biblioteca del Congreso (LC)
- MODS es un conjunto de metadatos que tiene el mismo objetivo que DC, pero que busca superar los problemas y limitaciones detectados en la aplicación de DC. Para ello, toma como punto de partida el formato MARC, del que selecciona diferentes En SEDICI se utilizan:
 - **mods.location** (URL de acceso al catalogo de la biblioteca, o información de acceso al recurso físico)
 - **mods.origenInfo.place** (Institución donde se creo el documento)
 - **mods.recordInfo.recordContentSource** (Registra el nombre de la entidad y/o sitio web desde donde se obtuvo la información del registro)

Categories for the Description of Works of Art (CDWA)

Directrices para la documentación de los bienes patrimoniales desarrolladas por un grupo de trabajo dirigido por el Getty y la College Art Association of America (CAA),

Cataloguing Cultural Objects (CCO)

Manual para documentar arte, arquitectura y objetos culturales desarrollado y diseminado por el Visual Resources Association (VRA).

Metadatos - Esquema SEDICI

—SEDICI utiliza un formato de metadatos propio: dichos metadatos están traducidos al inglés y al portugués, y el sistema los transforma al formato [Dublin Core](#) para que sean accedidos por otros repositorios vía [OAI-PMH](#).

Ej.: [tesis](#)

dc.format.extent	110 p.	es
dc.language	es	es
dc.title	Familias y estrategias de crianza	es
dc.type	Tesis	es
sedici.title.subtitle	Prácticas de tránsito de niños y niñas por diversos grupos de crianza en sectores populares de Viedma	es
sedici.creator.person	Belmartino, María Belén 🇺🇦	es
sedici.subject.materias	Trabajo Social 🇺🇦	es
sedici.subject.decs	Familia 🇺🇦	es
sedici.subject.decs	Crianza del Niño 🇺🇦	es
sedici.subject.decs	Población Suburbana 🇺🇦	es
sedici.description.fulltext	true	es
mods.originInfo.place	Facultad de Trabajo Social 🇺🇦	es
sedici.subtype	Tesis de maestria	es
sedici.rights.license	Creative Commons Atribución-NoComercial-CompartirIgual 2.5 Argentina (CC BY-NC-SA 2.5)	
sedici.rights.uri	http://creativecommons.org/licenses/by-nc-sa/2.5/ar/	
sedici.contributor.director	Colangelo, María Adelaida 🇺🇦	es
sedici.contributor.juror	Colangelo, María Adelaida	es
sedici.contributor.juror	Fuentes, Pilar	es
sedici.contributor.juror	Velurtas, Marcela	es
thesis.degree.name	Magister en Trabajo Social 🇺🇦	es
thesis.degree.grantor	Facultad de Trabajo Social 🇺🇦	es
sedici.date.exposure	2012-06-04	

Sección 2



PREMIS

PREMIS son las siglas de “PREservation Metadata: Implementation Strategies” que es el nombre de un grupo 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 la v2 del 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”. Sitio web de PREMIS: <http://www.loc.gov/standards/premis/>

Preservación y metadatos de preservación

Se definió a la **preservación digital** como el conjunto de prácticas de naturaleza política, estratégica y acciones concretas, destinadas a asegurar el acceso a los objetos digitales a largo plazo.

En la clase precedente se ha hablado de metadatos y sus categorías de acuerdo a las funciones que cumplen.

Los ***metadatos de preservación*** soportan las actividades cuyo objetivo es asegurar la utilización a largo plazo de un recurso digital.

Ejemplos de actividades de preservación

- Un contenido debe almacenarse de forma segura, sin que nadie lo modifique inadvertidamente (por mala intención/o no). Una verificación posible es el checksum, el cual puede almacenarse en un metadato asociado.
- Un archivo debe almacenarse en un soporte actualizado, para acceder a los datos, los metadatos sirven para registrarlos.
- Los formatos de los archivos se vuelven viejos y hay que transformarlos para seguir usando los datos.
- Hay que mantener el origen y la cadena de custodia del archivo, para mantener la trazabilidad del mismo, los cambios permitidos y quiénes realizaron esos cambios.

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.

Las funciones de preservación pueden variar de un repositorio a otro, pero, incluyen acciones para mantener viables, recuperables, sin alteración y con los cambios documentados a los ODs.

El DD NO define todos los elementos posibles de los metadatos de preservación, sólo los que necesita la mayoría de los repositorios.

Diccionario de datos PREMIS

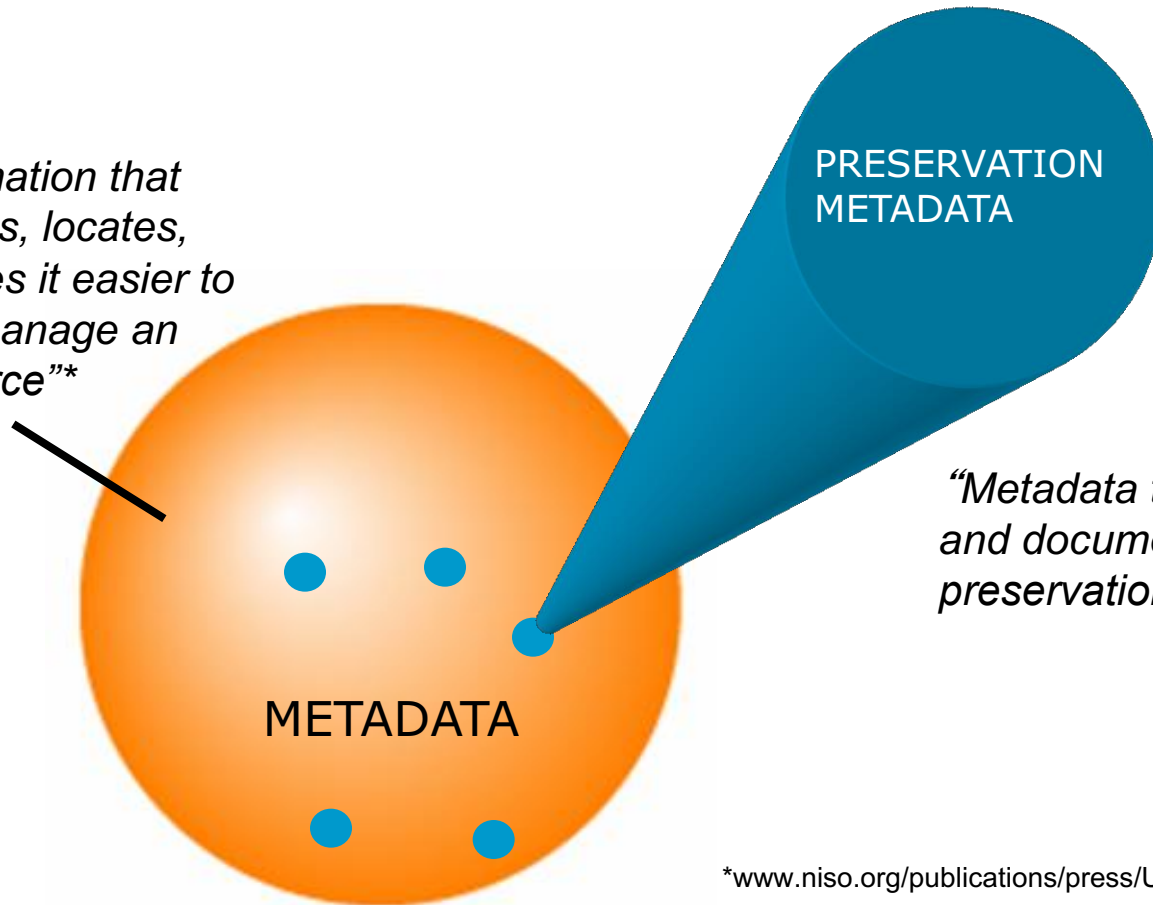
El diccionario NO define algunas categorías de metadatos:

- Los metadatos que pertenecen sólo a un formato de archivo, o una clase de formato como audio, video, etcétera.
- Los metadatos de políticas o reglas de negocios del RI.
- Los metadatos descriptivos.
- La información detallada de soporte o de hardware.
- Información adicional sobre agentes.
- Información de derechos que no afecte estrictamente a la preservación, por ejemplo una Licencia de libre distribución.

[Diccionario de Datos de Metadatos de Preservación: PREMIS Versión 2.0](#)

Preservation metadata

*“Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource”**



“Metadata that supports and documents the digital preservation process”

*www.niso.org/publications/press/UnderstandingMetadata.pdf

Cómo utilizar PREMIS

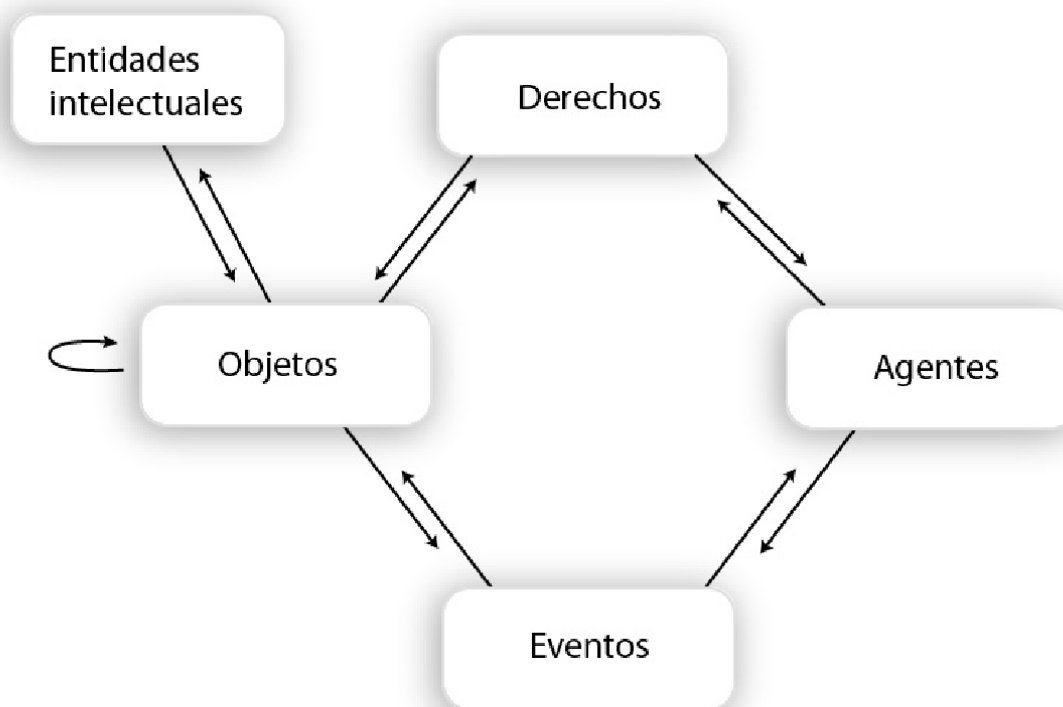
- Se utiliza fundamentalmente en la etapa de diseño de un repositorio, para su evaluación y para el intercambio de los paquetes de la información archivada entre repositorios de preservación.
- Quienes diseñan software para repositorios lo pueden utilizar como guía.
- Para un repositorio de preservación PREMIS es útil como un checklist.
- En los casos de exportación de la información almacenada para la ingesta en otro repositorio. Permite acordar. Especialmente PREMIS XML.

Modelo de Datos PREMIS

Las entidades que este modelo define se denominan:

- Entidades intelectuales
- Objetos
- Derechos
- Agentes
- Eventos

Modelo de datos PREMIS



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.

Entidad intelectual

El conjunto de contenidos que se considera una única unidad intelectual a efectos de gestión y descripción, por ejemplo, un libro, un mapa, una fotografía o una base de datos.

Una Entidad Intelectual puede comprender otras Entidades Intelectuales. Por ejemplo, un sitio web puede incluir una página web o una página web puede incluir una imagen. Una Entidad Intelectual puede tener una o más representaciones digitales.

Objetos

Unidades discretas de información en forma digital, que se clasifican en tres tipos: *archivo (file)*, *representación (representation)* y *cadenas de bits (bitstream)*. El objeto *archivo* es tal cual se entiende normalmente, p.e. un archivo PDF de un capítulo de un libro. El objeto *representación* es el conjunto de todos los archivos que se necesitan para representar la entidad ***Intelectual*** (p.e. un libro), incluyendo los metadatos estructurales. Los objetos *cadenas de bits* son subconjuntos de archivo con propiedades útiles a la preservación, en el ejemplo del libro el archivo JPEG de la tapa puede tener sus propios identificadores y metadatos (tal vez se podría querer cambiarla). 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.

Diccionario de Datos: Objeto

1.1 objectIdentifier (identificador del objeto) (O, R)

1.1.1 objectIdentifierType (tipo del identificador del objeto) (O, NR)

1.1.2 objectIdentifierValue (Valor del identificador del objeto) (O, NR)

1.2 objectCategory (categoría del objeto) (O, NR)

Diccionario de Datos: Objeto

1.3 preservationLevel (nivel de preservación) (NO, R) [representación, fichero]

1.3.1 preservationLevelValue (valor del nivel de preservación) (O, NR) [representación, fichero]

1.3.2 preservationLevelRole (función del nivel de preservación) (NO, NR)
[representación, fichero]

1.3.3 preservationLevelRationale (fundamentos del nivel de preservación) (NO, R) [representación, fichero]

1.3.4 preservationLevelDateAssigned (fecha asignada al nivel de preservación)
(NO, NR) [representación, fichero]

Diccionario de Datos: Objeto

1.4 significantProperties (propiedades significativas) (NO, R)

1.4.1 significantPropertiesType (tipo de propiedades significativas) (NO, NR)

1.4.2 significantPropertiesValue (valor de las propiedades significativas) (NO, NR)

1.4.3 significantPropertiesExtension (extensión de las propiedades significativas) (NO, R)

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.

Diccionario de Datos: Evento

2.1 eventIdentifier (identificador del acontecimiento) (O, NR)

2.1.1 eventIdentifierType (tipo de identificador del acontecimiento) (O, NR)

2.1.2 eventIdentifierValue (valor del identificador del acontecimiento) (O, NR)

2.2. eventType (tipo de acontecimiento) (O, NR)

2.3 eventDateTime (tiempo de la fecha del acontecimiento) (O, NR)

2.4 eventDetail (detalle del acontecimiento) (NO, NR)

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).

Diccionario de Datos: Agente

3.1 agentIdentifier (identificador del agente) (O, R)

3.1.1 agentIdentifierType (tipo de identificador del agente)

3.1.2 agentIdentifierValue (valor del identificador del agente)

3.2 agentName (nombre del agente) (NO, R)

3.3 agentType (tipo de agente) (NO, NR)

Derechos

Corresponden a la declaración de uno o varios derechos o permisos pertenecientes a un Objeto o Agente.

En un repositorio puede quedar registrada una gran variedad de información sobre los derechos, desde resúmenes de declaraciones de derecho hasta los permisos y derechos de agentes externos y objetos no depositados en el repositorio. Sin embargo, el corpus mínimo de información de derechos que un repositorio de preservación debería conocer, es el que hace referencia a los derechos o permisos con los que un repositorio tiene que cumplir y que corresponde a las acciones que afectan a los objetos dentro del repositorio. Esta información puede quedar garantizada por la ley de copyright, por una norma, o por un acuerdo de licencia con el titular de los derechos.

Diccionario de Datos: Agentes

4.1 rightsStatement (mención de derechos) (NO, R)

4.1.1 rightsStatementIdentifier (identificador de la mención de derechos) (O, NR)

4.1.1.1 rightsStatementIdentifierType (tipo de identificador de la mención de derechos) (O, NR)

4.1.1 rightsStatementIdentifierValue (valor del identificador de la mención de derechos) (O, NR)

4.1.2 rightsBasis (bases de los derechos) (O, NR)

Encabezado diapositiva

4.1.3 copyrightInformation (información sobre el copyright) (NO, NR)

4.1.3.1 copyrightStatus (estado del copyright) (O, NR)

4.1.3.2 copyrightJurisdiction (jurisdicción del copyright) (O, NR)

4.1.3.3 copyrightStatusDeterminationDate (determinación de la fecha del estado del copyright) (NO, NR)

4.1.3.4 copyrightNote (nota sobre el copyright) (NO, R)

4.1.4 licenseInformation (información sobre la licencia) (NO, NR)

4.1.4.1 licenseIdentifier (identificador de la licencia) (NO, NR)

PREMIS - XML

<xmlData>

<object xmlns=http://www.loc.gov/standards/premis/v1

xsi:schemaLocation="http://www.loc.gov/standards/premis/v1

http://www.loc.gov/standards/premis/v1/PR...">

<objectIdentifier>

<objectIdentifierType>URI</objectIdentifierType>

<objectIdentifierValue>http://www.nps.gov/rich/fall99.html</objectIdentifierValue>

Ejemplo

- <http://gredos.usal.es/jspui/handle/10366/115664>

Ejemplo

</objectIdentifier>

<objectCategory>file</objectCategory>

<objectCharacteristics>

<compositionLevel>0</compositionLevel>

<fixity>

<messageDigestAlgorithm>SHA-1</messageDigestAlgorithm>

<messageDigest>4ea7325ecef266792a03e5f82ce67762970e14a9</messageDigest>

Ejemplo

</fixity>

<size>31264</size>

<format>

<formatDesignation>

<formatName>text/html</formatName>

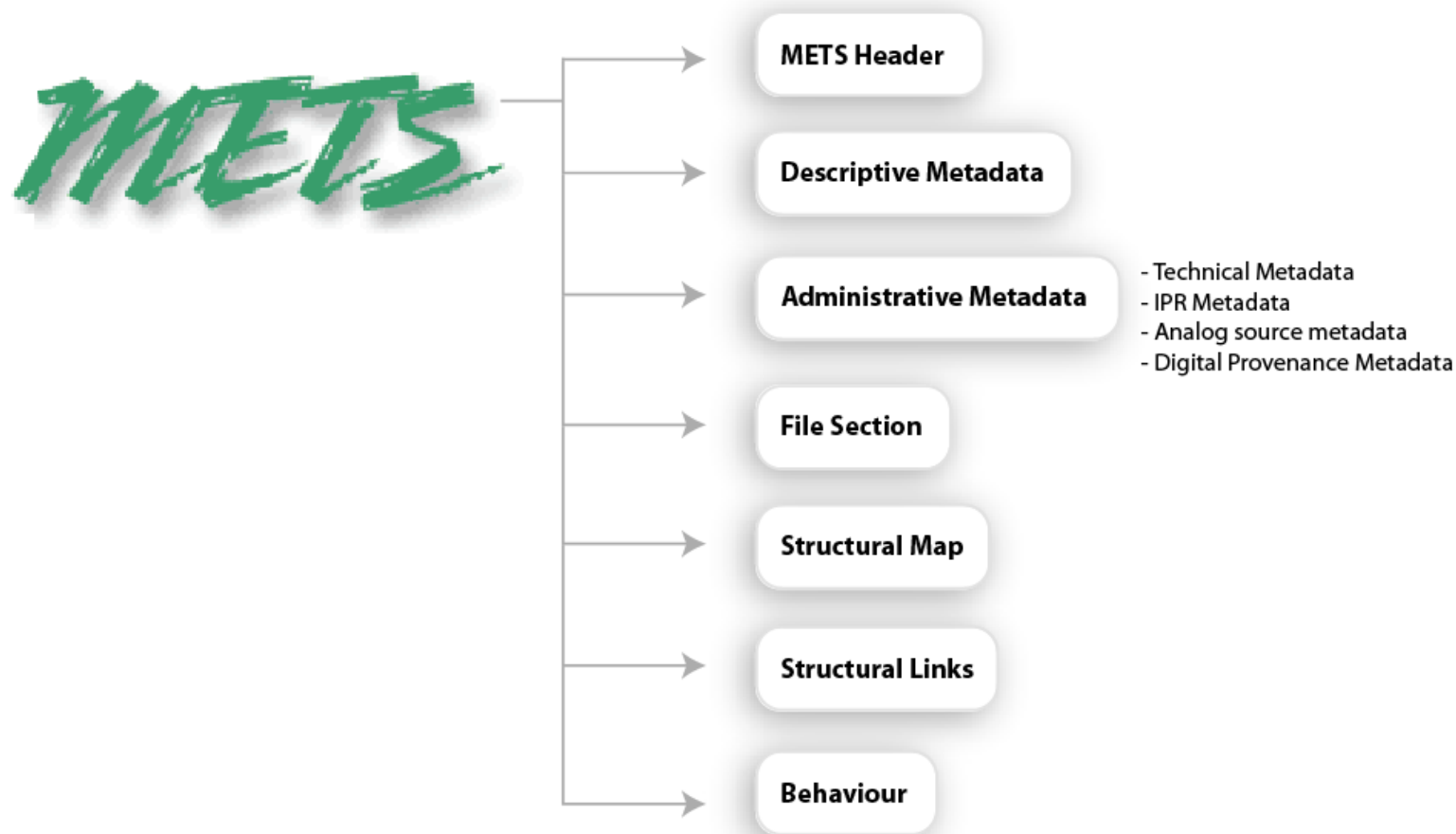
<formatVersion>4.0</formatVersion>

</formatDesignation>

METS

La gestión de objetos digitales requiere el uso de metadatos sobre esos objetos, estos son importantes porque permiten preservar información estructural y referente a las imágenes y los archivos de texto que conforman el OD. También son necesarios metadatos técnicos relativos al proceso de digitalización y otros que permitan la gestión interna, de modo tal que sea posible refrescar y migrar regularmente los contenidos y asegurar la preservación de los recursos. METS (Metadata Encoding and Transmission Standard, es un esquema para describir objetos de bibliotecas digitales complejas que utiliza el lenguaje XML schema y asocia metadatos administrativos y descriptivos.

METS



Secciones en METS

Cabecera METS - contiene metadatos que describen el propio documento METS: fecha de creación, creador, editor, incluso agente.

Metadatos Descriptivos - Esta sección puede: a) apuntar a metadatos descriptivos externos al documento METS (por ejemplo, un registro MARC en un OPAC o un documento EAD disponible en un servidor web); b) contener internamente los metadatos descriptivos, o c) combinar ambas aproximaciones. En la sección Metadatos Descriptivos se pueden incluir ambos.

Metadatos Administrativos - ofrece información sobre cómo se crearon y almacenaron los archivos que conforman el objeto digital, derechos de propiedad intelectual, metadatos sobre el objeto original a partir del cual se obtuvo la representación digital, e información sobre la procedencia de los archivos que conforman el objeto digital (es decir, relaciones entre copias maestras y derivadas, migraciones y transformaciones). Al igual que sucede con los metadatos descriptivos, los metadatos administrativos pueden ser externos o codificarse dentro del propio documento METS.

Cabecera METS

```
<metsHdr CREATEDATE="2003-07-04T15:00:00"  
  RECORDSTATUS="Complete"> <agent ROLE="CREATOR"  
  TYPE="INDIVIDUAL"> <name>Jerome McDonough</name> </agent>  
  <agent ROLE="ARCHIVIST" TYPE="INDIVIDUAL"> <name>Ann  
  Butler</name> </agent> </metsHdr>
```

- En este ejemplo el elemento <metsHdr> contiene dos atributos: CREATEDATE y RECORDSTATUS. Indican respectivamente la fecha y hora en que se creó el documento METS y su estado. Se listan dos agentes que han trabajado en este documento: la persona responsable de su creación y un archivero responsable del material original. Los atributos ROLE y TYPE del elemento <agent> toman sus valores de vocabularios controlados. Los valores permitidos para el atributo ROLE son: "ARCHIVIST," "CREATOR," "CUSTODIAN," "DISSEMINATOR," "EDITOR," "IPOWNER" y "OTHER." Los valores permitidos para el atributo TYPE son: "INDIVIDUAL," "ORGANIZATION" y "OTHER."

Secciones en METS

Sección Archivo - lista todos los archivos con contenidos que forman parte del objeto digital.

Mapa Estructural - es la parte principal de un documento METS. Recoge la estructura jerárquica del objeto digital, y enlaza sus secciones con los archivos de contenido y los metadatos correspondientes.

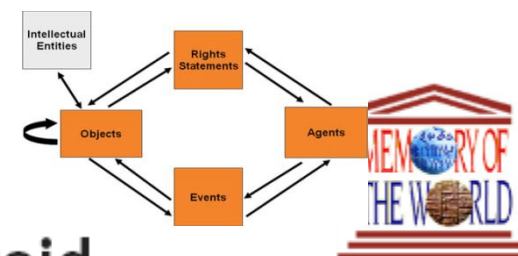
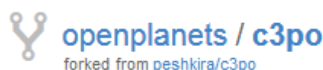
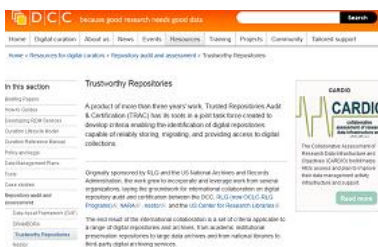
Enlaces Estructurales - permite registrar la existencia de hiperenlaces entre las secciones del mapa estructural. Tiene gran valor cuando se usa METS para archivar sitios web.

Comportamientos - se puede usar para vincular comportamientos ejecutables con los contenidos del documento METS. Cada comportamiento tiene una definición de interfaz y un "mecanismo" que identifica un módulo de código ejecutable que implementa y ejecuta el comportamiento definido de forma abstracta por la interfaz.

Sección 3



Herramientas



Droid

DROID (Digital Record and Object Identification)

The **technical registry**

PRONOM



Open Planets Foundation

United Kingdom <http://openplanetsfound...> info@openplanetsfound...



ISO 14721: 2012
ISO 16363: 2012



Plato 4.4

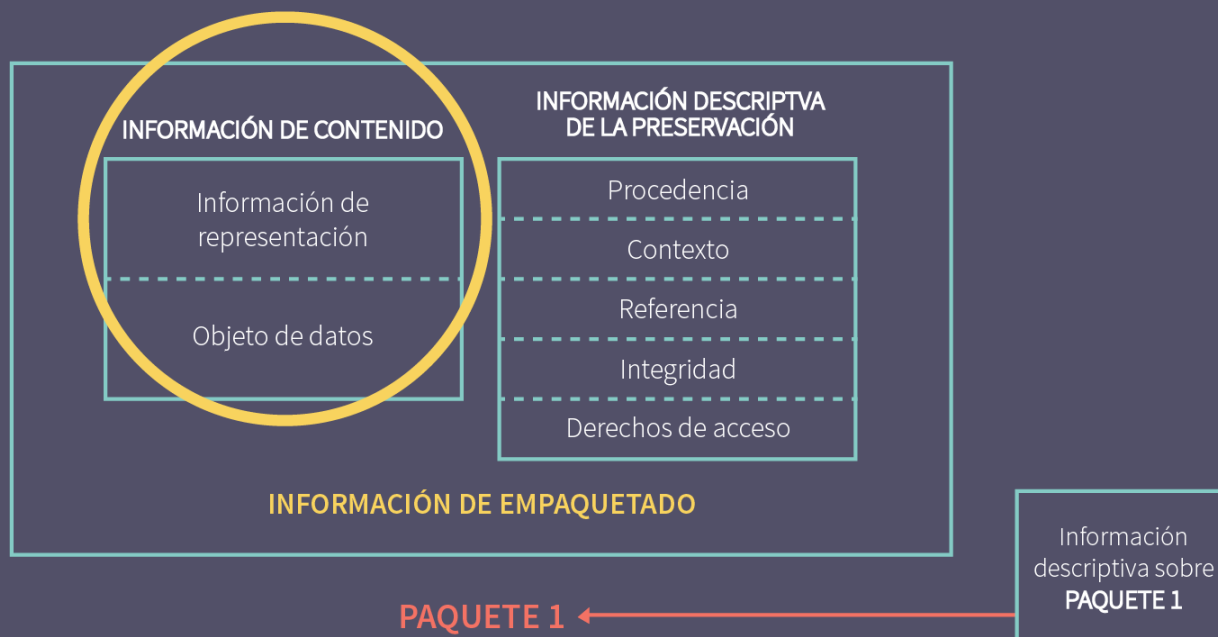


DRAMBORA interactive

Digital Repository Audit Method
Based on Risk Assessment



El paquete de información en el OAIS



DROID: Identificación de formatos

File profiling tool (DROID)

DROID stands for Digital Record Object Identification. It's a free software tool developed by The National Archives that will help you to automatically profile a wide range of file formats. For example, it will tell you what versions you have, their age and size, and when they were last changed. It can also provide you with data to help you find duplicates.

Profiling your file formats helps you to manage your information more effectively. It helps you to identify risks (and therefore plan mitigating actions). It can also help you to save money, for example by supporting data reduction.

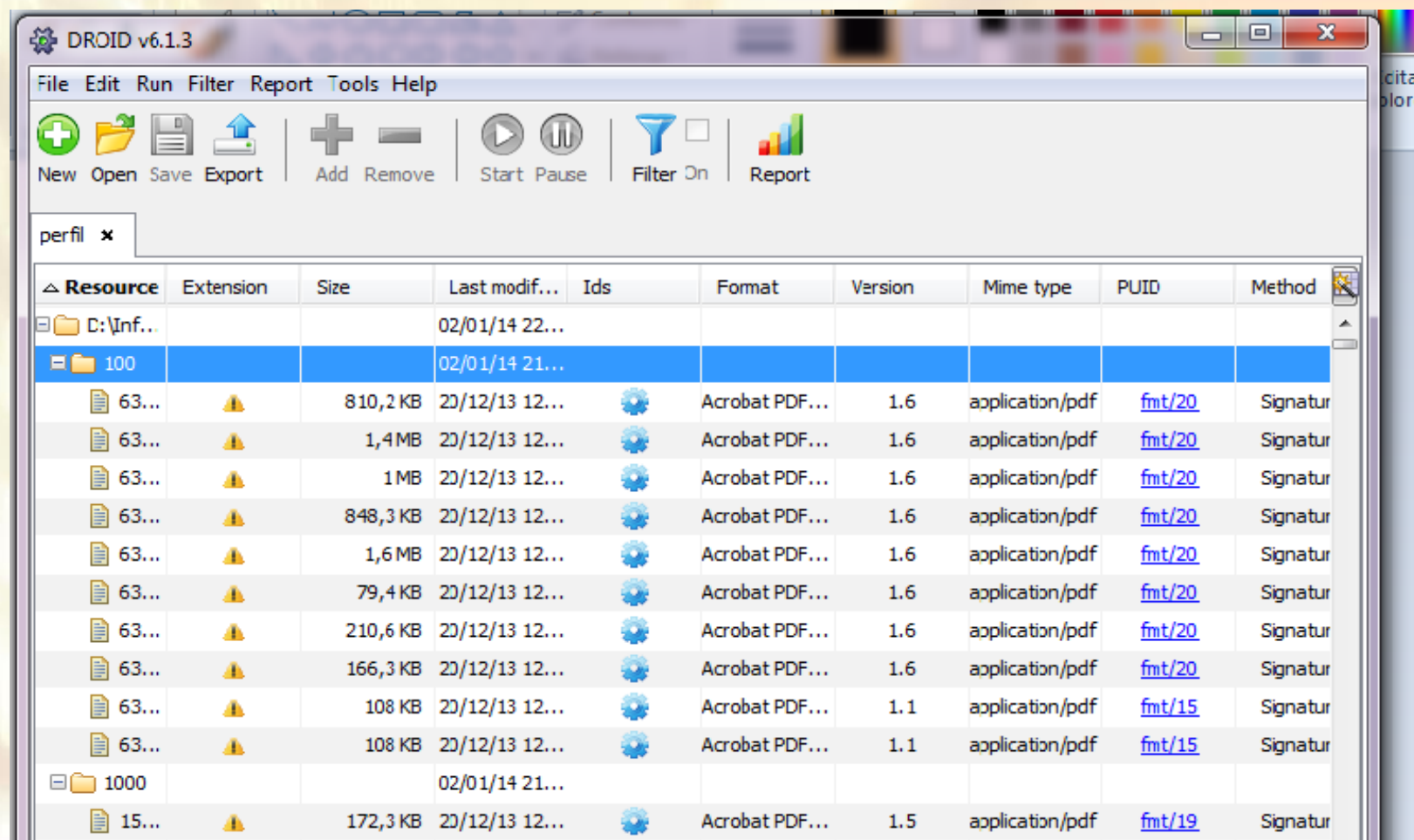
You can [download our latest version of DROID for free](#). For previous versions go to <http://droid.sourceforge.net/>. For more information, see our [PRONOM resource](#).

Also in Managing digital continuity

- [What is digital continuity?](#)
- [Step-by-step guidance](#)
- [Guidance by role](#)
- [Risk assessment](#)

Perfil en DROID

A partir del perfil se revisó el registro PRONOM para ver los riesgos de los formatos



DROID v6.1.3

File Edit Run Filter Report Tools Help

New Open Save Export Add Remove Start Pause Filter On Report

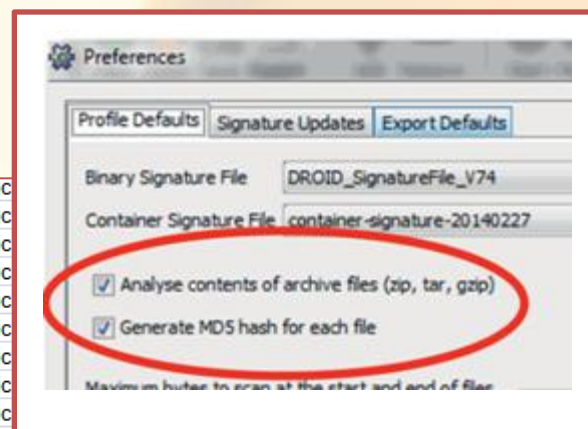
perfil x

Resource	Extension	Size	Last modif...	Ids	Format	Version	Mime type	PUID	Method
D:\Inf...			02/01/14 22...						
100			02/01/14 21...						
63...	⚠	810,2 KB	20/12/13 12...	⚙	Acrobat PDF...	1.6	application/pdf	fnt/20	Signatur
63...	⚠	1,4 MB	20/12/13 12...	⚙	Acrobat PDF...	1.6	application/pdf	fnt/20	Signatur
63...	⚠	1 MB	20/12/13 12...	⚙	Acrobat PDF...	1.6	application/pdf	fnt/20	Signatur
63...	⚠	848,3 KB	20/12/13 12...	⚙	Acrobat PDF...	1.6	application/pdf	fnt/20	Signatur
63...	⚠	1,6 MB	20/12/13 12...	⚙	Acrobat PDF...	1.6	application/pdf	fnt/20	Signatur
63...	⚠	79,4 KB	20/12/13 12...	⚙	Acrobat PDF...	1.6	application/pdf	fnt/20	Signatur
63...	⚠	210,6 KB	20/12/13 12...	⚙	Acrobat PDF...	1.6	application/pdf	fnt/20	Signatur
63...	⚠	166,3 KB	20/12/13 12...	⚙	Acrobat PDF...	1.6	application/pdf	fnt/20	Signatur
63...	⚠	108 KB	20/12/13 12...	⚙	Acrobat PDF...	1.1	application/pdf	fnt/15	Signatur
63...	⚠	108 KB	20/12/13 12...	⚙	Acrobat PDF...	1.1	application/pdf	fnt/15	Signatur
1000			02/01/14 21...						
15...	⚠	172,3 KB	20/12/13 12...	⚙	Acrobat PDF...	1.5	application/pdf	fnt/19	Signatur

Perfil exportado con MD5

para detección de duplicados

1801	15757	3328885	ead803a90d5b190fd4b4935e6a852fc3	1	fmt/17	applic	nt Format	1.3
1802	6355	175675	4564d6e8bc2f282435429c9560e841b1	1	fmt/17	applic	nt Format	1.3
1803	6354	180887	cc0601b8518aa7d0addc22bab2925e1	1	fmt/17	applic	nt Format	1.3
1804	6356	260190	6402659c609ce5bb4d22d585a596517	1	fmt/17	applic	nt Format	1.3
1805	6361	105889	6b4f05561604f71946f3814ed2a313ee	1	fmt/17	applic	nt Format	1.3
1806	6359	105889	6b4f05561604f71946f3814ed2a313ee	1	fmt/17	applic	nt Format	1.3
1807	6360	105889	6b4f05561604f71946f3814ed2a313ee	1	fmt/17	applic	nt Format	1.3
1808	15834	5596903	0d02c816f3b71e08c270dc332fc580a9	1	fmt/17	applic	nt Format	1.3
1809	6362	196539	c45cc1c6dd997d059e318859ae77ee8	1	fmt/17	applic	nt Format	1.3
1810	6363	169421	6aa3fe9e89bce079595fad164a99424	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1811	6364	165625	c25df5f2549f1de3f8082152e5f2c8bc	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1812	6367	120563	9c1db44a3a75b067398aac160c247	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1813	6365	491302	b8769fc53e969d10e37a3a40af15f672	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1814	6366	1089885	ae988ddc448cfe7b526ca79a0cab8bd0	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1815	6372	71355	283535028f1300c050d62da17d520372	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1816	15874	503792	d2adf4cf28a2331e3fc8034a241251a0	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1817	15875	877958	1d2c9ecf09361cee408def76b4cf2a4d	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1818	15877	825731	c1c6d475af3dd9c6ae887741b44ba6cc	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1819	15880	218787	a59e7f27ba35490150c4b530d25a509c	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1820	15888	631271	c1c1e03084b4bdf06a91c19d56e68fcd	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1821	15892	151251	c1788870df8b8e0defddea522a065357	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1822	15891	246622	28a30b03071aed0478ba12e8a6e0a68	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3



Reporte en DROID exportado

File count and sizes

Report field		Grouping fields			
FILE_SIZE					
Filter fields:					
Field		Operator		Values	
RESOURCE_TYPE		NONE_OF		"Folder"	
Profile	Count	Sum	Min	Max	Average
perfil	18522	26863674031	1120	701906423	1450365
Profile totals	18522	26863674031	1120	701906423	1450365

File sizes per extension

Report field		Grouping fields			
FILE_SIZE		FILE_EXTENSION			
Filter fields:					
Field		Operator		Values	
RESOURCE_TYPE		NONE_OF		"Folder"	
Profile	Count	Sum	Min	Max	Average
perfil	18522	26863674031	1120	701906423	1450365
Profile totals	18522	26863674031	1120	701906423	1450365

Group totals

Count	Sum	Min	Max	Average
18522	26863674031	1120	701906423	1450365

File sizes per PUID

Report field		Grouping fields			
FILE_SIZE	PUID	FILE_FORMAT	FORMAT_VERSION	MIME_TYPE	

You are here: [Home](#) > [Information management](#) > [Our projects and work](#) > [Digital preservation](#) > PRONOM



The **technical registry** PRONOM

 Welcome : [About](#)  [Add an entry](#)
 [Search](#) ? [Help](#)  [Information resources](#)

Welcome to PRONOM

[PRONOM changes and DROID signature file release notes.](#)

[DROID signature files.](#)

Find out more about our plans to make PRONOM's data available in a linked open data format on [The National Archives Labs](#).

The online registry of technical information. PRONOM is a resource for anyone requiring impartial and definitive information about the file formats, software products and other technical components required to support long-term access to electronic records and other digital objects of cultural, historical or business value. Find out about the future of PRONOM on our [Information resources](#) page.

 **Search PRONOM** >

Interpretación del reporte de DROID

SÓLO ALGUNOS ARCHIVOS

El perfil elaborado por DROID queda plasmado en el reporte adjunto en la tesis denominado "Comprehensive Breakdown", una síntesis del mismo muestra:

- **47 archivos Macromedia Flash 5**, formato que se corresponde con el PUID fmt/108 de PRONOM.
- **4 archivos Windows Bitmap 3.0**, formato que se corresponde con el PUID fmt/116 de PRONOM.
- **5 archivos MS PPT 1997-2002**, correspondiente al PUID fmt/126 de PRONOM.
- **583 archivos MPEG 1/2 Audio Layer 3**, correspondiente al PUID fmt/134 de PRONOM.
- **17660 archivos PDF**, de los cuales:
 - 1 archivo Acrobat PDF 1.0, formato que se corresponde con el PUID fmt/14 de PRONOM...hasta Acrobat PDF 1.7 que se corresponde con el PUID...
- **188 archivos JPEG**, formato que se corresponde con el PUID ...

JHOVE



JSTOR/Harvard Object Validation Environment

[HOME](#)[TUTORIAL](#)[DOCUMENTATION](#)[DISTRIBUTION](#)[LINKS](#)[MAILING
LIST](#)[JHOVE
PROJECT](#)

JHOVE - JSTOR/Harvard Object Validation Environment

1 Introduction

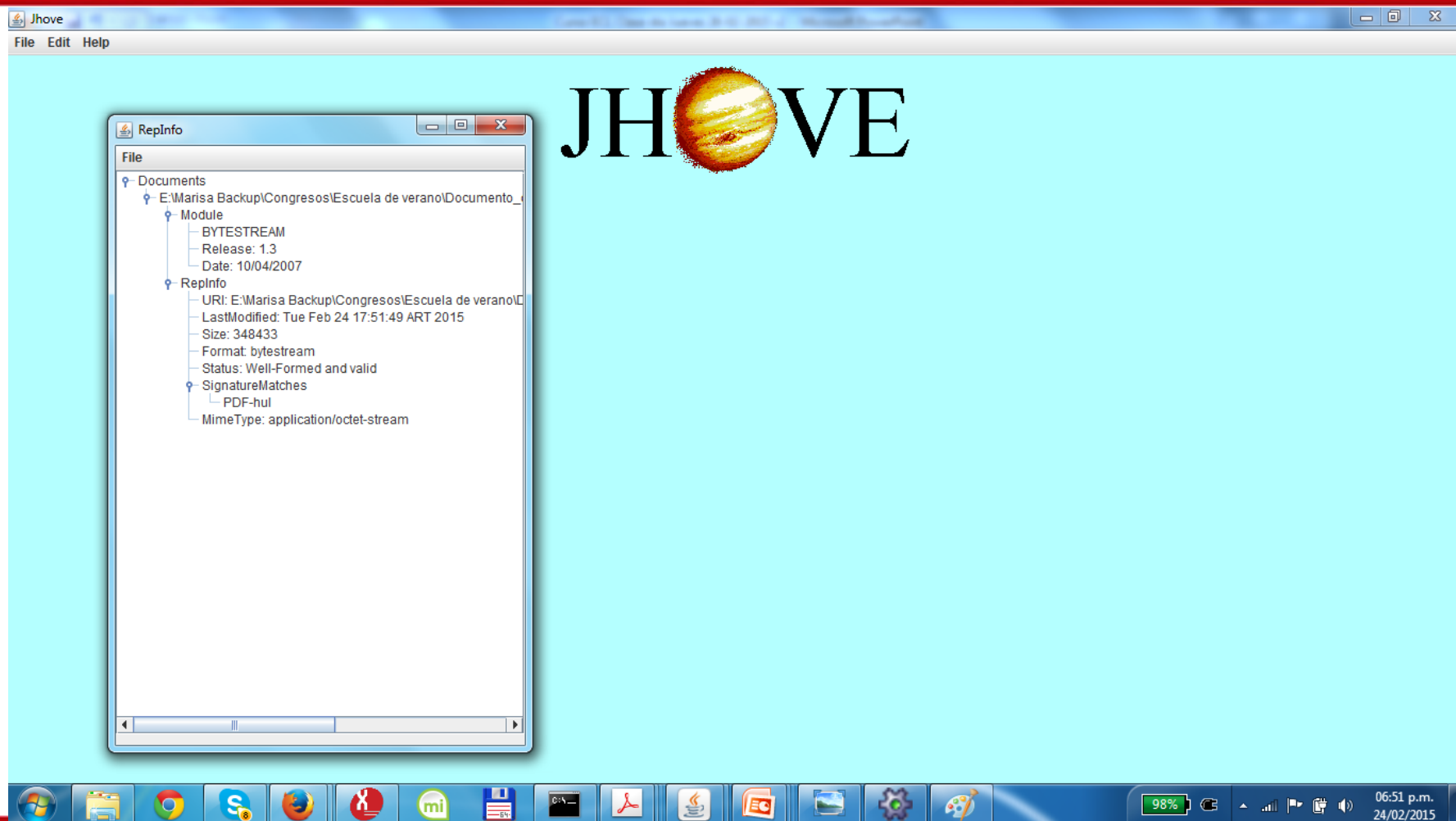
The concept of representation format, or type, permeates all technical areas of digital repositories. Policy and processing decisions regarding object ingest, storage, access, and preservation are frequently conditioned on a per-format basis. In order to achieve necessary operational efficiencies, repositories need to be able to automate these procedures to the fullest extent possible.

JSTOR and the [Harvard University Library](#) are collaborating on a project to develop an extensible framework for format validation: JHOVE (pronounced "jove"), the JSTOR/Harvard Object Validation Environment.

JHOVE provides functions to perform format-specific identification, validation, and characterization of digital objects.

- Format *identification* is the process of determining the format to which a digital object conforms; in other words, it answers the question: "I have a digital object; what format is it?"

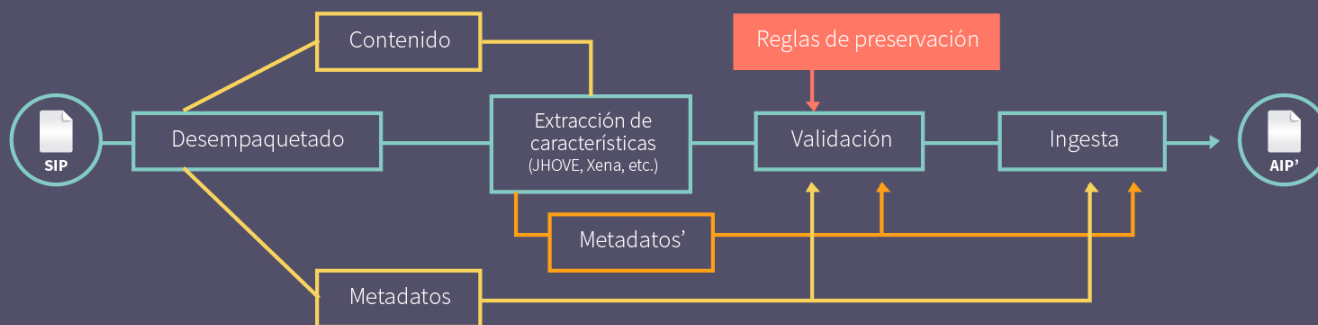
JHOVE



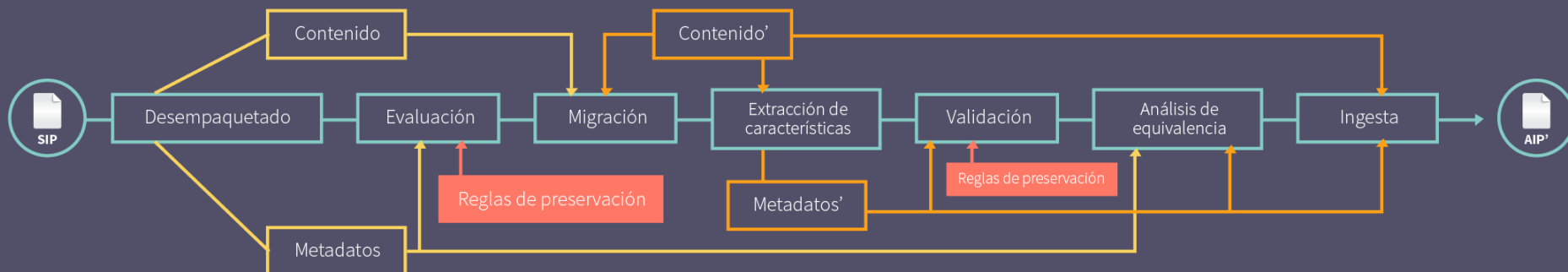
Trabajos futuros

Introducción en el flujo de trabajo de herramientas que permitan la verificación y validación de formatos:

En la ingesta: al ingresar el SIP realizar el desempaquetado y que la herramienta de extracción de características realiza la caracterización del objeto digital.



En el AIP, se evalúa si es necesaria la migración, en caso de serlo se reingresa un nuevo AIP y metadatos.



Metadata Extraction

Enriching the archival record with embedded metadata.

Most popular resources

[Archivematica](#)

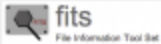
[The BagIt Library](#)

[ICA-AtoM](#)

[NLNZ Metadata Extraction Tool](#)

[JSTOR/Harvard Object Validation](#)

[Environment \(JHOVE\)](#)

	Tool	Implementation	Cost	Platform	Installation	User interface	API
	FITS	Download	Free	Lin Mac Win	Moderate	CL	Yes
	JHOVE2	Download	Free	Lin Mac Win	Simple	CL	Yes
	NLNZ Metadata Extraction Tool	Download	Free	Lin Win	Moderate	CL GUI	Yes
	OpenDataForge	Download	Free Purchase Subscription	Lin Mac Win	Simple	CL GUI	No

Generación automática de metadatos PREMIS-XML y METS-XML

- <http://pim.fcla.edu/>
- Utilizar esta herramienta, básicamente su opción de DESCRIBIR para generar los metadatos PREMIS
- Utilizar esta herramienta para convertir el archivo PREMIS en un archivo METS-XML

C3PO

Clever, Crafty Content Profiling of Objects

Overview

Content profiling consists of three high-level steps: meta-data gathering, processing & aggregation and meta-data analysis. The first step transforms the data in a model that supports faster and scalable analysis and stores it. Post-processing solves issues, such as conflict resolution, due to the normalisation of data provided by different tools and aggregation provides a machine readable overview of the data. The last part of profiling offers the planning expert a service on top of the data. It helps the analysis of the subtleties of the objects and partitioning the content into smaller sets fit for a specific preservation action.



Clever, Crafty, Content Profiling of Objects (c3po) is a software tool prototype, which uses [FITS](#) generated data of a digital collection as input and generates a profile of the content set in an automatic fashion. It is designed in a way so that different meta data formats originating from other tools can be easily integrated. The tool follows the proposed three part profiling process and



[Introduction](#) [Documentation](#) [Cite](#) [Case Studies](#) [Events](#) [Media](#) [History](#) [Team](#)

[Enter Plato 4](#)

What is Plato?

Digital content is short-lived, yet may prove to have value in the future. How can we keep it alive?
Finding the right action to enable future access to digital content in a transparent way is the task of Plato.

The mission of digital preservation is to ensure continued, authentic long-term access to digital objects in a usable form for specific user communities. This requires preservation actions to be carried out when the original environment of digital objects is unavailable. A variety of preservation actions exist, but each shows specific peculiarities, and a variety of factors influence the decision.

The **mission of preservation planning** is to ensure authentic future access for a specific set of objects and designated communities by defining the actions needed to preserve it.

The planning tool **Plato** is a decision support tool that implements a solid preservation planning process and integrates services for content characterisation, preservation action and automatic object comparison in a service-oriented architecture to provide maximum support for preservation planning endeavours.

What's new?

September 2014: Best Demonstration at Digital Libraries 2014

Plan de preservación

Planes de preservación en etapa de inicio del repositorio SEDICI



Plato 4.4



 marisadg

Welcome to Plato 4.4! [M. Kraxner06.03.2014 13:14]

X

A copy has been created: *Plan for electronic papers - marisadg's copy of: This is an example plan. The project was created for the DELOS Summer School 2008 and revised afterwards. (originally created by admin)*

X

It is marked as playground. If you want to use it for serious planning, please change this in Plan Settings. [PLATO11.03.2014 12:59]

My Plans

ID	Name	Description	Author	State	Action
506268	SEDICI Draft Preservation Plan	This is a proof related with SEDICI real preservation plan. My target here is only test this feature and after I'll see how to prepare a real preservation plan for our institutional repository SEDICI	Marisa De Giusti	Initialised	 
506321	SEDICI Draft Presevation Plan	This is a proof whose target is only know something about Plato	Marisa De Giusti	Initialised	 
506425	SEDICI Draft Presevation Plan	This is a test for exploratory reasons. The results will be incorporated it the DSPACE- SEDICI Institutional Repository.	Marisa De Giusti	Records Chosen	 
539484	Plan for electronic papers	marisadg's copy of: This is an example plan. The project was created for the DELOS Summer School 2008 and revised afterwards. (originally created by admin)	Christoph Becker, Andreas Rauber	Weights Set	 
653374	SEDICI-UNLP-001	Plan de prueba para entender la herramienta Plato. A futuro será útil para la migración en masa de documentos PDF al formato standard para preservación a largo plazo PDF/A.	Marisa De Giusti	Experiments Performed	 





DAITSS Digital Preservation Repository Software

[Home](#)[Documentation](#)[Download](#)[News Archive](#)[Contact](#)

Primary links

- [Home](#)
- [Documentation](#)
- [Download](#)
- [News Archive](#)
- [Contact](#)

07/19/2012 - 13:22

[Updated Chapter 5: The DAITSS Archiving Process](#)

An updated version of the DAITSS documentation Chapter 5: The DAITSS Archiving Process is now available from the Documentation section of the DAITSS website. The SIP

Welcome to the DAITSS website!

DAITSS is a digital preservation software application developed by the Florida Center for Library Automation (FCLA) with some support from the IMLS. DAITSS is used by the [Florida Digital Archive \(FDA\)](#), a long-term preservation repository service provided by the [Florida Virtual Campus](#) for the use of the libraries of the eleven publicly-funded universities in Florida.

DAITSS provides automated support for the functions of Submission, Ingest, Archival Storage, Access, Withdrawal, and Repository Management. It is architected as a set of RESTful Web Services and micro-services but enforces strict controls to ensure the integrity and authenticity of archived content. It implements active preservation strategies based on format-specific processing including, where necessary, normalization and forward migration. It is particularly well suited for materials in text, document, image, audio and video formats.

DAITSS was written for a multi-user environment and supports consortial as well as institutional preservation repositories.

DAITSS is available for use through a GPLv3 license. This website provides access to a fully configured VM version of DAITSS that can be downloaded to run under any VM manager, along with sample SIPs (submission packages) and documentation.

For more general information about DAITSS, see:

- [DAITSS, an OAIS-based Preservation Repository](#)
- [DAITSS Grows Up: Migrating to a Second-Generation Preservation System](#)

Evaluación y sistemas de auditoría

- Los sistemas especializados en preservación digital presentan particularidades de planificación y gestión que no son revisadas adecuadamente por normas para sistemas de gestión general, como por ejemplo ISO 9000 ó la ISO 27001 de auditoría de un sistema de información.

Sistemas de auditorías especializadas

- Drámbora (Reino Unido)
 - <http://www.repositoryaudit.eu/>
- Nestor 2 (Alemania)
 - http://www.langzeitarchivierung.de/Subsites/nestor/EN/Home/home_node.html;jsessionid=152978E15E5458B991F2387FDCA85107.prod-worker2
- TRAC (Estados Unidos) reconocido como Norma ISO/IEC 16363:2012

¡GRACIAS!

Dra. Marisa R. De Giusti

marisa.degiusti@sedici.unlp.edu.ar

