SEDICI
REPOSITORIO INSTITUCIONAL DE LA UNLP

13 CONGRESO NACIONAL DE BIBLIOTECOLOGÍA Y CIENCIA DE LA INFORMACIÓN
20 JORNADA NACIONAL Y V INTERNACIONAL DE ACTUALIZACIÓN Y CAPACITACIÓN DE BIBLIOTECAS MÉDICAS

Las actividades de preservación en un repositorio digital destinadas a dar acceso a lo largo del tiempo a sus contenidos

Bogotá, Colombia, Octubre de 2014

Prof. Ing. Marisa R. De Giusti
Directora PREBI-SEDICI
Universidad Nacional de la Plata

sedici.unlp.edu.ar

sedici.unlp sedici_unlp info@sedici.unlp.edu.ar

UNLP

Índice

Parte 1 - Conceptos básicos

Repositorio, interoperabilidad, preservación, guías, proyectos

Parte 2 - Metadatos de preservación

Parte 3 - Directrices sobre preservación

PREMIS, Modelo de datos PREMIS, METS

Otros esquema de metadatos y más posibilidades en la preservación

Parte 4- OAIS

Parte 5- DSPACE

Modelo de datos, OAIS en Dspace

¿Qué es un repositorio institucional?

Un RI es un conjunto de servicios Web centralizados, creados para organizar, gestionar, preservar y ofrecer acceso libre a la producción científica, académica o de cualquier otra naturaleza cultural, en soporte digital, generada por los miembros de una institución. Su creación es un proceso MULTIDISCIPLINARIO. Las principales características de un repositorio institucional son:

- Su naturaleza institucional, entendiendo por institución a una organización educativa y de investigación y habiendo tenido como punto de partida a las universidades.
- Su carácter acumulativo y perpetuo.
- Su carácter abierto e interoperable con otros sistemas.
- El uso de mucha tecnología.

¿Qué es un repositorio institucional?

Un Repositorio brinda un conjunto de servicios a una comunidad, destinados a recopilar, gestionar, difundir y preservar contenidos a través de una colección organizada y accesible en abierto que debe estar provista de facilidades que le permiten interoperar con otros repositorios similares.

El desarrollo de los servicios en un repositorio no es parejo: gestión y difusión avanzadas, recopilación compleja, preservación: reciente.

Via verde - Los repositorios institucionales

Preservación

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.

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

El primer paso: qué preservar

La “Guía Unesco: *Directrices para la preservación del Patrimonio Digital*” [1], sostiene en su artículo 5.2.1 sobre Patrimonio que **“El patrimonio digital está constituido únicamente por aquellos que se considera que poseen un valor permanente”.** Este es el primer punto de un plan de preservación: definir materiales y tiempos de preservación de cada uno, definir formatos y versiones.
Definir qué se quiere preservar (datos y funcionalidad, apariencia y esencia). Algunas decisiones pueden afectar los formatos que se recepcionan en el repositorio.

[1] UNESCO “Directrices para la preservación del Patrimonio digital | Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura,” 2012.

Amenazas

1. La propia naturaleza de los objetos digitales los hace efímeros.
2. Los OD siempre están mediados por la tecnología que cambia constantemente: equipos y programas, hardware, software y formatos.
3. Pérdidas por desastres.
4. Barreras de acceso en los archivos: claves, cifrado.
5. Desconocimiento de las responsabilidades por la curatela.
6. Falta de conocimientos, de planificación, de recursos.
7. Problemas legales: permisos para transformar las obras.
8. Descripción inadecuada: imposibilidad de recuperación.
9. Pérdida de información sobre el contexto.

La preservación de contenidos de otros RI's. Problemas adicionales

La clase de documentos puede ser muy variada:

- Proyectos de investigación
- Datos, resultados
- Tesis
- Artículos
- Materiales digitalizados
- Informes
- Planos
- Material administrativo
- Imágenes, obras de arte
- Fondos patrimoniales

Preservación de contenido

Estrategias

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

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

Christoph Becker, Hannes Kulovits, Mark Guttenbrunner, Stephan Strodl, Andreas Rauber & Hans Hofman. “Systematic planning for digital preservation: evaluating potential strategies and building preservation plans”. International Journal on Digital Libraries, Volume 10, issue 4, pp 133-157. Diciembre de 2009.

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
- Trabaja sobre el objeto
- 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.
- Trabaja sobre el entorno del objeto.
- Una vez que se archivaron los datos, sólo hay que asegurarse que el soporte físico utilizado siga siendo accesible.
- Se puede usar un mismo emulador para múltiples objetos del mismo tipo.

Preservación de contenido

Encapsulamiento

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

Como elementos a encapsular podemos tener:

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

Preservación de contenido

Metadatos de preservación

Generalmente considerados como metadatos administrativos

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

Varias iniciativas:
- PREMIS: PREservation Metadata: Implementation Strategies
- OAIS: Open Archival Information System
- NEDLIB: Networked European Deposit Library
http://www.dnb.de/EN/Wir/Projekte/Abgeschlossen/nedlib.html

Preservación de contenido

Políticas de backup

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

- Incendios
- Inundaciones
- Robos
- Fallas de hardware

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

- Incremental
- Espejo

Aproximaciones a la preservación

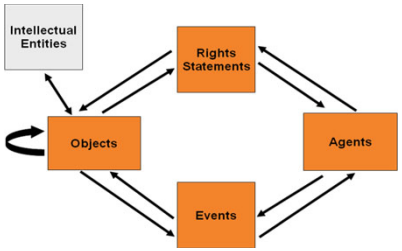
MEMORY OF THE WORLD

The Library of Congress » Standards » PREMIS Home

Standards Pages

PREMIS

PRESERVATION METADATA MAINTENANCE ACTIVITY



Droid

DROID (Digital Record and Object Identification)

ISO 14721: 2012

The technical registry PRONOM

JISC KeepIT

Plato 4.4

Índice

Parte 1 - Conceptos básicos

Repositorio, interoperabilidad, preservación, guías, proyectos

Parte 2 - Metadatos de preservación

Parte 3 - Directrices sobre preservación

PREMIS, Modelo de datos PREMIS, METS

Otros esquema de metadatos y más posibilidades en la preservación

Parte 4- OAIS

Parte 5- DSPACE

Modelo de datos, OAIS en Dspace

Metadatos y metadatos de preservación

Metadatos tradicionales

Metadatos de preservación

- Procedencia,
- Autenticidad
- Contexto
- Gestión de derechos
- Actividades de preservación

Metadatos y metadatos de preservación

•**Metadatos tradicionales:** describen, explican, localizan, facilitan la localización y gestionan un recurso, pueden ser descriptivos, administrativos o estructurales (hay modos diferentes de clasificarlos).

•**Metadatos de preservación**: documentan el proceso de preservación de los recursos digitales, también son descriptivos, administrativos y estructurales (si se sigue la misma taxonomía previa) aunque tradicionalmente se los considera administrativos.

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.

OD y metadatos de preservación: Los metadatos que son útiles a la preservación pueden verse más fácilmente en su utilidad en esta gráfica donde se representa junto al Objeto de Información (OD), las acciones a realizar para su preservación de las cuales se desprenden los metadatos necesarios.

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

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

Debe conocerse su **Creador**

Deben guardarse las **relaciones** que vinculen al objeto con otros

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

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

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

Autenticidad
Mediante la documentación de su procedencia

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

Índice

Parte 1 - Conceptos básicos

Repositorio, interoperabilidad, preservación, guías, proyectos

Parte 2 - Metadatos de preservación

Parte 3 - Directrices sobre preservación

PREMIS, Modelo de datos PREMIS, METS

Otros esquema de metadatos y más posibilidades en la preservación

Parte 4- OAIS

Parte 5- DSPACE

Modelo de datos, OAIS en Dspace

PREMIS

PREMIS es un grupo de trabajo internacional patrocinado por *Online Computer Library Center* (OCLC) y *Research Libraries Group* (RLG) que 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".

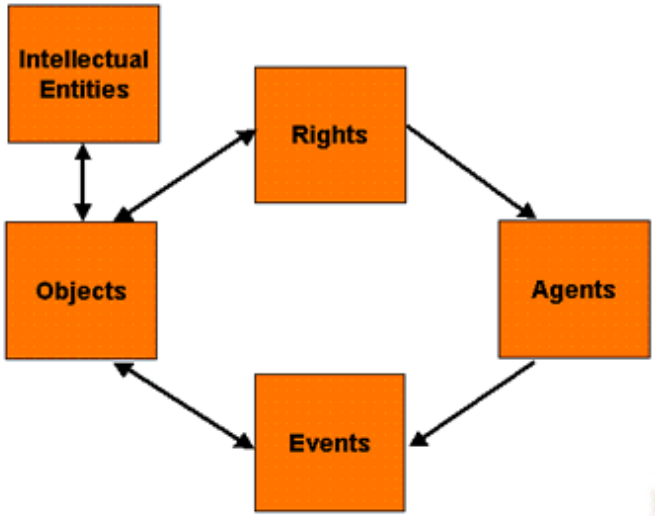
El diccionario de datos PREMIS define un conjunto de *unidades semánticas*, propiedades, e información que la mayoría de los repositorios necesita conocer de sus entidades para asegurar la preservación.

Tanto el *Framework como el Diccionario de Datos PREMIS se adaptan al modelo de referencia OAIS- ISO 14721.*

http://www.bne.es/es/Micrositios/Guias/DiccionarioPremis/resources/images/docs/PREMIS_es.pdf

PREMIS PRESERVATION METADATA MAINTENANCE ACTIVITY

PREMIS data model



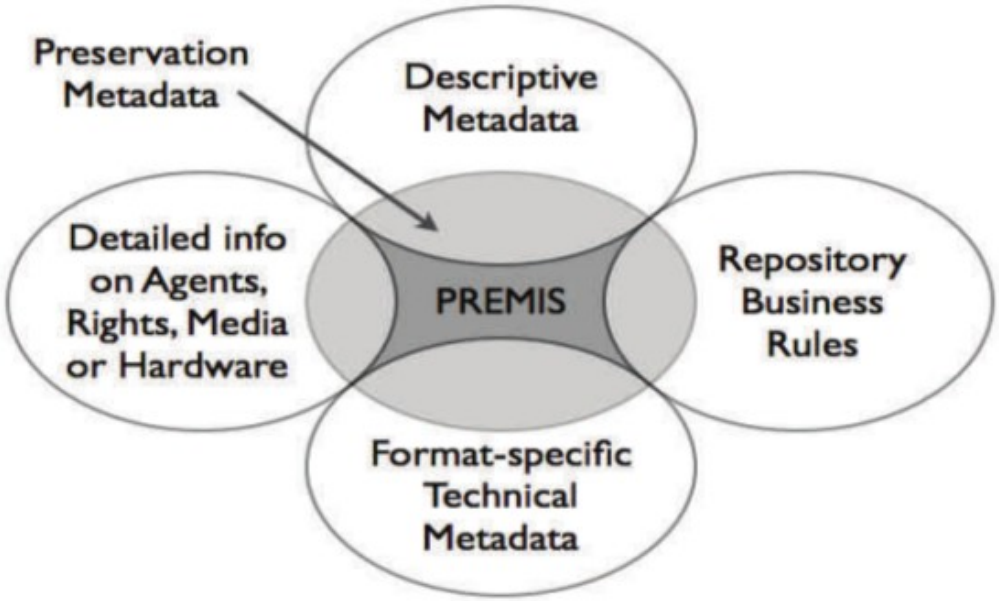


Imagen de Priscilla Caplan, Understanding PREMIS (2001)

PREMIS y OAIS

El Framework PREMIS puede ser considerado una aplicación práctica del modelo de referencia OAIS a través del mapeo de los metadatos de preservación a esa estructura conceptual; el Diccionario de Datos PREMIS puede ser entendido como la traducción del Framework a un conjunto de unidades semánticas.

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

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 entendemos 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*** (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. 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).

Unidades semánticas

"El Diccionario de Datos PREMIS define unidades semánticas. Cada unidad semántica especificada en el Diccionario de Datos se mapea a una de las entidades del modelo de datos. En este sentido, una unidad semántica puede entenderse como una propiedad de una entidad. Por ejemplo, la unidad semántica *size es una propiedad de la entidad Objeto. Las unidades semánticas poseen valores: para un objeto concreto el valor de size puede ser «843200004»".*

"" Tomada del diccionario de datos PREMIS.

Índice

Parte 1 - Conceptos básicos

Repositorio, interoperabilidad, preservación, guías, proyectos

Parte 2 - Metadatos de preservación

Parte 3 - Directrices sobre preservación

PREMIS, Modelo de datos PREMIS, METS

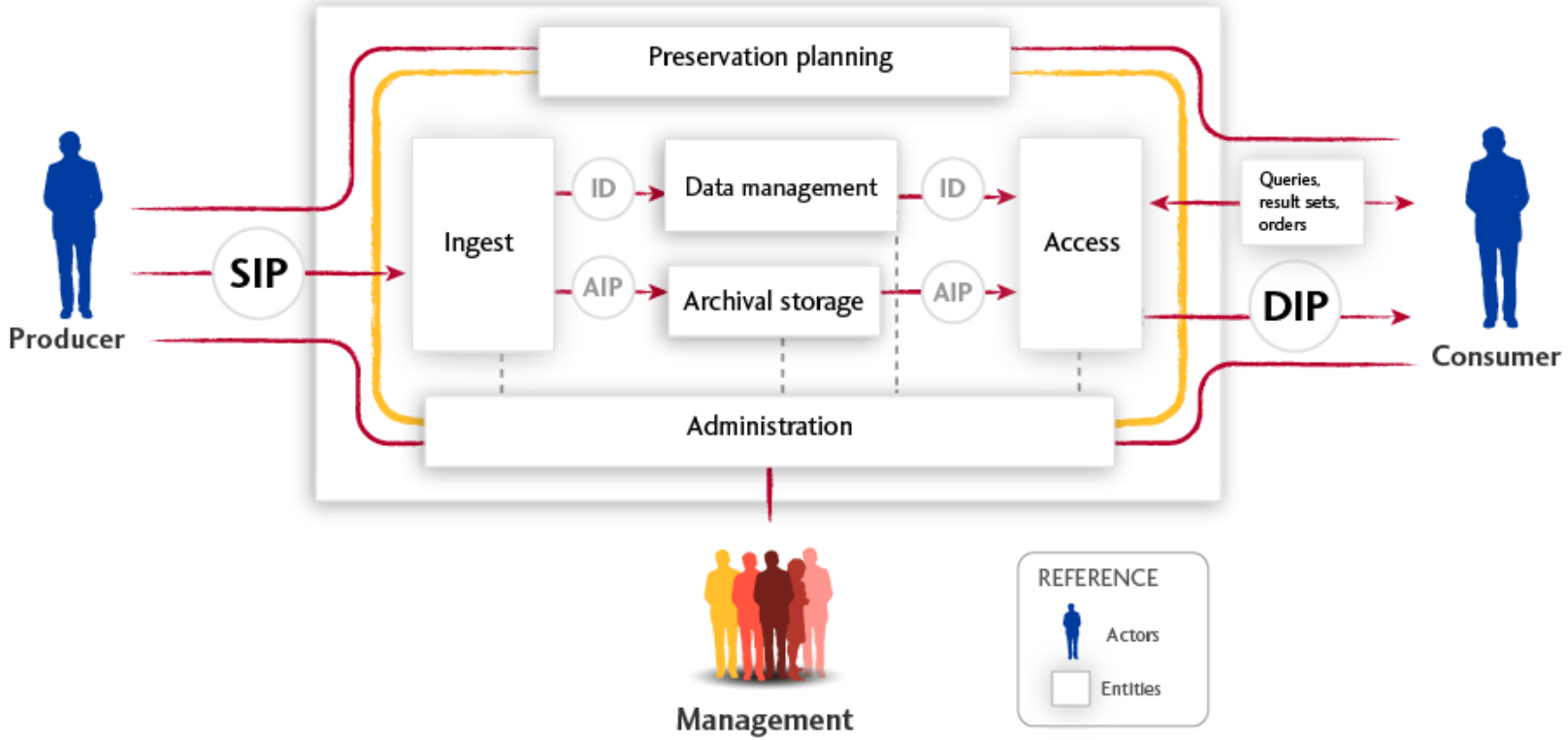
Otros esquema de metadatos y más posibilidades en la preservación

Parte 4- OAIS

Parte 5- DSPACE

Modelo de datos, OAIS en Dspace

Directrices de la Norma ISO 14721. Última versión Junio de 2012. ISO Reference Model for an Open Archival information System (OAIS)



Preservation planning: ¡¡sólo en la mente!!

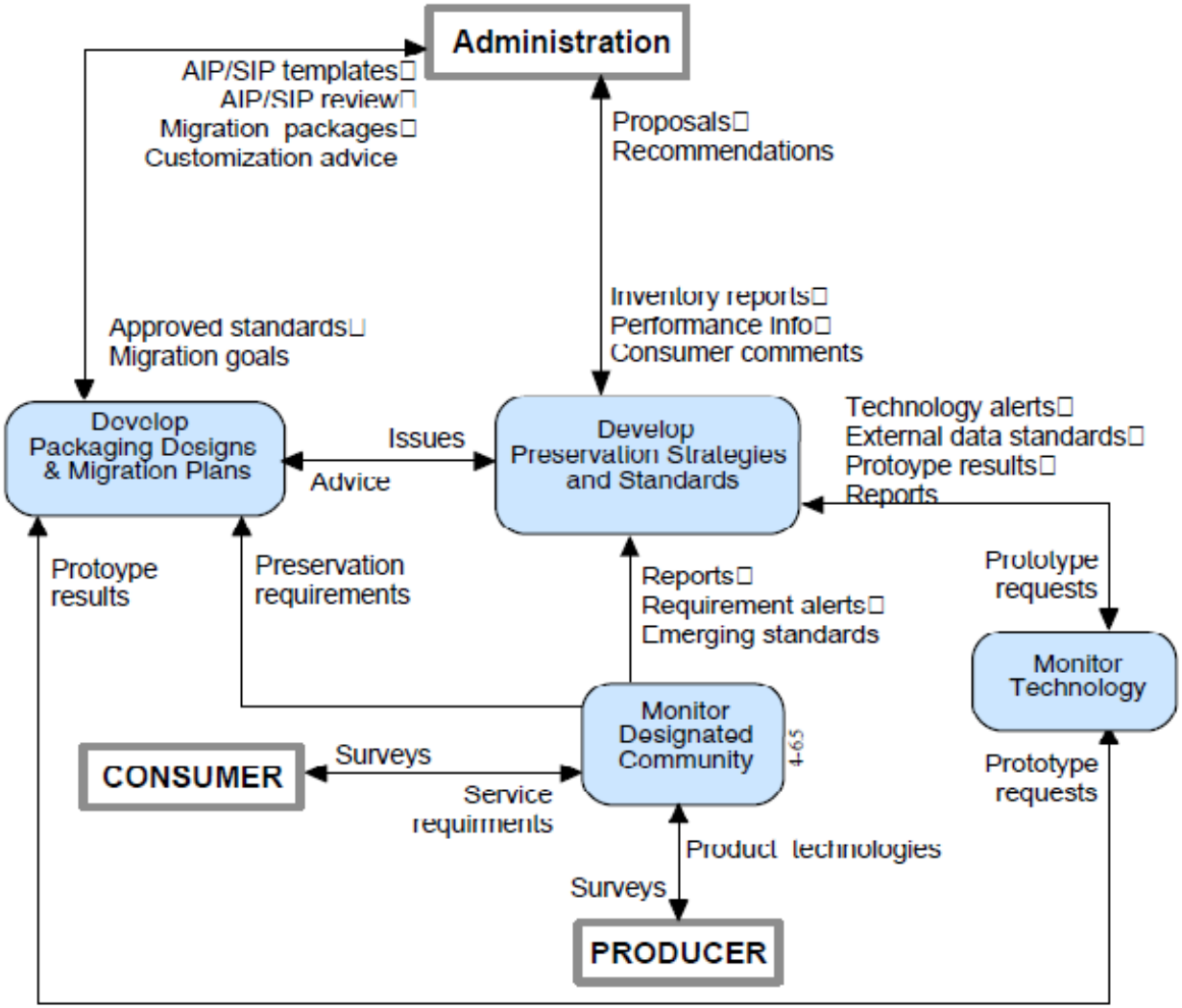


Figure 4-6: Functions of Preservation Planning

OAIS (Open Archival Information System)

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.

Funciones del Modelo de Referencia

Las dos funciones principales del modelo son conservar la información y garantizar el acceso a la misma.

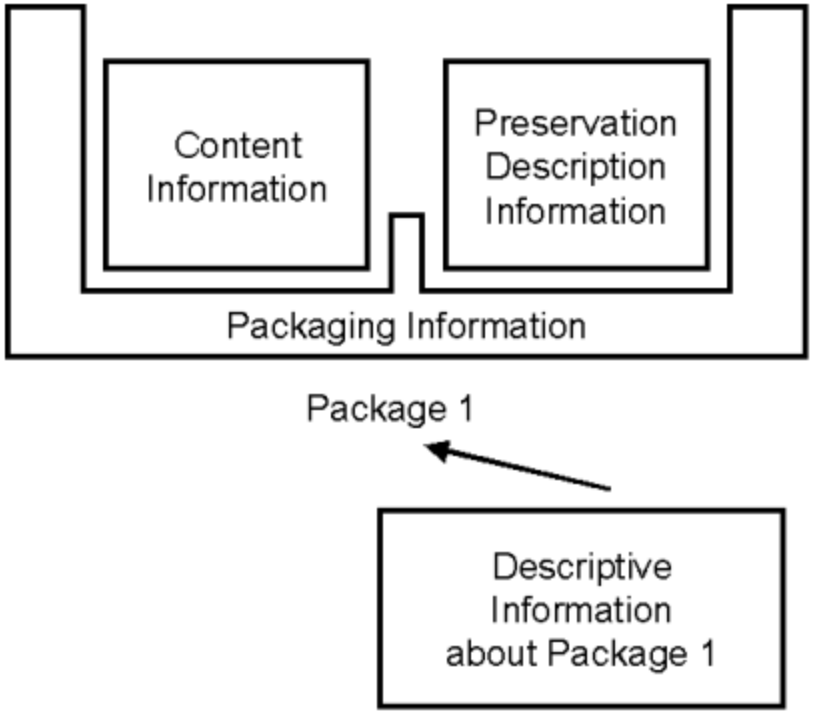
El modelo funcional OAIS, que se propone lograr estos objetivos amplios, en cierta medida, define la arquitectura aproximada de cualquier tipo de sistema de software diseñado para cumplir con esta norma y con todo tipo de flujos de trabajo asociados con el repositorio.

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 (CI)
  - De descripción de preservación (PDI)

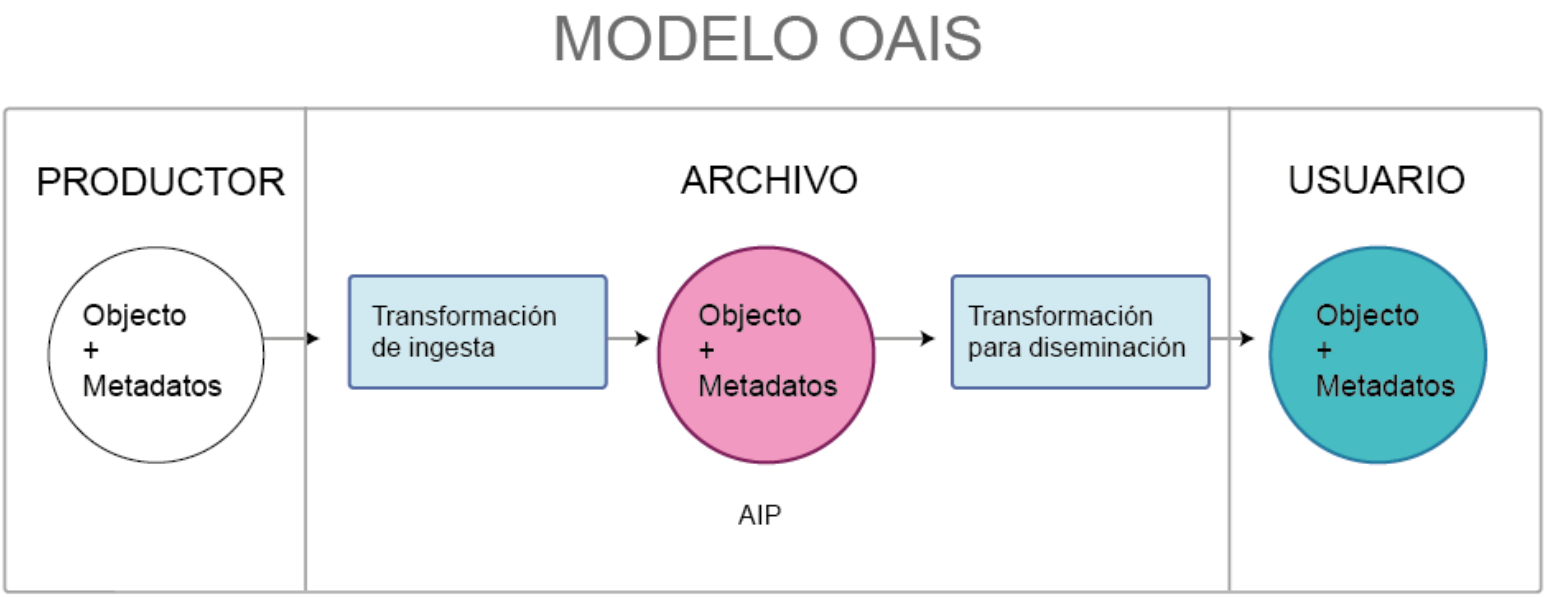
La información de contenido y la PDI pueden verse como encapsuladas e identificables por medio de la información de empaquetado.

- El paquete resultante es recuperable en virtud de la información descriptiva: DI.



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

Variantes del IP: SIP, AIP y DIP



Información de contenido

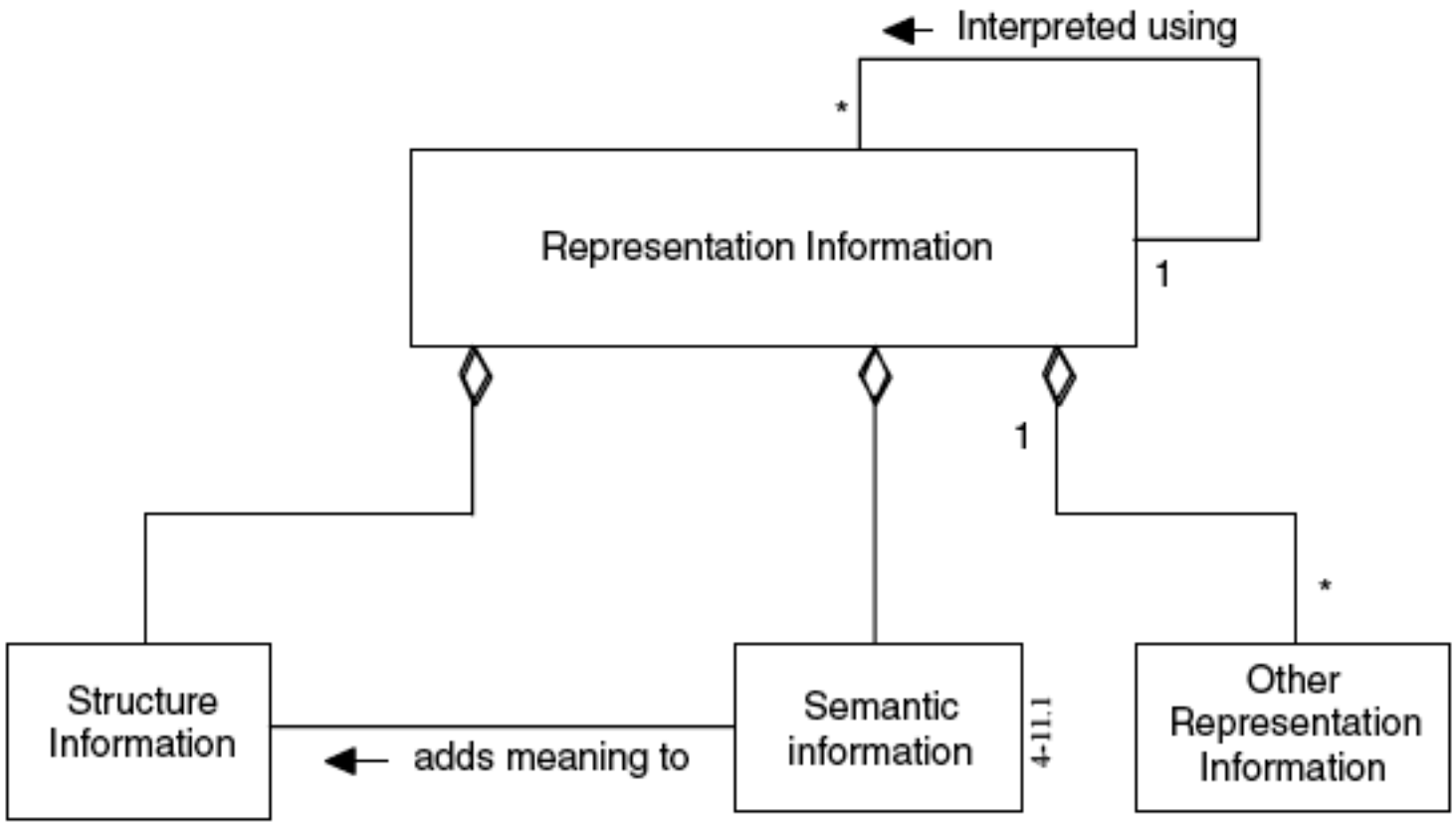
La información de contenido es el conjunto de información que es el objetivo original de la preservación de la OAIS.

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 de representación



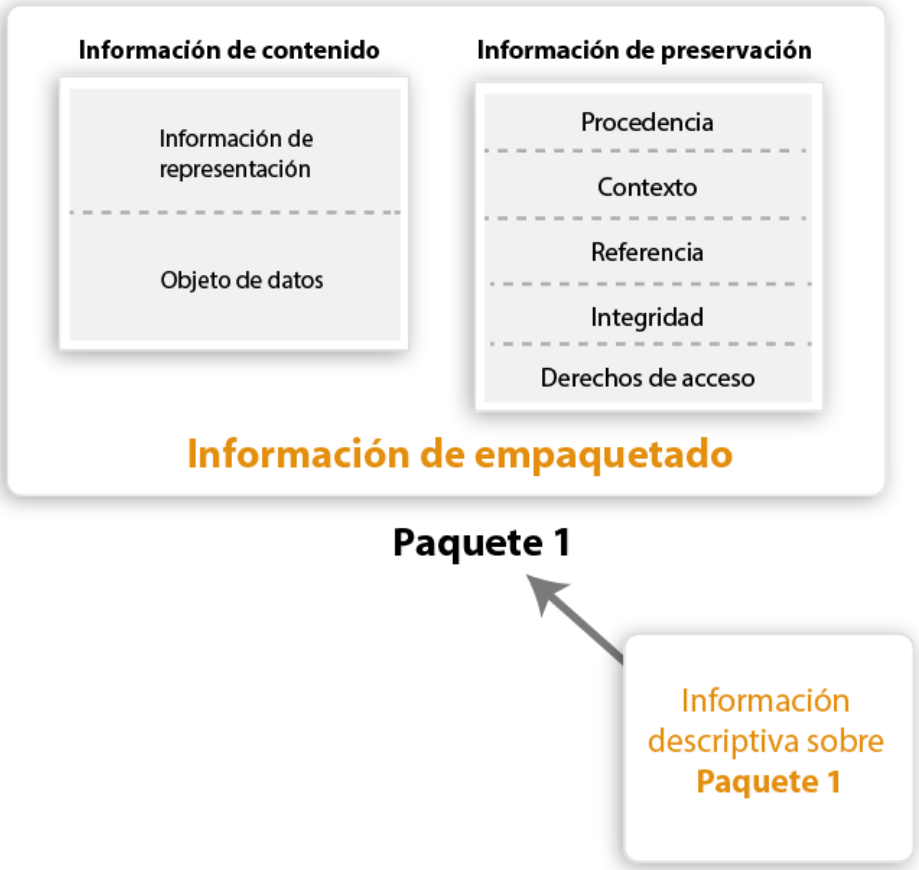
Información de representación: item: http://sedici.unlp.edu.ar/handle/10915/25088

Archivos del ítem

Nombre	Descripción	Formato	Ver	Orden
Bloque: TEXT				
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)
Bloque: ORIGINAL				
Tesina de Licenciatura - Almazan Maria Belen.pdf (principal)	Documento completo	Adobe PDF	[Ver]	1 (Anterior:1)
	Presentación	Adobe		2

Información descriptiva de preservación (PDI)

Debe brindar datos suficientes sobre la **procedencia**, el **contexto**, la **referencia**, la **integridad** y **los derechos de acceso**. 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.



La información de empaquetado es la 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.

Índice

Parte 1 - Conceptos básicos

Repositorio, interoperabilidad, preservación, guías, proyectos

Parte 2 - Metadatos de preservación

Parte 3 - Directrices sobre preservación

PREMIS, Modelo de datos PREMIS, METS

Otros esquema de metadatos y más posibilidades en la preservación

Parte 4- OAIS

Parte 5- DSPACE

Modelo de datos, OAIS en Dspace

DSpace 3 grupos lógicos de metadatos para el contenido archivado

1) **Metadatos descriptivos:** soporta múltiples formatos de metadatos no jerárquicos y permite mantener para un mismo ítem, metadatos de diversos formatos, habilitando de esta forma la utilización de uno o más Application Profiles. Para colecciones y comunidades, sólo permite gestionar una serie de campos fijos básicos aunque probablemente en las próximas versiones de Dspace se permita configurar los metadatos como sucede con los items.

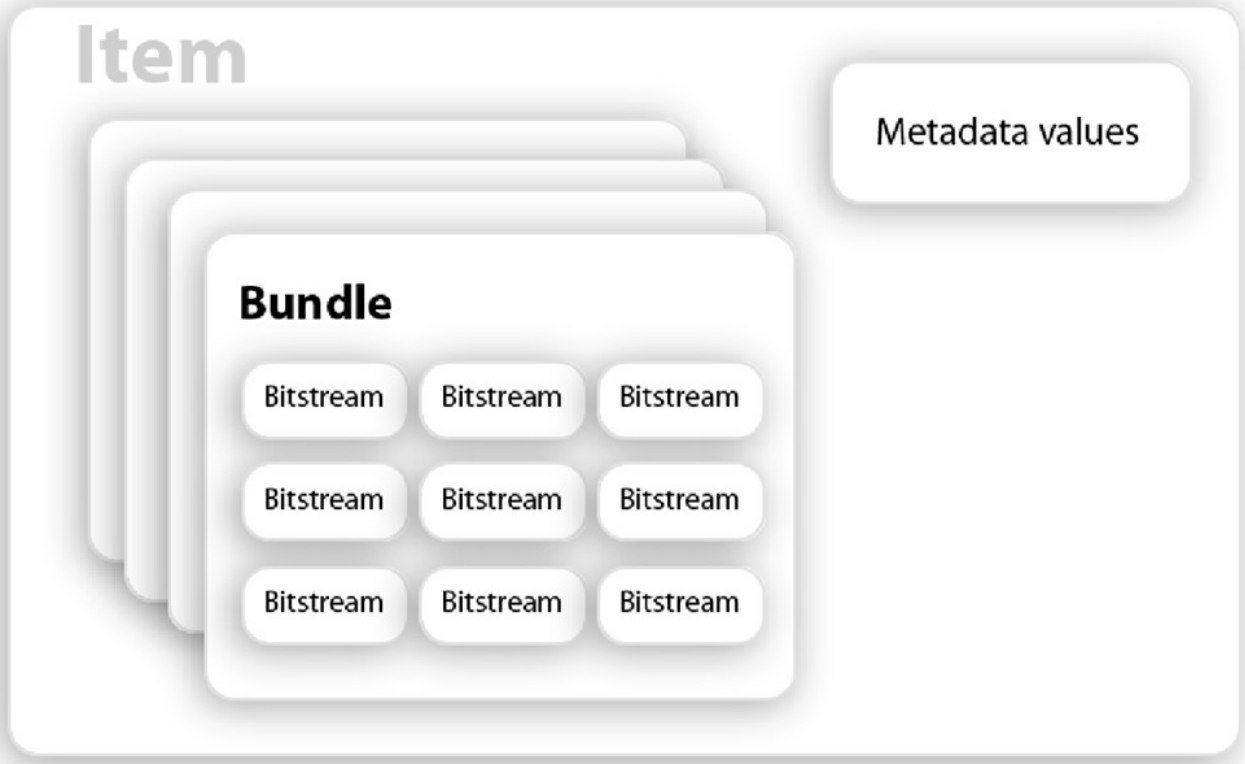
DSpace 3 grupos lógicos de metadatos para el contenido archivado

2) **Metadatos Administrativos:** 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.

DSpace 3 grupos lógicos de metadatos para el contenido archivado

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.

Modelo de datos en DSPACE



OAIS en 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 y metodología. 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 (http://www.handle.net/). Sin embargo, 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).

EXPERIMENTACIÓN

Acciones propuestas

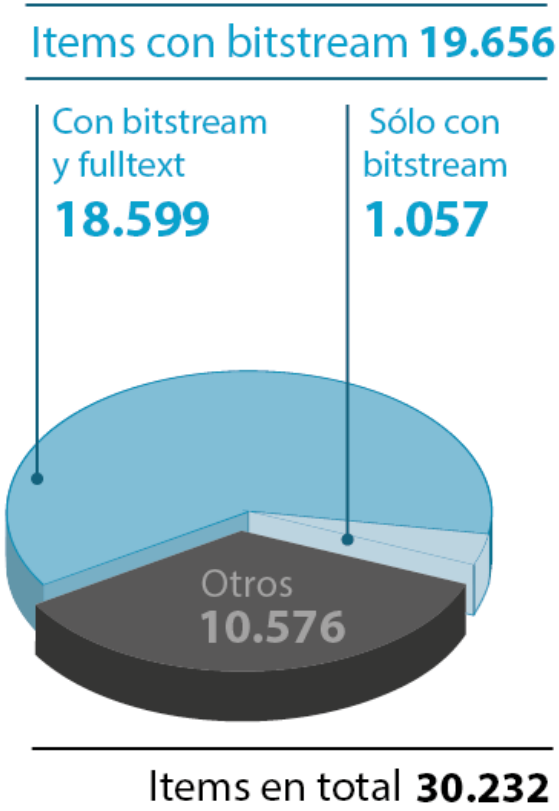
Las acciones propuestas están vinculadas justamente a los elementos constitutivos del paquete de información: la información de contenido (CDO), la información sobre la representación de ese contenido (RI), la información descriptiva de preservación (PDI), la información de empaquetado (PI) y la información descriptiva (DI).

Acciones

La propuesta tiene como fin generar un reporte sobre el estado de los objetos del repositorio, considerándolos como paquetes de información. Se reportará que están bien formados (o no) y serán preservables (o no) si cuentan con todos los elementos que la norma define para el paquete de información; también se validará que cada elemento esté bien formado en el sentido de que cumple con una serie de estándares o criterios que se van a definir a continuación. Si los paquetes de información en el repositorio se adecúan a los criterios establecidos, los objetos digitales del repositorio y por tanto el repositorio mismo “pasan” la evaluación.

SEDICI en números

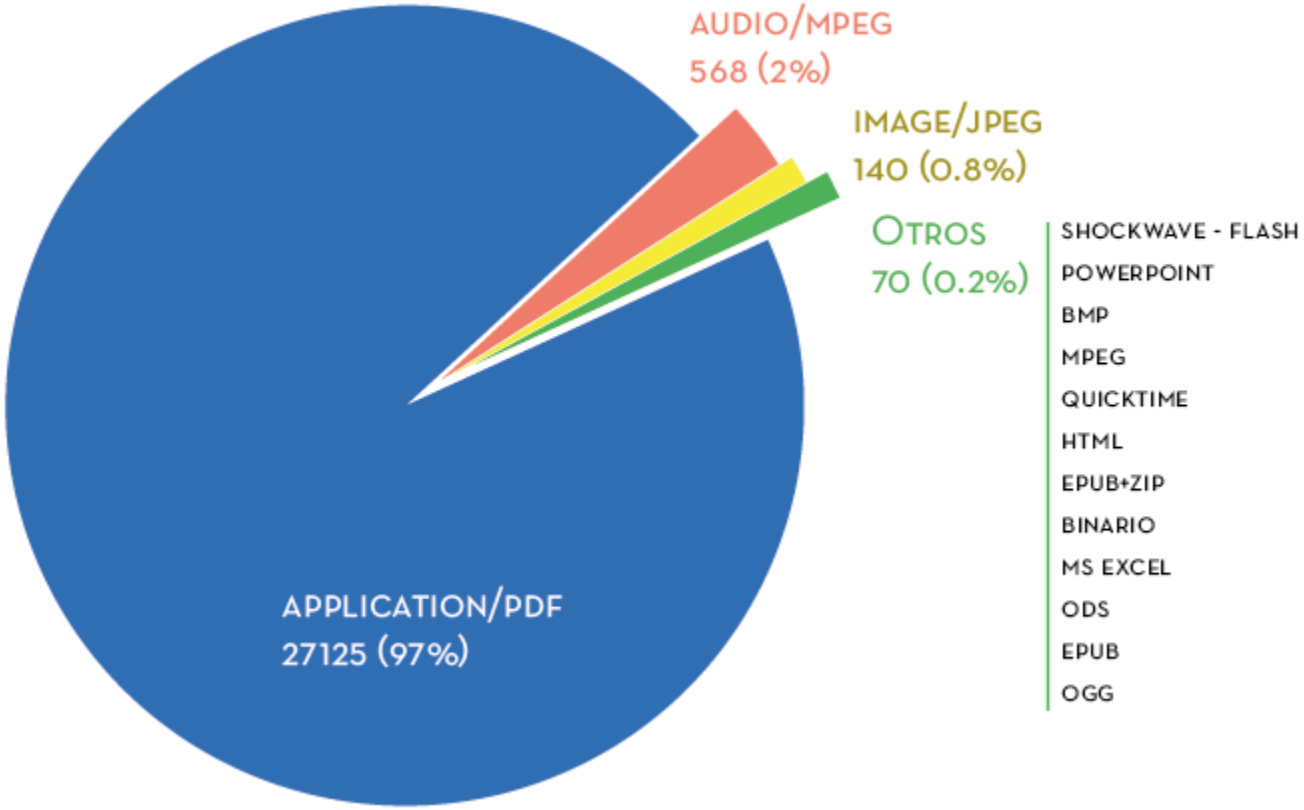
- **ITEMS TOTALES EN EL REPOSITORIO**: **30.323**
- **ITEMS CON BITSTREAM**: **19.656**
- **ITEMS CON BITSTREAM y CON FULLTEXT**: **18.599**



AssetStore en análisis

El conjunto de archivos analizados de SEDICI comprende 17359 PDFs en diferentes versiones, 47 archivos en Flash, 4 BMPs, 5 PPTs, 583 audios en MPEG, 1 archivo en MOBI, 200 JPEGs, 1 ePub, 1 archivo de Adobe Illustrator, 4 archivos en x-msvideo.

Formatos-Mime Types



Objetivo de la evaluación

La presente evaluación va a analizar la estructuración de contenidos bajo el criterio de usabilidad específico que es la accesibilidad:

- **Content** (del RI)
  - Accesibilidad de los contenidos para la comunidad designada de usuarios.

Qué verificar/medir/chequear: Metadatos

1) Sobre el perfilamiento y el registro PRONOM próximas hojas.

2) La información descriptiva de preservación (PDI) para OAIS debe brindar datos suficientes sobre la **procedencia (provenance)**, el **contexto (context)**, la **referencia (reference),** la **integridad (fixity) y los derechos de acceso (access rights)**.

3) La información de empaquetado de los objetos deberá seguir las indicaciones de la Norma OAIS. Se anotan a continuación algunos puntos destacados de la Norma al respecto de esto que se encuentran en el Capítulo 4.2 de la Norma titulado “Modelo Funcional”.

4) Metadatos descriptivos: Las Directrices Driver, desarrolladas en el ámbito del proyecto DRIVER.

¿Qué hacer?

- Con la experimentación propuesta en los cuatro pasos precedentes se construirá la evaluación, la cual deberá además brindar un reporte adecuado y una recomendación de acciones a seguir en el caso de objetos que no cumplan con los requerimientos.

Extensión de Driver

Si bien las directrices Driver están dadas pensando en recursos textuales, se van a extender aquí a recursos no textuales que brinda el RI, es decir para todos los recursos se verificará si cumplen con todos los metadatos que pide Driver para la interoperabilidad de los repositorios. Se podrían mirar también las Directrices OpenAIRE que sí son extensibles, pero el hecho que se use un subconjunto de metadatos descriptivos, se justifica por el Modelo OAIS en sí mismo, que dice que se refiere *"a la información descriptiva necesaria para la localización".*

Directrices Driver

Basic element	Status	Encoding schemes
Title	M	None, free text
Creator	M	APA bibliographic writing style as in a reference list. Syntax: surname, initials (first name)
Subject	MA	Choice of keywords and classifications can be free text (preferably in English) and defined by an URI scheme (preferably info:eu-repo/classification)
Description	MA	None, free text. Recommended practice is to include an abstract in English. "Abstract" is the default interpretation to the value for dc:description
Publisher	R	None
Contributor	O	APA bibliographic writing style as in a reference list. Syntax: surname, initials (first name)
Date	M	Date | ISO 8601 W3C-DTF - "Published" is the default interpretation to the value for dc:date
Type	M	Publication type and Version type can be free text (preferably in English) and defined by an URI scheme (preferably info:eu-repo/semantics).
Format	R	IANA registered list of Internet Media Types (MIME types)
Identifier	M	URI scheme, linking to persistent identifier (URN, handle, DOI), full text document or human start page.
Source	O	Guidelines for Encoding Bibliographic Citation Information in Dublin Core Metadata as in dcterms:bibliographicCitation
Language	R	ISO 639-3
Relation	O	None
Coverage	O	"Period" is the default interpretation to the value for dc:coverage Encoding: DCMI Period [http://dublincore.org/documents/2000/07/28/dcmi-period/] For more ncoding schemas see Chapter 5 Use of vocabularies and semantics.
Rights	R	None
Audience	O	None. "Eduction level" is the default value for dc:audience.

Qué tomar de Driver

Título, Creador, Fecha, Tipo e ~~Identificador~~ (este último será parte de la PDI, entonces no vuelve a considerarse aquí). Los elementos Description (en el archivo DC XML de SEDICI) y Subject (en el metadata sedici xml) ya se tienen y serán considerados, manteniendo el criterio de que por esos dos (aunque no estén) no se evalúe negativo, pero tal vez sea importante chequear qué porcentaje lo tienen y cuántos no.

Se atenderá además al punto de Driver: Normalización de descripción: Interoperabilidad semántica

p.e. Creator* y Contributor: sintaxis: Apellido, iniciales (primer nombre) APA. bibliographic writing style as in a reference list.

Driver y algo más

•En algunos casos resulta obvio pensar si se acepta o no el objeto digital, por e.j. si no tiene un identificador persistente; en otros aspectos esto se hace más problemático, por ejemplo qué eventos se van a considerar imprescindibles.

•PREMIS recomienda que se registren siempre las acciones que modifican un objeto y la entrada para el tipo de acontecimiento (*eventType*) del Diccionario de Datos proporciona una "lista de entrada" de tipos de acontecimientos importantes con el objetivo de promover que los repositorios registren estos acontecimientos de manera consistente. Esa será la guía que se utilizará para determinar la validez o no del objeto en cuanto a los eventos.

¿Cómo aplicar la metodología?

1) Perfilamiento automático del repositorio y reporte. Herramienta: DROID.
Debido a la necesidad/posibilidad de generar tareas de curation para los ítems que no cumplan con todos los requisitos, se hace necesario conocer sobre qué ítems deberán realizarse las tareas, dado que DROID no muestra sus reportes para objetos individuales sino en global, se realizarán consultas directas para contrastar los resultados y completar las acciones. **Ver archivo de resultados**.

2, 3 y 4) Revisión de la PDI de SEDICI, revisión de la Información de empaquetado y de los metadatos descriptivos y reporte que de cuenta de la adecuación o no con los patrones establecidos: PDI de OAIS, Información de empaquetado de OAIS y metadatos descriptivos según Directrices Driver. Validador según desarrollo propio.

¿Cómo aplicar la metodología?

5) Reporte final basado en los reportes parciales de 1,2,3,4.

6) Recomendación de acciones:

6.1) Formato antiguo - herramienta de migración automática (CRIB, Preservation **action services** de la Universidad de Minho y/o el workflow de Taverna y MiniMEE la herramienta de migración de Plato).

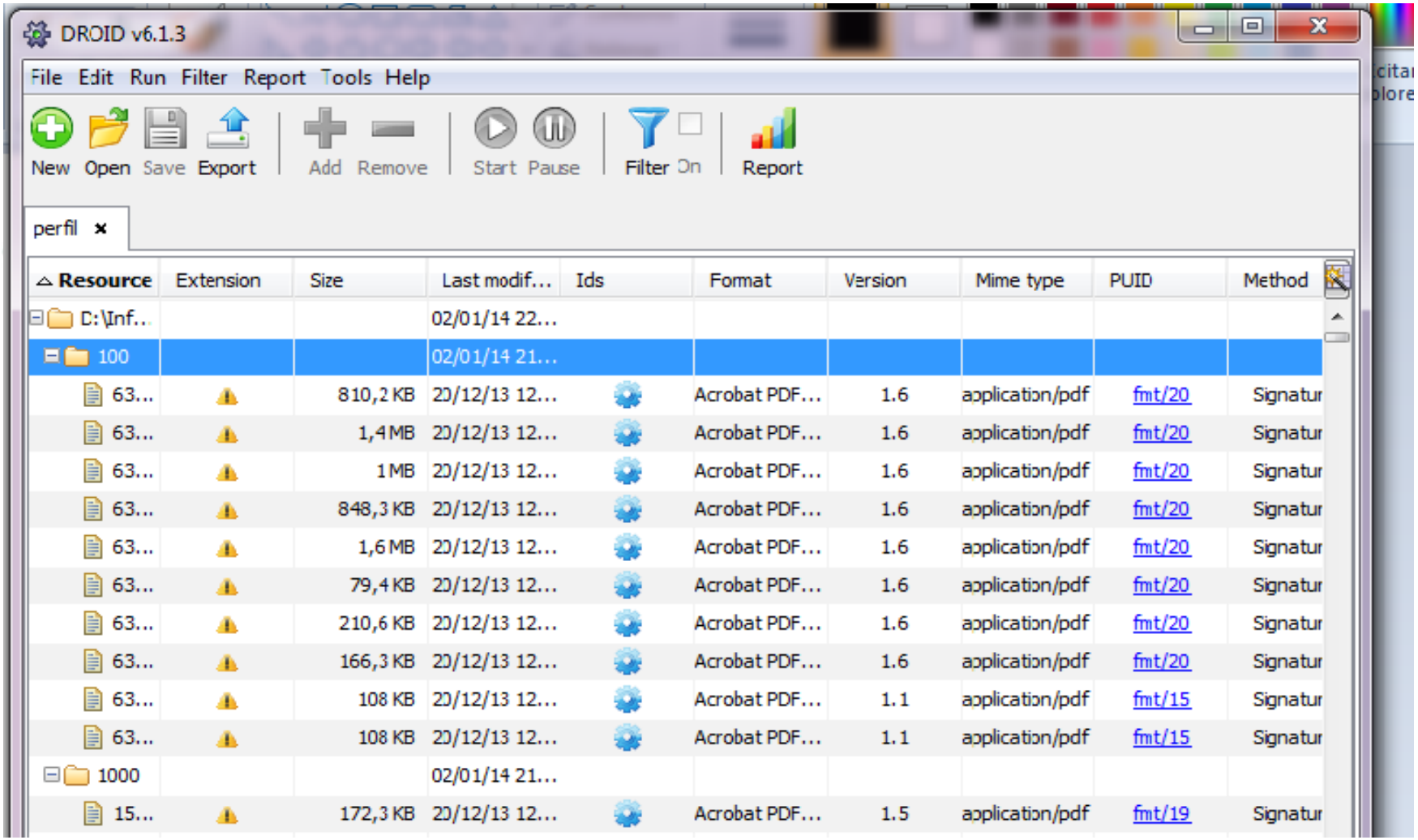
6.2) PDI defectuosa---> tarea de curation (automática) generada en SEDICI.

7) Propuesta de un Plan de Preservación para SEDICI cuya herramienta de creación es Plato (en experimentación)

Experimento

Se analizaron 19230 archivos de un AssetStore completo de Diciembre de 2013. De entre esa totalidad de archivos 18522 son archivos reales y el resto folders. Filtrados los folders, se caracterizan los archivos por su formato PUID (Persistent Unique Identifier) y se analiza el riesgo, como tarea colateral se observa si hay repetición en el checksum (MD5) y de acontecer esto se revisan los casos. Luego, con los casos sospechosos se generan las tareas de análisis, revisión y corrección para que los administradores de SEDICI resuelvan el problema.

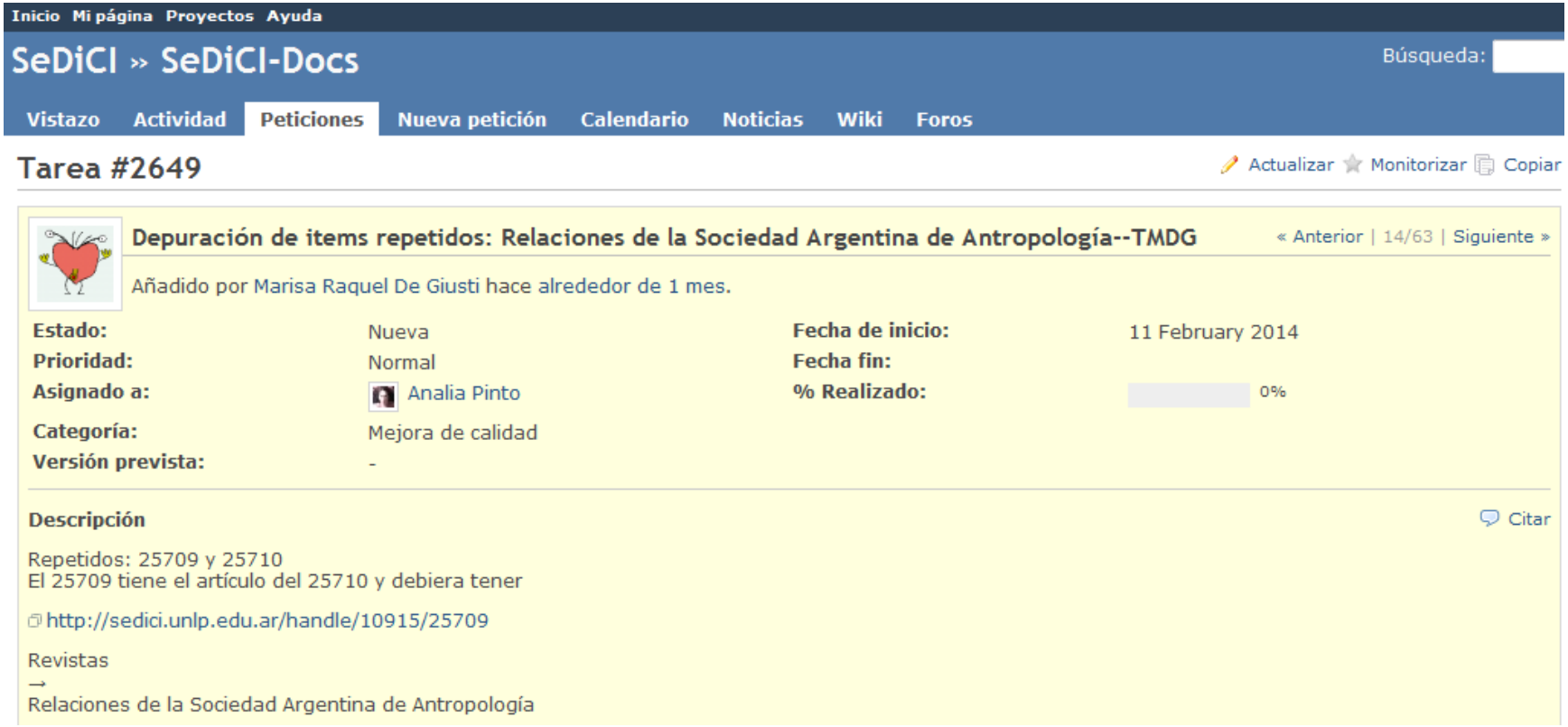
Perfil en DROID



Perfil exportado con MD5 para detección de duplicados

1801	15757	3328885	eaf803a90d5b190fd4b4935e6a852fc3	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1802	6355	175675	4564d6e8bc2f282435429c9560e841b1	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1803	6354	180887	cc0601b8518aa7d0addc22bab2925e1	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1804	6356	260190	6402659c609ce5bb4d22d585a596517	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1805	6361	105889	6b4f05561604f71946f3814ed2a313ee	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1806	6359	105889	6b4f05561604f71946f3814ed2a313ee	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1807	6360	105889	6b4f05561604f71946f3814ed2a313ee	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1808	15834	5596903	0d02c816f3b71e08c270dc332fc580a9	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1809	6362	196539	c45cc1c6dd997d059e318859ae77ee8	1	fmt/17	application/pdf	Acrobat PDF 1.3 - Portable Document Format	1.3
1810	6363	169421	6aa3fe9e89bcce079595fad164a99424	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	9c1db44a3a75b067398aacec160c247	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

Generación de tickets para mejora de calidad



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**

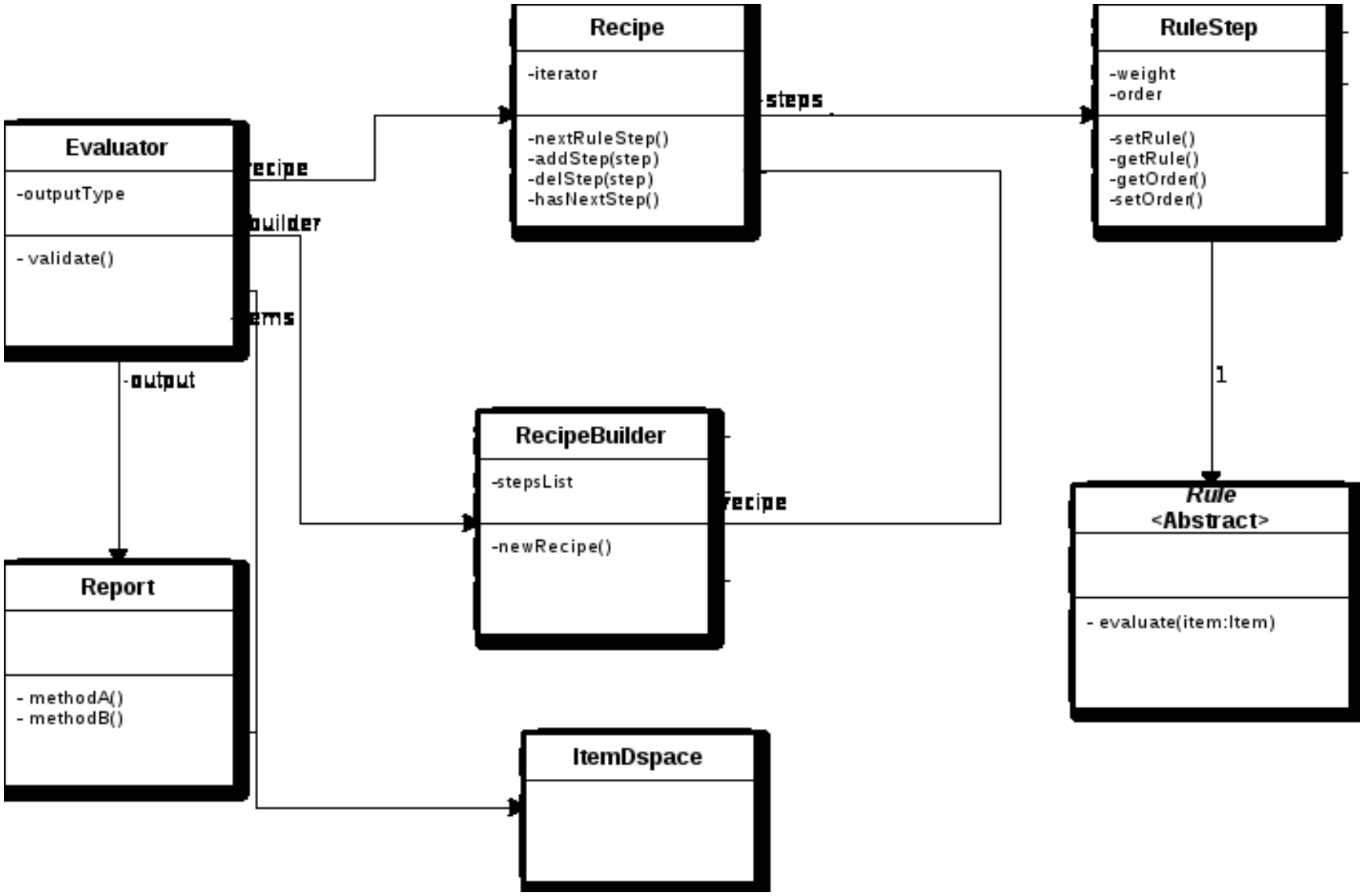
Group totals

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

Modelo del Validador Desarrollado



Plato 4.4

SCAPE

marisadg

Welcome to Plato 4.4! [06.03.2014 13:14] X

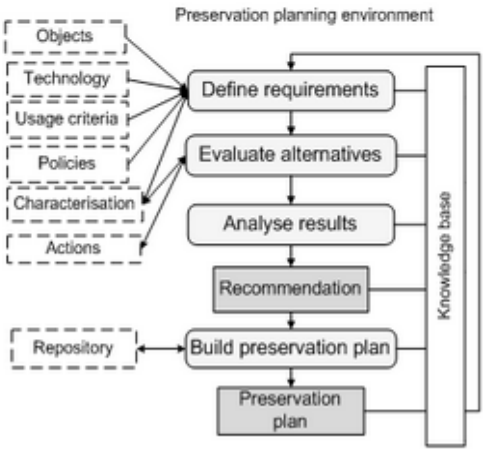
A copy has been created: *Plan for electronic papers - marisadg's copy of: This is an example plan. The project was created for the DELOS Summer School 2008 and revised afterwards. (originally created by admin)*
It is marked as playground. If you want to use it for serious planning, please change this in Plan Settings. [11.03.2014 12:59] X

How to start?

1. Take a look at the definition of the preservation plan at the documentation page,
2. Read through the description of the preservation planning procedure (below), and then
3. Create a *demo plan* in the list of plans, and walk through the steps to familiarise yourself with the procedure and tool.
4. If you have any questions, comments, or ideas, please let us know!

The planning procedure is completely supported by Plato, relying on a variety of information sources and services. When you load a plan, you will find four menu items on the top which correspond to the four planning phases:

1. Define requirements,
2. Evaluate alternatives,
3. Analyse results, and
4. Define preservation plan.



Servicio de Difusión de la Creación Intelectual, Argentina :: sedici.unlp.edu.ar

Proyecto de Enlace de Bibliotecas
prebi.unlp.edu.ar

SERVICIO DE DIFUSIÓN DE LA CREACIÓN INTELECTUAL
sedici.unlp.edu.ar

UNIVERSIDAD NACIONAL DE LA PLATA
www.unlp.edu.ar

¡Muchas gracias!

Marisa R. De Giusti marisa.degiusti@sedici.unlp.edu.ar
Ariel J. Lira alira@sedici.unlp.edu.ar
Gonzalo Villarreal gonzalo@sedici.unlp.edu.ar
Agustín Terruzzi agustinterruzzi@sedici.unlp.edu.ar
Facundo Adorno facundo@sedici.unlp.edu.ar

Portal de Revistas Científicas de la UNLP

revistas.unlp.edu.ar/cientificas/

PORTAL DE **REVISTAS**
PORTAL DE **CONGRESOS**

revistas.unlp.edu.ar
congresos.unlp.edu.ar

Puede acceder a esta presentación y a mucho más material del SEDICI en http://sedici.unlp.edu.ar/handle/10915/25293