

Las Actividades y el Planeamiento de la Preservación en un Repositorio Institucional

De Giusti, Marisa R.

Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata (SeDiCI), Argentina. Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC), Argentina
marisa.degiusti@sedici.unlp.edu.ar

Lira, Ariel J.; Oviedo, Néstor

Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata (SeDiCI), Argentina
{alira, nestor}@sedici.unlp.edu.ar

Villarreal, Gonzalo L.

Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata (SeDiCI), Argentina. CONICET, Argentina
gonzalo@sedici.unlp.edu.ar

Texier, Jose

Universidad Nacional Experimental del Tachira (UNET), Venezuela

Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata (SeDiCI), Argentina. CONICET, Argentina
jtexier@unet.edu.ve

Resumen:

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 desarrollo de los repositorios institucionales, el crecimiento de sus contenidos y el reconocimiento de que la actividad institucional se canaliza, cada vez más, en soporte digital, obliga a los repositorios a acompañar su desarrollo con actividades destinadas a la preservación. En este trabajo se presentan el estándar 14721 (OAIS), los metadatos PREMIS y las directrices para la preservación, en conjunto con el esquema METS, para finalmente, explorar los metadatos en esquemas muy utilizados en la tarea normal de un repositorio (MODS, DC) y señalar los que resultan útiles a los fines de la preservación, proponiendo su reutilización. Un segundo objetivo práctico es mostrar qué herramientas de preservación ofrece el desarrollo DSpace que sustenta al repositorio SeDiCI - UNLP.

Palabras Clave: preservación digital, repositorios, esquemas de metadatos, software para la gestión de repositorios, Dspace, OAIS, PREMIS.

Abstract

Digital preservation is defined as a set of political and strategic practices and concrete actions deployed to ensure long term access to digital objects.

The development of institutional repositories, growth of their content and awareness that institutional activity is increasingly channeled through digital media require the deployment of preservation activities to sustain the evolution of repositories. This work presents ISO standard 14721 (OAIS), PREMIS metadata and guidelines for preservation, along with the METS schema to finally explore metadata in widely-used repository-related schema (MODS, DC) and point out those useful for preservation purposes, proposing their reuse. A second practical purpose is to show the preservation tools offered by the DSpace development which supports the SeDiCI - UNLP repository.

Keywords: digital preservation, repositories, metadata schema, software for repository management, DSpace, OAIS, PREMIS.

Introducción

En este trabajo se presenta una propuesta de naturaleza técnico-práctica cuyo objetivo es proponer un plan inicial de desarrollo de una estrategia de preservación de contenidos de un repositorio institucional. Comienza con conceptos básicos y luego se introducen estándares de preservación, haciendo hincapié en los metadatos que soportan y documentan el proceso de preservación digital.

La propia naturaleza de los objetos digitales hace que sean dependientes de tecnologías que cambian constantemente. La incompatibilidad entre sistemas nuevos y antiguos sumado a que los formatos, medios de soporte, software y hardware quedan obsoletos en poco tiempo, hace necesario documentar a través de metadatos técnicos cómo acceder a los objetos. Los objetos digitales cambian, y dichos cambios deben registrarse y validarse, 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.

Finalmente, debido a que algunos problemas en lo relativo a la preservación, derivan de una parametrización deficiente del desarrollo que sustenta un repositorio, se cotejan las facilidades de DSpace con el modelo de preservación OAIS. En el directorio de repositorios ROAR (<http://roar.eprints.org/>), DSpace representa el 46,37% de los repositorios registrados y en el directorio OpenDOAR (<http://www.opendoar.org/>) hay un 40,60% de repositorios registrados que usan DSpace, datos tomados en octubre del 2012 con lo cual se espera sea de utilidad.

Sección 1: Conceptos básicos

Un Repositorio [1], [2] es una infraestructura web capaz de brindar 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 no es parejo; mientras la gestión y la difusión de contenidos están muy avanzadas, la recopilación es compleja pues involucra a una comunidad que no conoce el acceso abierto y tiene dudas sobre sus derechos y los permisos que puede delegar; la preservación es más reciente, y no es sustentada por todas las plataformas o no se conoce el potencial de la herramienta.

En un repositorio se debe asegurar la **interoperabilidad** [3], [4], es decir, la capacidad de sus sistemas y servicios para trabajar en conjunto intercambiando datos con otros de manera transparente. El basamento técnico, yace en el uso de estándares abiertos para la interacción y en la descripción de los recursos para asegurar su reutilización desde herramientas de **búsqueda y acceso**. La Iniciativa de Archivos Abiertos (<http://www.openarchives.org/>) desarrolla normas de interoperabilidad que facilitan la difusión eficaz de los contenidos.

Sección 2: Directrices de la Norma ISO 14721

2.1. Modelo para la Preservación

En el año 2003 se publicó la Norma ISO:14721 [5], denominada *ISO Reference Model for an Open Archival Information System (OAIS)*, que modela las partes que componen un Sistema Abierto de archivo de información, como por ejemplo un repositorio institucional, sin definir su implementación. La Figura 1 muestra las seis entidades funcionales y los paquetes de información del modelo. Las entidades funcionales son descriptas en detalle en la Tabla de la Sección 4.2.

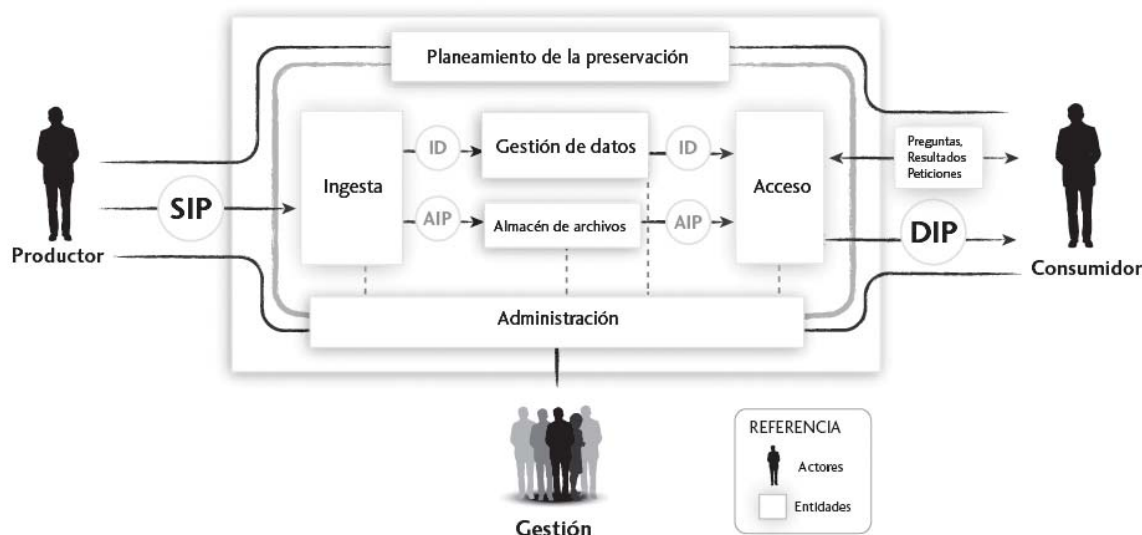


Figura 1. Modelo OAIS. Fuente: ISO 14721:2003

Las dos funciones principales del modelo OAIS son conservar la información y garantizar el acceso a la misma. En la figura 1, se observan claramente los tres actores y las seis entidades. El proceso puede iniciarse cuando el productor (un actor) suministra el recurso (paquete de entrada) llamado SIP a través del *ingest*, que luego se convierte en AIP terminando en la entidad de *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 ha sido preservada en el sistema previamente. 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 ingest, *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. 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.

El elemento central de la norma es el **Information Package (IP)** que refiere al conjunto que conforman el objeto digital y todos sus metadatos. El IP puede subclasificarse en tres subtipos según su función en el proceso de archivo:

- 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.
- 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.
- Dissemination Information Package (DIP): es el paquete que se entrega a un Consumidor en respuesta a una solicitud.

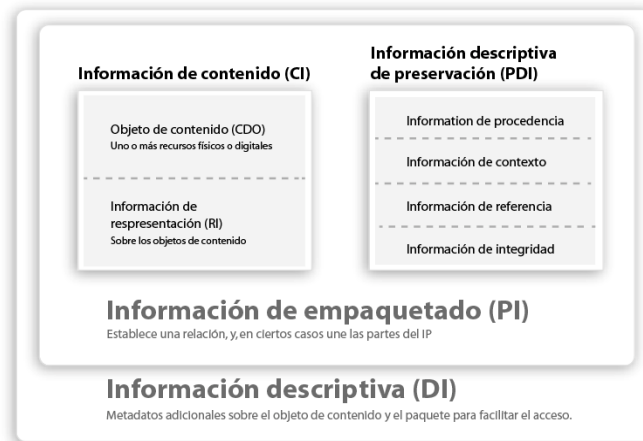


Figura 2. Estructura del Paquete de Información

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** y la **integridad**. 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 o no. La figura 2 se complejiza para poner de manifiesto las partes constitutivas del paquete de información así como la información descriptiva que debe guardar el OAIS para su búsqueda y recuperación. El proceso de ingesta (*ingest*) a través del generador de AIP, produce la información descriptiva que es enviada al módulo de gestión de datos (*data management*) que se encarga de proveer los servicios para poblar, mantener y dar acceso a la información descriptiva que identifica y documenta los archivos y los datos administrativos que se utilizan para administrarlos. 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.

2.2. Aproximaciones a la Preservación

Existen numerosas estrategias para asegurar la preservación de la información y evitar los problemas de obsolescencia, entre los que podemos mencionar: migración, encapsulamiento, emulación, políticas de backups, y más.

La “Guía Unesco: Directrices para la preservación del Patrimonio Digital” [6], 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.

Tránsito Ferreras Fernández [7] afirma que los esfuerzos para preservar los objetos digitales no se acaban con lo técnico, sino que existen otros: legales (permiso de los autores), económicos e institucionales (compromiso permanente) para llegar al resultado de perdurabilidad.

Catherine Lupovici y Julien Masanés [8] definen un conjunto mínimo de metadatos de carácter obligatorio para el propósito de preservación en un contexto de grandes volúmenes y tecnología variable. Proponen 8 metadatos y 38 sub-elementos siguiendo la taxonomía de OAIS. El último de

los elementos elegidos recoge la información de todos los cambios sucedidos en los previos. Proponen que la generación de metadatos se realice, en lo posible, de manera automática.

El proyecto denominado Repository preservation and interoperability (Preserv2) [9], se basa en el servicio de PRONOM provisto por The National Archives (TNA) y la herramienta DROID (Digital record object identification service) [10] que usa los perfiles de formato de más de 200 repositorios del registro PRONOM. Este último, 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. Para extender los servicios de identificación de formatos se utiliza un “*smart storage*” que combina un almacenamiento estándar con determinados servicios como el citado precedentemente y herramientas adicionales: programación basada en un calendario, historia de los eventos, etc.

JISC continuó el proyecto PReserv2 con el proyecto Keepit [11] para permitir gestión eficaz y mantenible a lo largo del tiempo de una amplia gama de contenido digital presente en los repositorios institucionales: artículos de investigación, datos primarios, artes, materiales didácticos y tesis.

2.3. Metadatos de Preservación

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:

1. Debe mantenerse en el repositorio de manera segura sin perderse ni ser modificado sin autorización.
2. Se debe conocer su creador.
3. Si cambia se debe conocer quién realizó el cambio.
4. Debe poder localizarse y entregarse al usuario.
5. Debe almacenarse en soportes que puedan leer los sistemas actuales de manera que el usuario pueda comprenderlos.
6. 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.
7. Soportar la autenticidad mediante la documentación de la *procedencia digital* a través de su cadena de custodia y el historial de cambios autorizados.
8. El repositorio debe disponer de los derechos suficientes como para llevar adelante las transformaciones necesarias para mantener el acceso al objeto.
9. Si el objeto está relacionado con otros del repositorio o de otros depósitos externos, estas relaciones deben guardarse.

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.

Muchos repositorios representan la información de preservación bajo PREMIS, METS o una combinación de ambos. En los siguientes inciso se realiza una breve descripción de cada uno.

2.3.1. PREMIS

PREMIS [12] 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 preservación en Archivos Digitales.

En 2008, este grupo elaboró el Diccionario de Datos PREMIS para Metadatos de Preservación [13], [14], el cual define los metadatos de preservación como “la información que utiliza un repositorio para dar soporte al proceso de preservación digital”.

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. Una gráfica muy aclaratoria (Fig. 3) que muestra lo que **no es** PREMIS [15]:

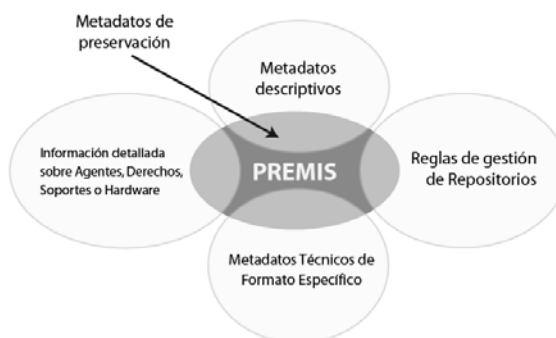


Figura 3. PREMIS como subconjunto de todos los metadatos de preservación.
Fuente: Caplan, Priscilla. "Entender PREMIS." (2009)

Los metadatos PREMIS se concentran sólo sobre los elementos que afectan directamente a la preservación, no mantiene vinculación con los metadatos de acceso, de recuperación de la información, ni siquiera con la información sobre derechos.

PREMIS plantea la necesidad de representar las unidades semánticas de forma abstracta, aunque no regula su implementación ni representación. Las unidades semánticas de PREMIS se corresponden con los metadatos definidos en el esquema PREMIS XML y pueden corresponderse parcialmente con algunos de de los formatos más usados.

2.3.1.1. Modelo de Datos PREMIS

Las entidades que este modelo define se pueden ver en el siguiente esquema:

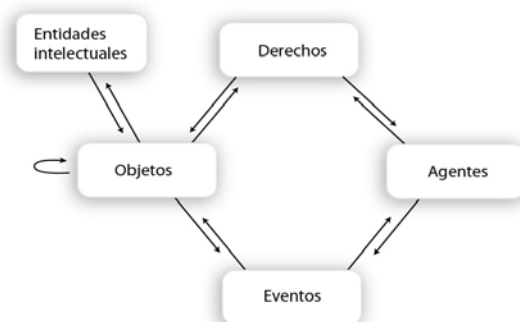


Figura 4. El modelo de datos PREMIS. Fuente: Caplan, Priscilla. "Entender PREMIS." (2009)

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.

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.

La entidad **Eventos** agrega información sobre acciones que un agente, o varios, lleva adelante sobre los objetos de los repositorios, el diccionario de datos aconseja como información: 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.

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).

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

2.3.2 METS

METS [16] está desarrollado por Network Development and MARC Standards Office de la Library of Congress. Es un formato que registra la estructura jerárquica de un objeto digital: nombre, archivos, ubicación, estructura y metadatos asociados.

Un documento METS posee un formato estandarizado para transmisión de metadatos archivos de datos, que se estructura XML. Eva Méndez lo define como una “estructura contenedora” extensible y modular que utiliza extensiones “wrappers” o “sockets” donde los elementos de otros esquemas pueden conectarse [17].

METS utiliza el esquema XML que le resulta útil para combinar distintos espacios de nombre.

Un documento METS consta de siete secciones:

1. Cabecera METS - contiene metadatos que describen el propio documento: identificación, creador, fecha de creación y agentes.
2. Metadatos Descriptivos: Esta sección puede: a) enlazar a metadatos descriptivos externos al documento METS; b) contener internamente los metadatos descriptivos posibles de registrar en formato XML o cualquier otro formato binario o textual, o c) combinar ambas modalidades.
3. Metadatos Administrativos: contiene metadatos técnicos con información relativa a la creación de los archivos que conforman al objeto digital, su formato y características de uso, sobre copyright e información sobre licencias, sobre el origen, y sobre la procedencia digital, es decir, sobre la relación entre el documento original y su representación digital, incluyendo la relación entre copias maestras y derivadas, migraciones y transformaciones.

4. Archivo: lista todos los archivos con contenidos que forman parte del objeto digital.
5. Mapa Estructural: recoge la estructura jerárquica del objeto digital, y enlaza sus secciones con los archivos de contenido y los metadatos correspondientes a cada una de ellas.
6. Enlaces Estructurales: permite registrar la existencia de hiperenlaces entre las secciones del mapa estructural.
7. Comportamientos: permite asociar el contenido del objeto con *comportamiento ejecutable*. Cada comportamiento posee un elemento que define su interface y un *mecanismo*, que identifica un módulo de código ejecutable que implementa y ejecuta los comportamientos definidos por la interface.

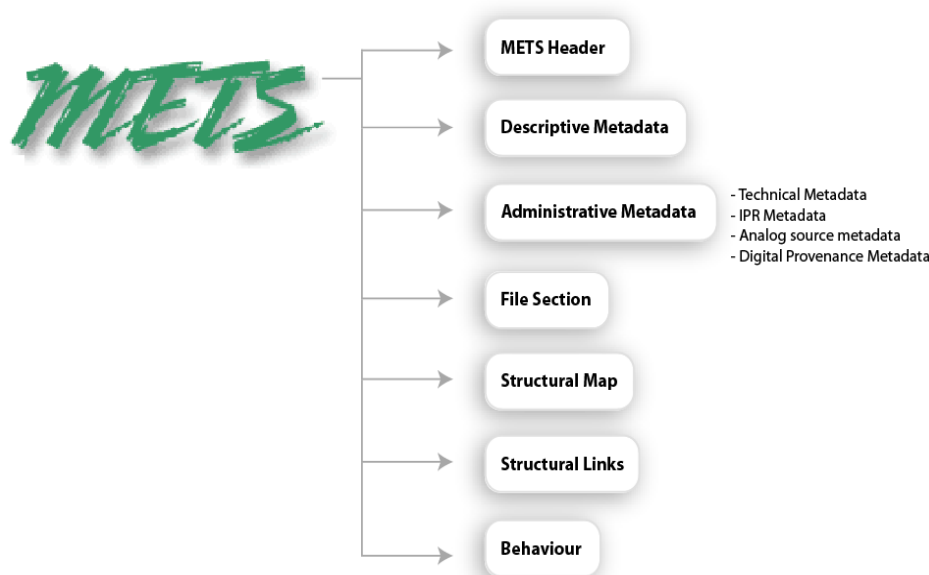


Figura 5. Estructura de un registro METS

Sección 3: Otros Esquemas de Metadatos y Sus Posibilidades en la Preservación

Si bien el par METS-PREMIS es muy adecuado a los fines de la preservación, los esquemas más utilizados de metadatos cuentan con metadatos útiles para ese cometido. Es importante tratar de detectar en el esquema usado por el repositorio, aquellos metadatos que resulten útiles para proveer datos técnicos sobre el formato, la estructura y el uso de los contenidos digitales, la historia de todas las acciones realizadas en los recursos, los cambios, la autenticidad, las responsabilidades y los permisos.

En algunos esquemas pueden surgir ambigüedades (algunas ampliamente referenciadas), sin embargo, una práctica más elaborada a la hora de incorporar los metadatos (por ejemplo el agregado de subelementos en el esquema DC) puede ayudar a la desambiguación. Es claro que los esquemas que no están dedicados específicamente a la preservación tienen limitaciones, aunque pueden ser un buen comienzo para contar con datos de preservación y además incorporar criterios para adoptar mejores prácticas.

A continuación se listan algunos metadatos de DC Simplificado [18] y MODS [19] que resultan útiles en la preservación según PREMIS [12] (Tablas 1 & 2):

DUBLIN CORE	PREMIS DD	DEFINITION
Identifier	Object Entity (1.1 objectIdentifier)	Unique reference to the resource in a given context.
Creator	Agent Entity	Responsible for creating the resource.
Description	Object Entity (1.10 relationship)	Summary of the resource.
Date	Object Entity (1.5.5 creatingApplication)	Time associated to an event in the resource's life cycle.
Format	Object Entity (1.5.4 format)	Physical or digital presentation of the resource.
Relation	Object Entity (1.10 relationship)	A reference to a related resource.
Rights	Right Entity (4.1 RightsStatements)	Information on the legal rights that affect the use of the resource.

Tabla 1. DC-PREMIS

MODS CORE Element	PREMIS Element	DEFINITION
Identifier	Object Entity (1.1 objectIdentifier)	Identifies the resource or work content.
name	Agent Entity	Associated in some way to the resource.
internetMediaType	Object Entity (1.5.4 format)	Physical or digital presentation of the resource.
relatedItem	Object Entity (1.10 relationship)	Summary of resource content.
location	Object Entity (1.7.1 contenLocation)	Identifies the institution or repository where the resource is or a remote location through a URL.
accessCondition	Right Entity (4.1 RightsStatements)	Information about restrictions imposed on access to a resource.

Tabla 2. MODS-PREMIS

Sección 4: OAIS y DSpace

Dado que cualquier repositorio asienta en gran parte su estructura, los formatos, los registros y del mismo modo la preservación en el software que utiliza, este trabajo considera el desarrollo DSpace que sustenta el repositorio SeDiCI - UNLP [20–22] y describe sus facilidades.

4.1. Descripción General

Dspace mantiene tres 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 [23]. 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 ítems.
- 2) Metadatos Administrativos: incluye metadatos de preservación, información de procedencia y permisos de acceso y uso sobre cada ítem, colección y comunidad. Se registran campos adicionales sobre varios elementos: tamaño, suma de comprobación y tipo Mime de cada archivo.
- 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.

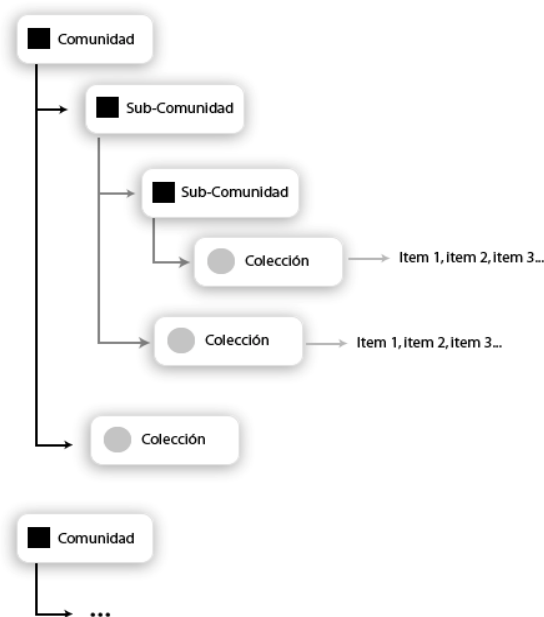


Figura 6. Jerarquía de comunidades y colecciones en Dspace.

Las *comunidades* y *colecciones* sirven a los fines administrativos para organizar los ítems dentro del repositorio, aunque en muchos repositorios se los considera como metadatos, dado que representan información estructural sobre los ítems.

La figura siguiente muestra el modelo de datos implementado por Dspace que se construye con los ítems y los bitstreams como conceptos fundamentales.

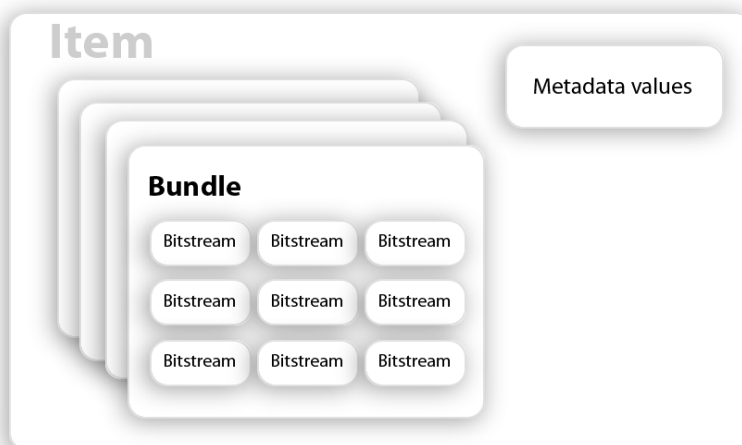


Figura 7. Estructura del ítem en Dspace.

Cada *ítem*, además de relacionarse con todos sus metadatos, está separado en paquetes (*bundle*) de secuencias de bits o archivos (*bitstreams*). Típicamente, se gestionan *bundles* para los recursos originales, Miniaturas, Licencia y texto extraído de los archivos, aunque es posible adicionar nuevos *bundles* de propósito específico. Para cada *bitstream* se registran datos propios del archivo físico como ser: su nombre, tamaño, suma de verificación y formato. Por último, los formatos poseen asociado un *nivel de soporte* (soportado, conocido y no conocido) que permite conocer al momento de la incorporación cuáles son las posibilidades de preservación a futuro en ese formato.

Inicialmente, todos los recursos que se crean en el repositorio son accesibles públicamente aunque

es posible regular completamente [24] qué usuarios o grupos de usuarios pueden acceder o gestionar cada bitstream, item, colección o comunidad.

Dspace registra algunos hitos sobre la vida de cada ítem almacenando en el metadato DC de procedencia `dc.description.provenance`. Para cada uno de los siguientes eventos de items registra la fecha y responsable que efectuó la acción: creación, restauración, retiro, aprobación, rechazo y modificación de bitstreams.

El proceso de exportación es configurable dado que posee módulos de software denominados *packager plugins* que permiten importar o exportar recursos. El *packager ingester* permite interpretar un paquete y crear un item, mientras que el *packager disseminator* sirve para armar el paquete. Para las transformaciones de metadatos, los *packager plugins* deben complementarse con los *Crosswalk plugins*, que se encargan de trasladar una representación en un formato de metadatos externo específico al interno del repositorio. Los *crosswalks* también se usan en el cosechador OAI-PMH y en el data provider OAI-PMH.

4.2 OAIS en DSpace

Dspace está influenciado por el modelo OAIS, en términos de metodología y funciones [25]. 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*).

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.

Como se mencionó en la sección anterior, Dspace permite exportar e importar paquetes DIP y SIP basados en METS [26], o, si se desarrolla un *packager plugin ad-hoc*, en cualquier otro formato. Adicionalmente, es posible importar y exportar AIPs [27] 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.

4.3 Tabla cruzada entre el modelo OAIS y las facilidades provistas por DSpace

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

<p>En DSpace: Se tiene un proceso claramente definido como <i>Ingest</i>, que esta compuesto por varios subprocesos. El mismo puede iniciarse por aportes externos al sistema o a través de la interfaz de usuario (vía web) del sistema. Se inicia un elemento <i>InProgressSubmission</i> que, dependiendo de la configuración, pasa por varios estadios durante los cuales se le agregan metadatos descriptivos y administrativos, y se realizan <i>checksums</i> a los contenidos. El resultado de esta secuencia es pasado al <i>item installer</i>, el cual debe convertir ese objeto en un <i>Archived Item</i> de DSpace.</p> <p>El <i>item installer</i> se encarga de asignar una fecha de adhesión, fecha de disponibilidad, de ser necesario fecha de publicación, procedencia, un identificador persistente, colección a la que pertenece, y crea los índices para su búsqueda y navegación.</p>
<p>Entidad OAIS <i>Archival Storage</i></p>
<p>Descripción: Provee los servicios y funciones para el almacenamiento, mantenimiento y recuperación de los AIPs.</p> <p>Recibe el AIP de la entidad <i>ingest</i> 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.</p> <p>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 un lugar físico separado. Provee copia de los AIPs almacenados a la entidad <i>access</i>.</p>
<p>En DSpace: provee dos tipos de almacén de datos para los bitstreams, uno dentro del sistema de archivos del servidor donde se aloja la aplicación y otro fuera, basado en SRB (Storage Resource Broker) que soporta el almacenamiento en forma distribuida y compartida. El resto de los datos del AIP son almacenados dentro de la Base de Datos relacional.</p> <p>Dspace ofrece la posibilidad de definir y ejecutar tareas de <i>curation</i> para realizar procesos desatendidos sobre los objetos del repositorio. Esto permite, por ejemplo, realizar validaciones periódicas, generar reportes, verificar los datos y los archivos e incluso realizar transformaciones.</p> <p>Para la verificación de los objetos almacenados en el sistema de archivos, provee una herramienta denominada checksum checker que compara la información de fijeza almacenada para el archivo con el el archivo propiamente.</p> <p>Permite efectuar tareas de duplicación sobre los ítems, colecciones y comunidades, generando un paquete AIP totalmente autocontenido. A continuación, los paquetes generados pueden ser restaurados en el almacén y usarse como reemplazo de los existentes.</p>
<p>Entidad OAIS <i>Data Management</i></p>
<p>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.</p> <p>Es responsable de la administración de la base de datos.</p> <p>Recibe solicitudes de la entidad <i>access</i> y genera un conjunto de resultados.</p> <p>Recibe pedidos de las entidades <i>ingest</i>, <i>access</i> y <i>administration</i> y genera reportes.</p> <p>También recibe actualizaciones de <i>ingest</i> y <i>administration</i>.</p>
<p>En DSpace: las funciones necesarias para almacenar y recuperar tanto la información descriptiva de los AIP como la información de soporte para las tareas administrativas (usuarios, permisos, etc), son provistas por los componentes asociados a la gestión de la base de datos.</p> <p>Asimismo, cuando se utiliza el módulo <i>Discovery</i> como proveedor del servicio de búsquedas para la interfaz web, se agrega un indexador de texto (denominado Apache Solr) como parte de esta entidad funcional, aportando un modo alternativo y eficiente de acceso a los metadatos descriptivos.</p>
<p>Entidad OAIS <i>administration</i></p>
<p>Descripción: Provee los servicios y funciones para la operación global del sistema de archivos.</p> <p>Solicita la información necesaria sobre los archivos y negocia los acuerdos con los Productores.</p> <p>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 <i>data management</i> y recibe reportes de esa entidad. Sumariza todos los reportes y provee</p>

<p>información sobre performance del OAIIS e inventario y envía esta info a <i>preservation planning</i> para establecer políticas y estándares. Recibe los paquetes de migración para <i>preservation planning</i>. Recibe los pedidos de cambio, procedimientos y herramientas para la actualización del archivo. Responsable de enviar un pedido de diseminación a <i>access</i>, actualizando los contenidos de los DIP y resumiendo los SIP a <i>ingest</i>. 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. Verifica que los AIP y SIP suministrados sigan las especificaciones. Verifica el PDI según los usos de la comunidad designada. Revisar periódicamente los contenidos del archivo para determinar si los datos están disponibles. Crea/mantiene/borra las cuentas de acceso de los consumidores.</p>
<p>En DSpace: La incorporación de los SIP está sujeta al cumplimiento de sus partes con las restricciones definidas en la configuración; es posible marcar metadatos requeridos y para algunos tipos de datos, el dominio válido de un metadato. Las autorizaciones de cada usuario dentro del sistema pueden configurarse con un gran nivel de granularidad, permitiendo definir permisos de lectura, adición, revisión, modificación y borrado a nivel de bitstream, ítem, colección y comunidad. Por ello, los usuarios encargados de comunidades pueden mantener de control sobre su contenido en el OAIIS, aliviando la carga administrativa central. Según la tarea de administración a realizar, será preciso utilizar la aplicación Web o ejecutar un comando directamente desde la consola del servidor. Prácticamente todas las tareas administrativas se realizan vía Web, incluyendo la gestión de ítems con sus archivos y metadatos, la gestión de usuarios y grupos, comunidades y colecciones, y control de ítems dentro de los workflows de carga. Solo algunas tareas deben ejecutarse desde consola, concretamente las creación de administradores, importación de estructura de colecciones y comunidades, eliminación de embargos. El un módulo de estadísticas provee información de uso y acceso sobre las comunidades, colecciones, ítems y bitstreams por parte de los visitantes, junto a algunos eventos notables como ser el acceso de usuarios registrados.</p>
<p>Entidad OAIIS <i>preservation planning</i></p>
<p>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 <i>administration</i>. Desarrolla nuevos IP y planes de migración y prototipos, para implementar políticas y directivas de administración de IPs.</p>
<p>En DSpace: no posee una correspondencia en absoluto con esta entidad funcional. Se debe principalmente a que esta entidad posee un componente evolutivo fundamental. Sin embargo, Dspace provee facilidades para afrontar una gran variedad de cambios que se pueden llevar adelante, por ejemplo: aplicar transformaciones de formatos en forma semi-automática, definir procesos de diagnóstico de los AIP, aplicar cambios en lote sobre metadatos, entre otros.</p>
<p>Entidad OAIIS <i>access</i></p>
<p>Descripción: Proporciona una interfaz única de usuario para el acceso a la información de los archivos. Tiene 3 categorías, los <i>query requests</i>, los <i>result sets</i> y los <i>report requests</i>. Acepta los requerimientos de los paquetes de diseminación recuperados de los AIP de la entidad <i>archival storage</i> y transmite un <i>report request</i> al <i>Data Management</i> generando un DIP. Entrega las respuestas en línea y fuera de línea de los consumidores</p>
<p>En DSpace: Se proveen de aplicaciones web para búsqueda y navegación: JSPUI o XMLUI. En ambos casos se disponen de mecanismos de exploración basados en la base de datos interna que permiten navegar el árbol de comunidades, colecciones e ítems del repositorio. En XMLUI se provee una herramienta de búsqueda más completa y eficiente basada en Apache Solr, denominada Discovery</p>

[28]. También permite habilitar algunos servicios adicionales como ser OpenSearch, suscripciones a colecciones.

Se provee un módulo que cumple con el rol de Data Provider del protocolo OAI-PMH que soporta mapeos configurables hacia cualquier formato de metadatos, haciendo uso de los CrosswalkPlugins.

Para la generación de DIP, se proveen varias alternativas de exportación:

- archivo separado por comas CSV.
- exportación de AIP de comunidades, colecciones e items.
- exportación de items en un formato simple denominado Simple Archive Format
- Mets Dip a partir del packager exporter, que incluye algunos metadatos según el diccionario de datos PREMIS, aunque no todos dado que solo se implementa una porción del modelo de datos PREMIS.

Adicionalmente, es posible extender los mecanismo de exportación e incorporar nuevos ExporterPlugins para generar DIP en otros formatos.

Visualización de información estadística generada por el módulo correspondiente.

Tabla 3. OAIS - DSpace

Sección 5: Conclusiones y Trabajos Futuros

El presente trabajo está orientado a llamar la atención sobre la importancia de la preservación y describir algunos modelos e implementaciones útiles a los fines de asegurar que un objeto digital esté siempre accesible. Está claro que la preservación no se restringe a las cuestiones técnicas o prácticas como las mencionadas en este escrito, hay cuestiones vinculadas a lo institucional y relativas tanto al material que se va a preservar, como al soporte económico. Desde el punto de vista práctico hay cuestiones de enorme importancia vinculadas a los derechos, las cuales deben manejarse con sumo cuidado. Un trabajo a futuro que resulta de interés para la comunidad de repositorios es relevar el estado, en cuanto a la preservación, de una buena parte de los repositorios más importantes que figuran en los registros internacionales. Finalmente el gran trabajo pendiente del repositorio SeDiCI-UNLP es el avance en el plan de preservación del repositorio institucional, el cual está apenas comenzando a hacerse realidad.

Referencias

- [1] R. K. Johnson, "Institutional repositories: Partnering with faculty to enhance scholarly communication," *D-Lib Magazine*, vol. 8, no. 11, 2002.
- [2] JISC, "JISC infoNet - What is a Repository," 2012. [Online]. Available: <http://www.jiscinfonet.ac.uk/infokits/repositories/what>. [Accessed: 13-Jul-2012].
- [3] DRIVER, *DRIVER guidelines 2.0*. 2008.
- [4] COAR, "COAR » Repository Interoperability," 2012. [Online]. Available: <http://www.coar-repositories.org/working-groups/repository-interoperability/>. [Accessed: 13-Jul-2012].
- [5] CCSDS, *Reference Model for an Open Archival Information System (OAIS): ISO 14721*. 2002.
- [6] UNESCO, "Directrices para la preservación del Patrimonio digital | Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura," 2012. [Online]. Available: <http://www.unesco.org/new/es/communication-and-information/resources/publications-and-communication-materials/publications/full-list/guidelines-for-the-preservation-of-digital-heritage/>. [Accessed: 13-Jul-2012].
- [7] T. Ferreras Fernández, "Preservación digital en repositorios institucionales: GREDOS," 2010. [Online]. Available: <http://gredos.usal.es/jspui/handle/10366/83130>. [Accessed: 13-Jul-2012].
- [8] C. Lupovici, J. Masanès, C. Lupovici, and J. Masanès, *Metadata for the Long Term Preservation of Electronic Publications*. 2000.
- [9] JISC, "Preservation Services for Digital Repositories -PRESERV 2," 2012. [Online]. Available: <http://www.jisc.ac.uk/whatwedo/programmes/preservation/preserv2.aspx#>. [Accessed: 13-Jul-2012].
- [10] T. N. Archives, "PRONOM | Welcome," 2012. [Online]. Available: <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>. [Accessed: 13-Jul-2012].
- [11] JISC, "KeepIt: Kultur, eCrystals, EdShare (and NECTAR) - Preserve It!," 2012. [Online]. Available: <http://preservation.eprints.org/keepit/>. [Accessed: 13-Jul-2012].
- [12] PREMIS, "PREMIS: Preservation Metadata Maintenance Activity (Library of Congress)," 2012. [Online].

- Available: <http://www.loc.gov/standards/premis/>. [Accessed: 13-Jul-2012].
- [13] B. F. Lavoie, "PREMIS With a Fresh Coat of Paint," *D-Lib Magazine*, vol. 14, no. 5/6, May 2008.
- [14] PREMIS, *PREMIS Data Dictionary for Preservation Metadata -- version 2.0*. 2008.
- [15] P. Caplan, "Entender PREMIS," 2009.
- [16] METS, "Metadata Encoding and Transmission Standard (METS) Official Web Site," 2012. [Online]. Available: <http://www.loc.gov/standards/mets/>. [Accessed: 13-Jul-2012].
- [17] E. Méndez, "Metadatos para la preservación digital," 2008. [Online]. Available: http://www.rebiun.org/export/docReb/ws_2008_EvaMendez.ppt. [Accessed: 13-Jul-2012].
- [18] DCMI, "Dublin Core Metadata Element Set, Version 1.1," 2012. [Online]. Available: <http://www.dublincore.org/documents/dces/>. [Accessed: 13-Jul-2012].
- [19] MODS, "Metadata Object Description Schema: MODS (Library of Congress)," 2012. [Online]. Available: <http://www.loc.gov/standards/mods/>. [Accessed: 13-Jul-2012].
- [20] SeDiCI, "SeDiCI - Repositorio de la Universidad Nacional de La Plata," 2012. [Online]. Available: <http://sedici.unlp.edu.ar/>. [Accessed: 13-Jul-2012].
- [21] M. R. De Giusti, A. Sobrado, A. J. Lira, M. M. Vila, and G. L. Villarreal, "SeDiCI (Servicio de Difusión de la Creación Intelectual)," *D-Lib Magazine*, vol. 14, Sep. 2008.
- [22] M. R. de Giusti, M. M. Vila, A. Sobrado, and A. J. Lira, "SeDiCI: Servicio de Difusión de la Creación Intelectual," *Revista Interamericana de Bibliotecología*, vol. 31, no. 2, pp. 187–202, Dec. 2008.
- [23] R. Heery and M. Patel, "Application profiles: mixing and matching metadata schemas," *Ariadne*, 24-Sep-2000. [Online]. Available: <http://www.ariadne.ac.uk/issue25/app-profiles/>. [Accessed: 13-Jul-2012].
- [24] DSpace, "Functional Overview - DSpace 1.8 Documentation - DuraSpace Wiki," 2011. [Online]. Available: <https://wiki.duraspace.org/display/DSDOC18/Functional+Overview>. [Accessed: 13-Jul-2012].
- [25] R. Tansley, M. Bass, and M. Smith, "DSpace as an Open Archival Information System: Current Status and Future Directions," in *Research and Advanced Technology for Digital Libraries*, vol. 2769, T. Koch and I. T. Sølvsberg, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 446–460.
- [26] DSpace, "DSpaceMETSSIPProfile - DSpace - DuraSpace Wiki," 2011. [Online]. Available: <https://wiki.duraspace.org/display/DSPACE/DSpaceMETSSIPProfile>. [Accessed: 13-Jul-2012].
- [27] DSpace, "DSpace AIP Format - DSpace 1.8 Documentation - DuraSpace Wiki," 2011. [Online]. Available: <https://wiki.duraspace.org/display/DSDOC18/DSpace+AIP+Format>. [Accessed: 13-Jul-2012].
- [28] DSpace, "Discovery - DSpace 1.8 Documentation - DuraSpace Wiki," 2011. [Online]. Available: <https://wiki.duraspace.org/display/DSDOC18/Discovery>. [Accessed: 13-Jul-2012].

Otros Recursos Consultados

Websites:

- PADI Preserving Access to Digital Information, consulted on July 1, 2012: <http://pandora.nla.gov.au/pan/10691/20110824-1153/www.nla.gov.au/padi/index.html>
- ERPANET, consulted on July 1, 2012: <http://www.erpanet.org/index.php>
- The Library of Congress. Digital Preservation, consulted on June 15, 2012: <http://www.digitalpreservation.gov>
- METS: Introduction and Tutorial. http://www.loc.gov/standards/mets/METSOverview_spa.html
- Digital Preservation Coalition, consulted on June 4, 2012: <http://www.dpconline.org/graphics/index.html>
- UK National Preservation Initiatives, consulted on June 16, 2012: <http://www.ifla.org/IV/ifla68/papers/131-109e.pdf>
- Digital Library Federation, consulted on June 15, 2012: <http://www.diglib.org/preserve.htm>
- National Digital Preservation Initiatives, consulted on June 16, 2012: <http://www.clir.org/pubs/reports/pub116/pub116.pdf>
- CEDARS Curl exemplars in Digital Archives, consulted on June 16, 2012: <http://www.leeds.ac.uk/cedars/>
- DELOS Network of Excellence on Digital Libraries, consulted on June 16, 2012: <http://www.delos.info/>
- Project no. 033572. CASPAR: Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Instrument: Information Society Technologies. Document identifier: CASPAR-D2301-RP-0101-1_3. Submission Date: 20-02-2008, consulted on June 15, 2012: <http://www.casparpreserves.eu/>
- Other works on preservation: <https://wiki.duraspace.org/display/Dspace/RelatedWork>

Articles:

- Chazarra Bernabé, Juan and Requena López, Víctor Manuel and Valverde Jerónimo, Sergio (2010) *Desarrollo de un repositorio de objetos de aprendizaje usando DSpace*. Work in progress, <http://eprints.ucm.es/11078/>.
- C. Becker, H. Kulovits, M. Guttenbrunner, S. Strodl, A. Rauber, H. Hofman: *Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans*. International Journal on Digital Libraries, (2009), 4; S. 133 - 157. http://publik.tuwien.ac.at/files/PubDat_180752.pdf ;
- Hitchcock, Steve, Tim Brody, Jessie M.N. Hey and Leslie Carr (2007). *Survey of repository preservation policy and activity. Preserv project*, January 2007 <http://preserv.eprints.org/papers/survey/survey-results.html>.
- Hitchcock, Steve, Tim Brody, Jessie M.N. Hey and Leslie Carr (2005). *Preservation for Institutional Repositories: practical and invisible. Ensuring Long-term Preservation and Adding Value to Scientific and Technical data (PV 2005)*. Edinburgh, November 21-23, 2005 <http://www.ukoln.ac.uk/events/pv-2005/pv-2005-final-papers/033.pdf>.
- Hitchcock, Steve (2005). *Capturing preservation metadata from institutional repositories. DCC Workshop on the Long-term Curation within Digital Repositories*, Cambridge, July 6, 2005 <http://preserv.eprints.org/talks/hitchcock-dcccambridge060705.ppt>.
- Strodl, Stephan y Rauber, Andreas. *Preservation Planning in the OAIS Model*. Vienna University of Technology, Vienna, Austria. IPRES 2007. Beijing, China. http://www.digitalpreservationeurope.eu/publications/presentations/strodl_final_3.pdf
- Yuan Li y Meghan Banach. 2011. *Institutional Repositories and Digital Preservation: Assessing Current Practices at Research Libraries*. <http://www.dlib.org/dlib/may11/yuanli/05yuanli.html>