

13

ACCESO ABIERTO,
PRESERVACIÓN DIGITAL
Y DATOS CIENTÍFICOS

COSTA RICA - 18 AL 17 DE OCTUBRE DE 2013

III Conferencia Bibliotecas y Repositorios
Digitales de América Latina (BIREDIAL 2013)

VIII Simposio Internacional de Bibliotecas
Digitales (SIBD 2013)



SEDICI

REPOSITORIO INSTITUCIONAL DE LA UNLP



Taller

“Modelo OAIS y las posibilidades de un repositorio en DSPACE”



Índice

13

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



El repositorio

Un Repositorio 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 en un repositorio no es parejo: gestión y difusión avanzadas, recopilación compleja, preservación: reciente.

Interoperabilidad

Capacidad de un sistema de información para comunicarse con otros sistemas heterogéneos e intercambiar datos de manera transparente, con mínima pérdida de valor y de funcionalidad y a través de una interconexión automática y sin perder la interfaz propia.

Interoperabilidad

DRIVER: intenta llegar a la interoperabilidad en dos capas:

- *La sintáctica* (y estructural): directrices para gestores y administradores de repositorios sobre cómo exponer los recursos científicos digitales utilizando los metadatos Dublin Core y el protocolo OAI-PMH y homogeneizando los outputs
- *La semántica* a través de la utilización de vocabularios que permitan un significado común.

Claves: estándares abiertos para la interacción y descripción de los recursos para asegurar su reutilización desde herramientas de búsqueda y acceso heterogéneas.



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.

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



No sólo lo técnico

Tránsito Ferreras Fernández [2] : “los esfuerzos para preservar los objetos digitales no se acaban con lo técnico, sino que existen otros retos: legales (permiso de los autores), económicos e institucionales (compromiso permanente) para llegar al resultado de perdurabilidad”.

[2] T. Ferreras Fernández, “Preservación digital en repositorios institucionales: GREDOS,” 2010. [Online].



Problemas en la preservación

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

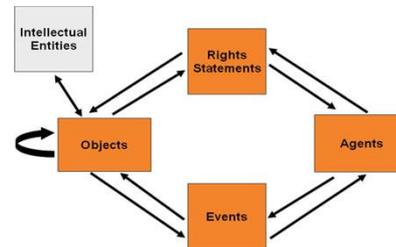
Preservación: motivos y tendencias

- Miles de objetos digitales sometidos a la obsolescencia tecnológica: fragilidad de los objetos digitales y obsolescencia de los medios informáticos
- La necesidad de metadatos apropiados para mantener el acceso a ese patrimonio.

Tendencias:

- ➔ Digital curation (ciclo de vida)
- Digital preservation (a largo plazo)
- Web archiving...

Aproximaciones a la preservación



Droid

DROID (Digital Record and Object Identification)

JISC KeepIT

The **technical registry**
PRONOM



Índice

Parte 1 - Conceptos básicos

Repositorio, interoperabilidad, preservación, guías, proyectos

13

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



Roles de los metadatos

- **Identificación y descripción de los recursos.**
- **Búsqueda, recuperación, ubicación.**
- Descripción de derechos patrimoniales y de uso.
- Preservación.
- Powder
- WAI
- Modalidades de acceso...

Metadatos y metadatos de preservación

Metadatos tradicionales

- 
- ID
 - Título
 - Autor
 - Idioma
 - Fecha
 - etc

documento

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)

OD y metadatos de preservación

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

Deben guardarse las relaciones que vinculen al objeto con otros

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

Autenticidad

Mediante la documentación de su procedencia

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



Debe conocerse su **Creador**

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

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

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

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.



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, como su nombre lo indica, se enfoca en estrategias de implementación de metadatos de preservación en Archivos Digitales. En 2008, este grupo elaboró el Diccionario de Datos PREMIS para Metadatos de Preservación el cual define los metadatos de preservación como “la información que utiliza un repositorio para dar soporte al proceso de preservación digital”.

Diccionario de datos PREMIS

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

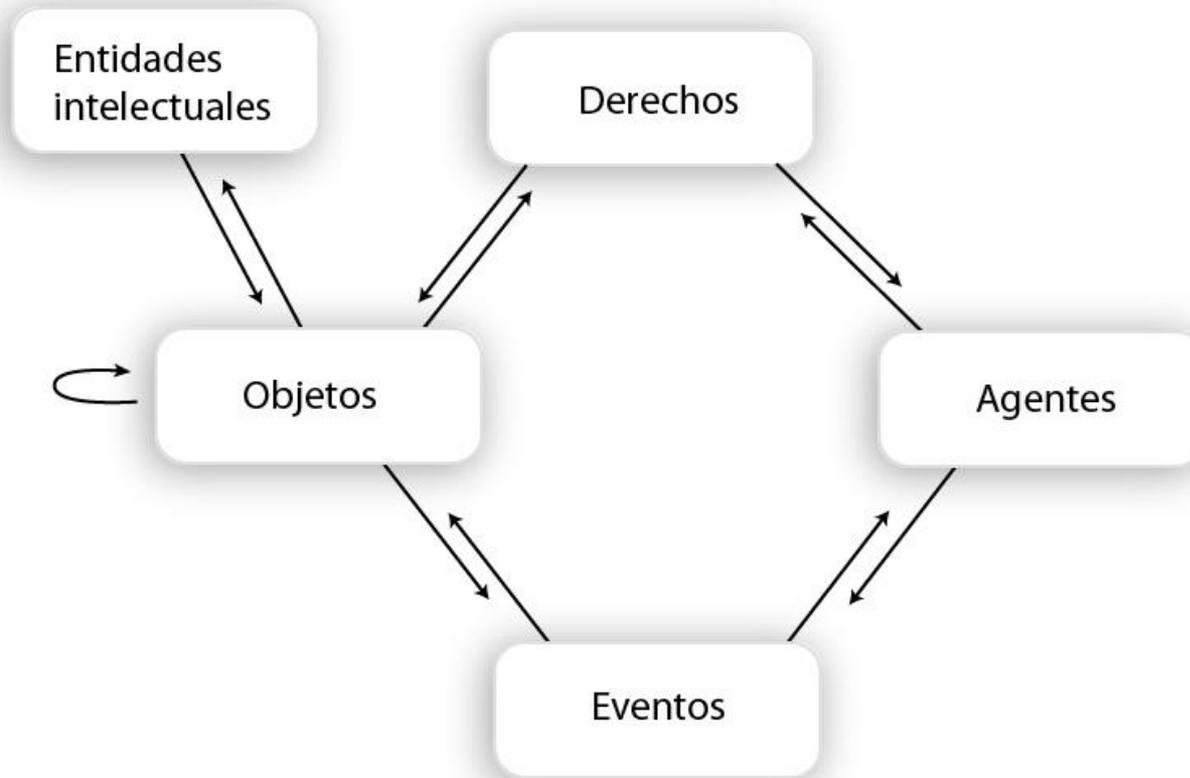
PREMIS plantea la necesidad de representar las unidades semánticas de forma abstracta, aunque no regula su implementación ni representación.

Modelo de Datos PREMIS

Las entidades que este modelo define se denominan:

- Entidades intelectuales
- Objetos
- Derechos
- Agentes
- Eventos

Modelo de datos PREMIS



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

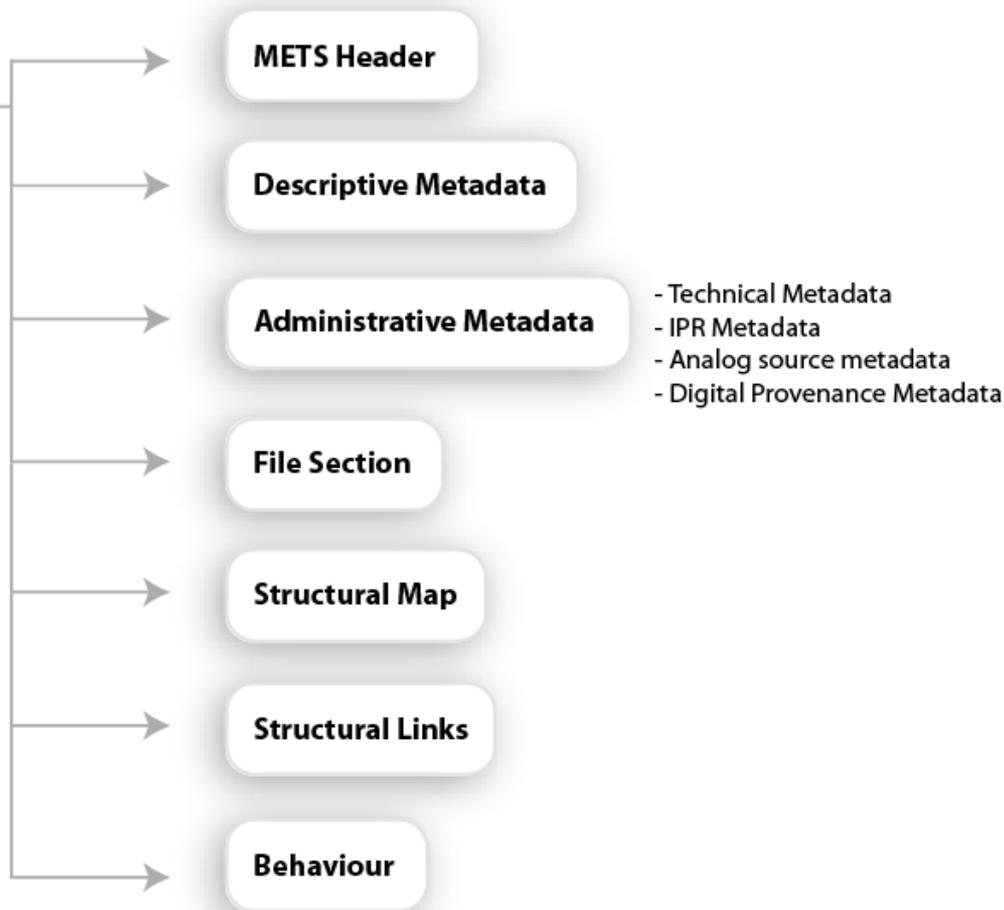
La entidad **Eventos** agrega información sobre acciones que un agente, o varios, lleva adelante sobre los objetos de los repositorios, por ejemplo: el identificador del acontecimiento (no repetible), el tipo (creación, migración, etc), la fecha de ocurrencia del evento, la descripción y el resultado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

Mets

METS



Otros esquemas y sus posibilidades

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.

Tratar de detectar en el esquema usado metadatos 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 permisos.

Otros esquemas y sus posibilidades

Si surgen ambigüedades (algunas ampliamente referenciadas), una práctica más elaborada a la hora de incorporar los metadatos, p.e. agregado de subelementos en DC ayuda a la desambiguación. Los esquemas que no están dedicados específicamente a la preservación tienen limitaciones, pero son un comienzo.

DC-PREMIS

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.

MODS-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.

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

13

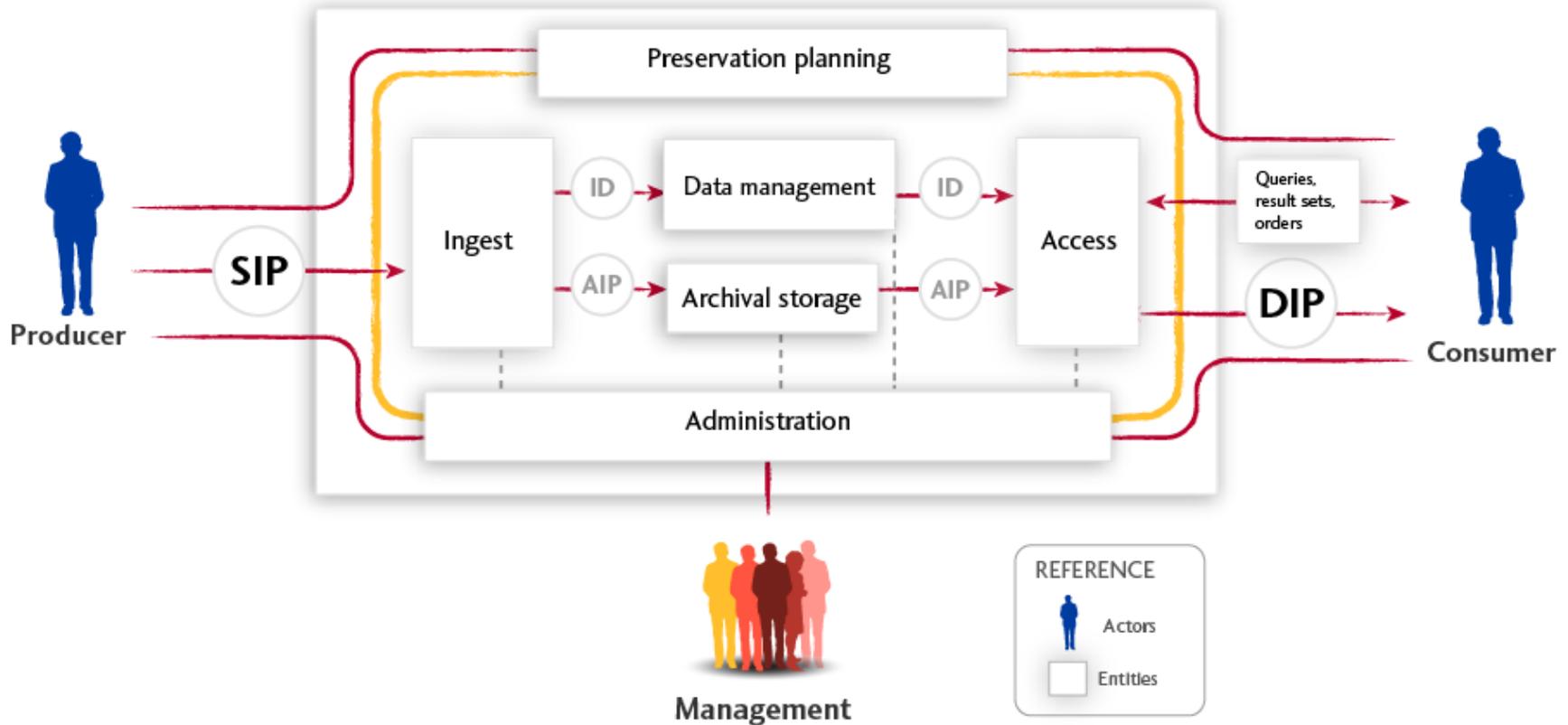
Parte 4- OAIS

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)



OAIS

Archivo que comprende una organización de personas y sistemas que han asumido el compromiso de preservar a largo plazo y hacer disponible un determinado corpus de información (cualquier tipo de conocimiento a intercambiar) para una comunidad designada.

Se refiere a la información analógica y a la digital, pero el foco está en esta última.

Open (abierto): se usa para indicar que esta recomendación ha sido realizada en foros abiertos. No significa que el archivo es de acceso gratuito o irrestricto. Puede ser cualquiera.



Secciones de la norma

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

OAIS

Sección 1

Justificación del modelo de referencia

Ninguna discusión sobre la conservación de repositorios y flujos de trabajo estaría completa sin al menos una breve introducción al modelo de referencia OAIS.

Una introducción a este modelo sirve para mostrar cómo implementa muchos de los procesos de flujos de trabajo y cómo se relaciona con la conservación digital.

Se recomienda como la mejor práctica actual, si bien no se ha establecido aún un consenso firme.

Funciones del Modelo de Referencia

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

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

Propósito y campo de aplicación

- Es aplicable para cualquier archivo, pero especialmente está enfocada en organizaciones con responsabilidad de hacer que la información esté disponible a largo plazo para una **comunidad designada**.
- Es de interés para aquellos que crean información que puede necesitar preservación a largo plazo, y aquellos que pueden precisar obtener información de tales archivos.
- No especifica un diseño o una implementación.
- El foco primario es la información inherentemente digital.
- Puede proveer servicios adicionales.
- El modelo se acomoda para información que no es inherentemente digital pero no se describe en detalle.

Propósito y campo de aplicación

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

OAIS

Sección 2

Conceptos en OAIS

El propósito de esta sección es motivar y describir varios conceptos clave, de alto nivel del OAIS. Un punto de vista más completo y una modelización formal de estos conceptos, se analiza en la Sección 4.2 de la norma pero no será expuesta en esta presentación.

Medioambiente y actores

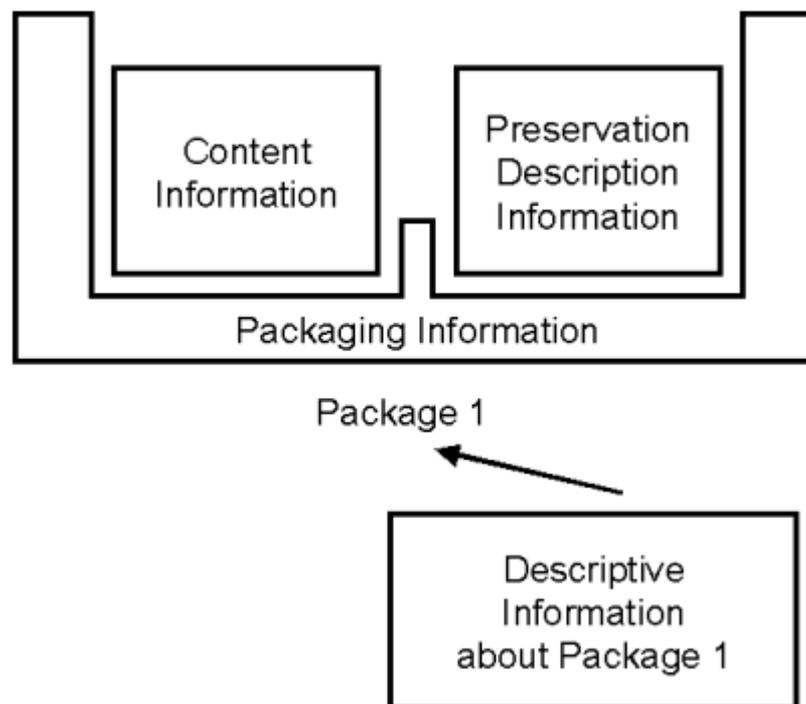
- Un productor que provee la información.
- Una política global de gestión (management), NO las operaciones diarias.
- La comunidad designada es el conjunto de los consumidores que son capaces de comprender la información preservada.

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

Paquete de Información: contenedor conceptual

Información de contenido (CI)

Objeto de contenido (CDO)
Uno o más recursos físicos o digitales

Información de
representación (RI)
Sobre los objetos de contenido

Información descriptiva de preservación (PDI)

Information de procedencia

Información de contexto

Información de referencia

Información de integridad

Información de empaquetado (PI)

Establece una relación, y, en ciertos casos une las partes del IP

Información descriptiva (DI)

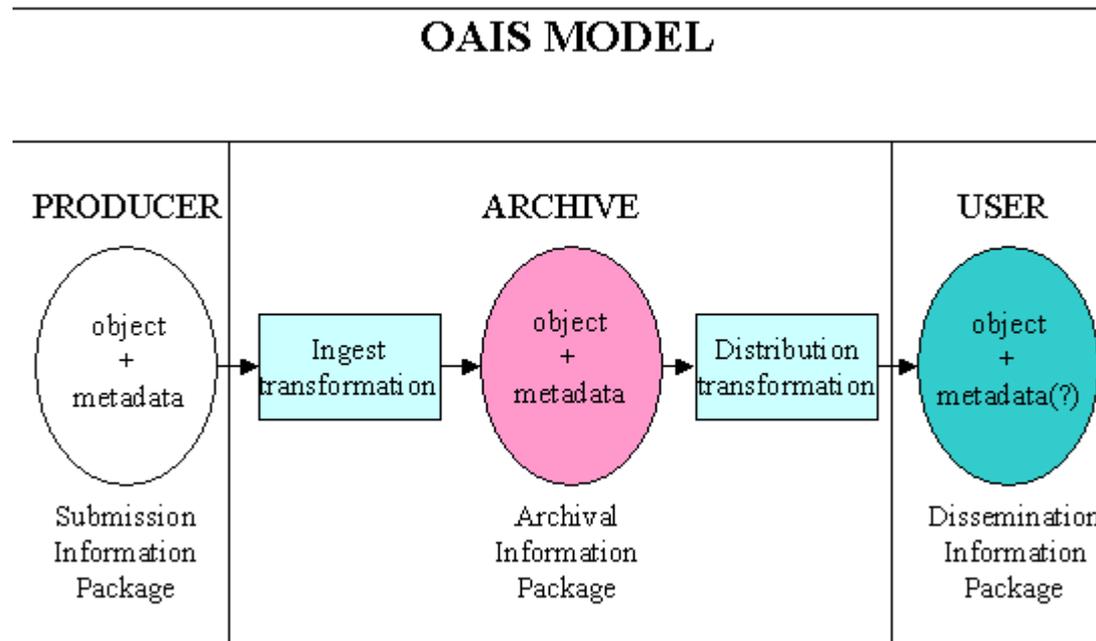
Metadatos adicionales sobre el objeto de contenido y el paquete para facilitar el acceso.

La información de empaquetado es la información que, ya sea real o lógicamente, une, identifica y relaciona la información del contenido y la PDI.

La información descriptiva es la información que se utiliza para descubrir qué paquete tiene la información de contenido de interés.

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 /no.

Variantes del IP: SIP, AIP y DIP



SIP: es el paquete que proviene del productor y se incorpora al OAIS. Suele contener menos información que el AIP.

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.

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

CCSDS RECOMMENDATION FOR AN OAIS REFERENCE MODEL

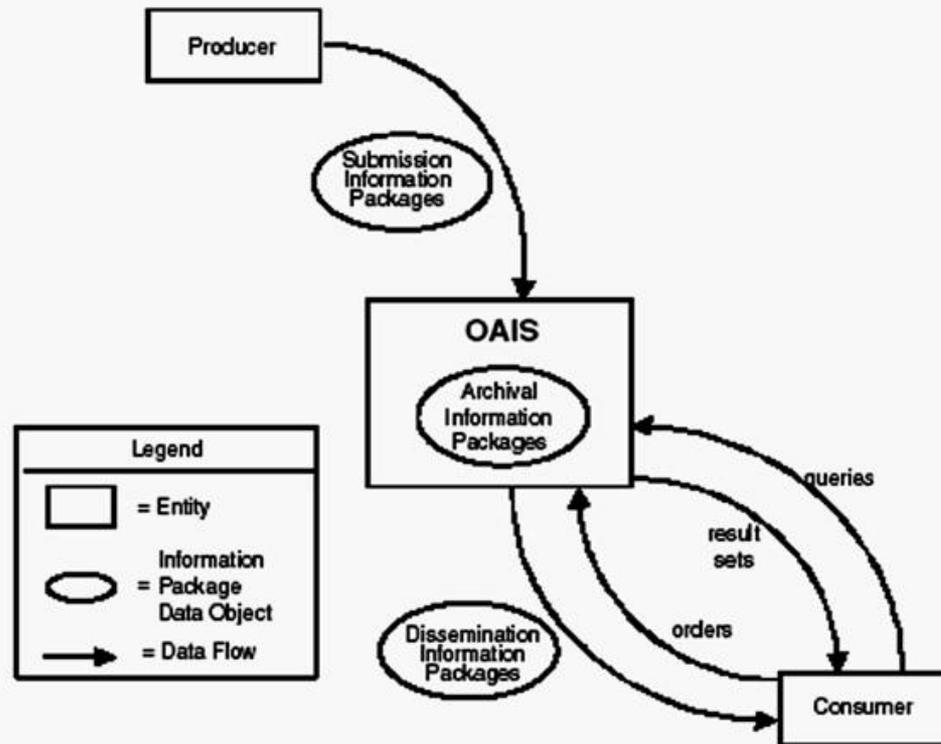


Figure 2-4: OAIS Archive External Data

Visión de alto nivel de las interacciones

Interacción de la gestión

- financiación, utilización de recursos, pagos, resolución de conflictos.

Interacción del productor

- los acuerdos de ingesta. Acuerdo por los SIPs que va a mandar, tiempo (acuerdo por data submission session)

Interacción de los consumidores

- Ayudas, descubrimiento de información, ordenamiento de la información. (Data dissemination session).

OAIS

Sección 3

Responsabilidades obligatorias

Negociar y aceptar la información adecuada de los productores de información.

Obtener un control de la información al nivel necesario para asegurar la conservación.

Determinar, por sí o con otras partes, cuáles comunidades deben convertirse en la comunidad designada y, entender la información.

Asegurar que la información que se conserva es comprensible.

Deslinde de responsabilidades: ejemplos

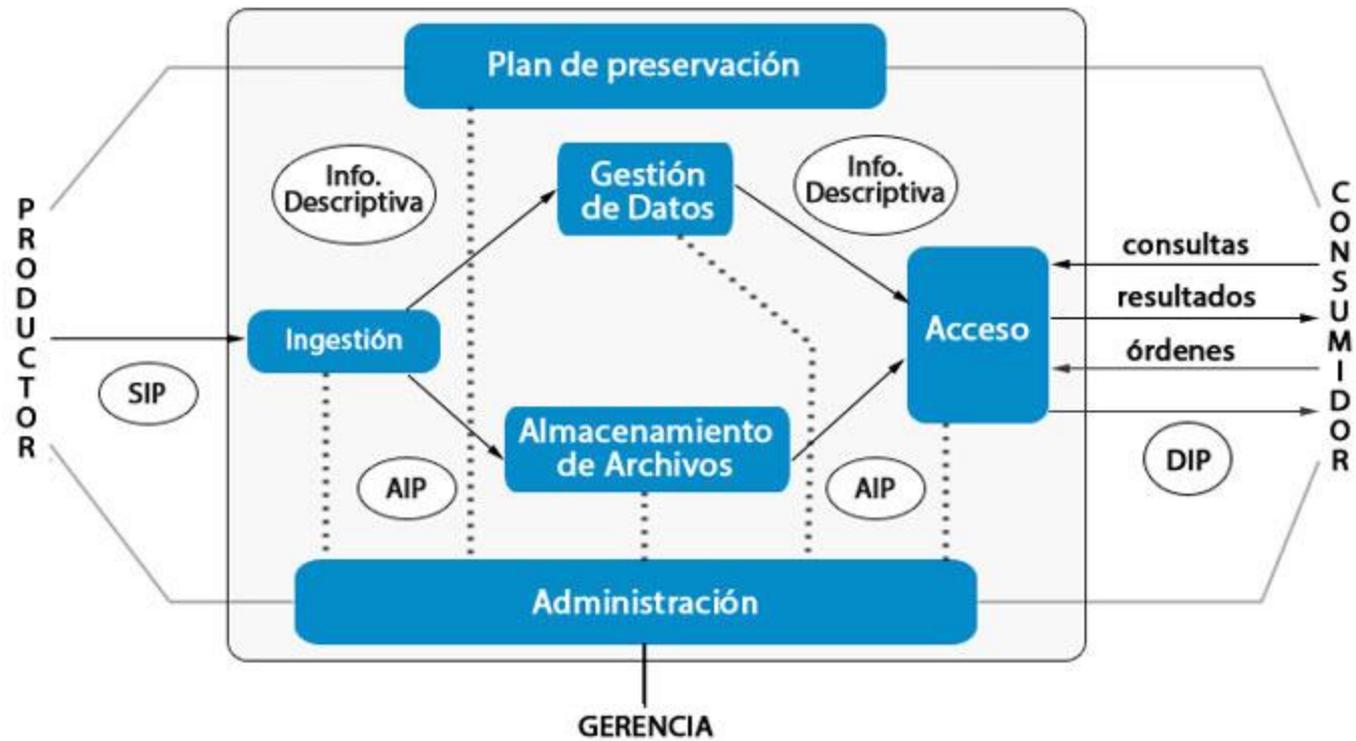
El control eficiente de la preservación significa:

- Atender cuestiones de copyright, propiedad intelectual y restricciones legales para el uso.
- Tener autoridad para modificar el modo de representación de la información.
- Realizar acuerdos con organizaciones externas (backups).

OAIS

Sección 4

Modelo funcional - Sección 4.1



Modelo OAIS

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

Modelo OAIS

Los datos relacionados con los documentos y el repositorio mismo se mantienen organizados a través de la entidad *data management*.

Luego hay una entidad *administration* dedicada a la administración adjunta a la gestión (administradores y responsable del repositorio) y esta entidad se relaciona con las secciones de *ingesta*, *gestión de datos*, *almacenamiento de archivos* y *planificación de la preservación*.

Esto permite una gestión estructural y ayuda a mantener los AIP a lo largo del tiempo.

Modelo OAIS

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

Ingesta/Ingest/presentación

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

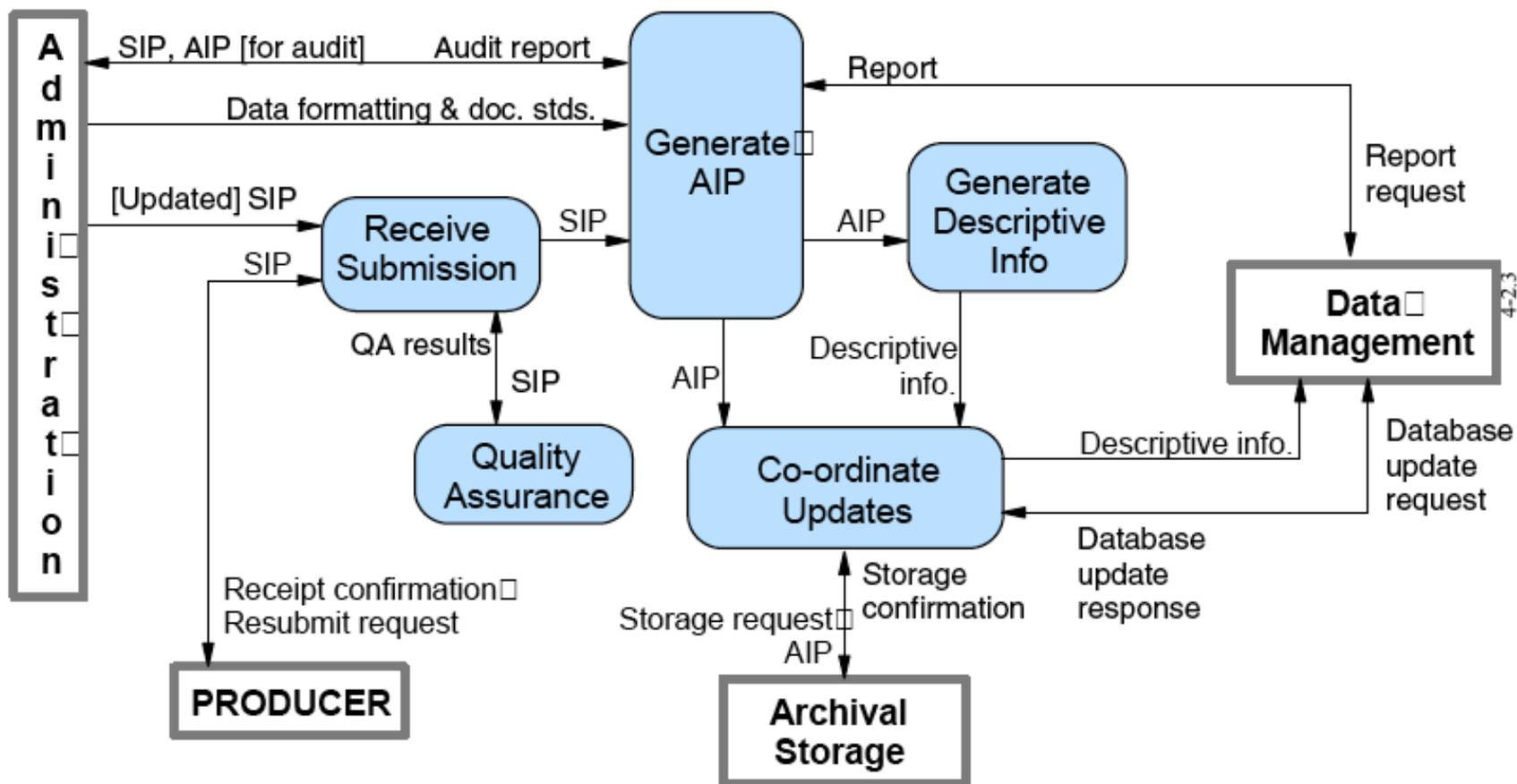
Realiza el aseguramiento de calidad/validación de los SIPs.

Genera el AIP que cumple con los estándares de formato de datos y documentos.

Extrae la información descriptiva y la envía al data management.

Coordina las actualizaciones en el archival storage y en el data management de la base de datos.

Functions of *Ingest*



Archival storage

Descripción: Provee los servicios y funciones para el almacenamiento, mantenimiento y recuperación de los AIPs.

Recibe el AIP de la entidad ingest y lo almacena. Gestiona las jerarquías de almacenamiento. Configura niveles especiales de servicio, seguridad y protección (por ejemplo backups). Provee estadísticas de inventario, capacidad disponible, etc.

Transforma los datos que constituyen la información de empaquetado para reproducir el AIP en el tiempo. Realiza una verificación de errores. Provee un mecanismo estándar para el seguimiento y verificación de la validez de los datos. Provee un mecanismo de duplicación de los contenidos en un lugar físico separado. Provee copia de los AIPs almacenados a la entidad *access*.

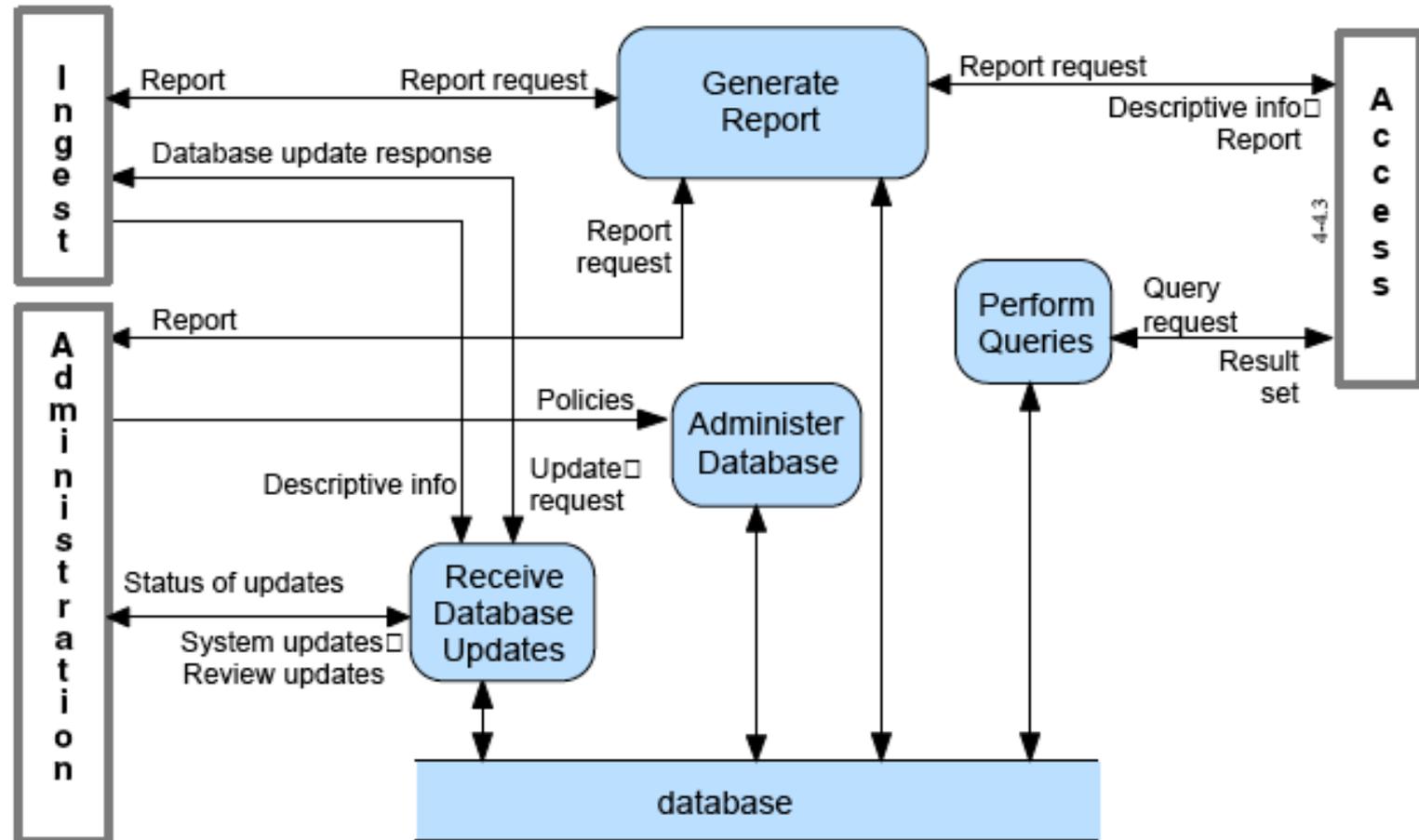
Data management

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

Recibe pedidos de las entidades *ingest*, *access* y *administration* y genera reportes.

También recibe actualizaciones de *ingest* y *administration*.

Functions of *Data Management*



Administration

Descripción: Provee los servicios y funciones para la operación global del sistema de archivos. Solicita la información necesaria sobre los archivos y negocia los acuerdos con los Productores. Monitorea la funcionalidad del sistema de archivos, controla los cambios de la configuración y mantiene su integridad y trazabilidad. Audita las operaciones del sistema, performance y uso. Envía reportes al *data management* y recibe reportes de esa entidad. Sumariza todos los reportes y provee información sobre performance del OAIIS e inventario y envía esta info a *preservation planning* para establecer políticas y estándares. Recibe los paquetes de migración para *preservation planning*.

Administration

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

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

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

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

Administration

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.

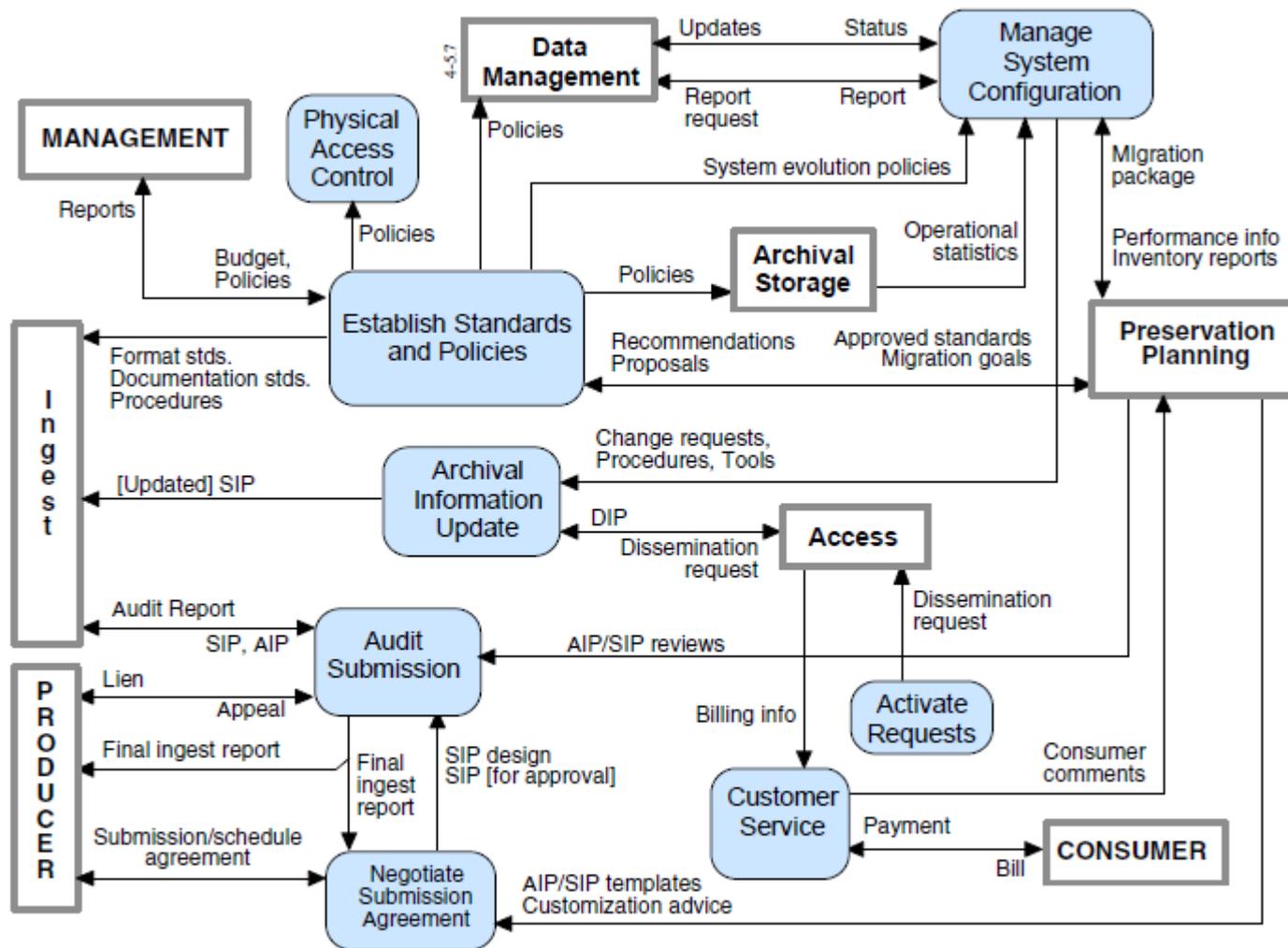


Figure 4-5: Functions of Administration

Preservation planning

Descripción: Interactúa con los consumidores y productores de archivos. Proporciona reportes, alertas de requisitos y estándares independientes. Identifica tecnologías que pueden causar obsolescencia. Desarrolla y recomienda estrategias y estándares, que envía a *administration*.

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

Preservation planning

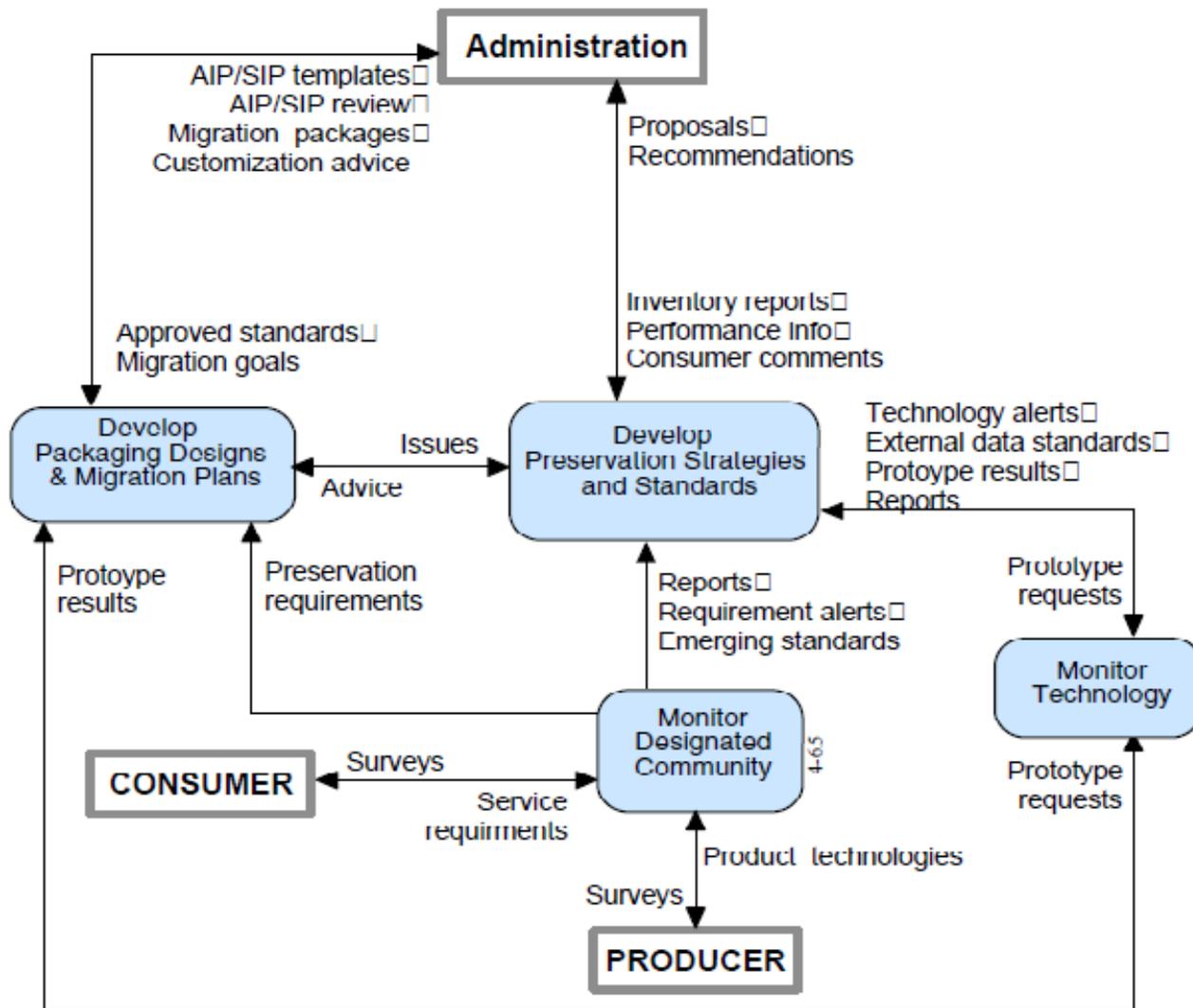


Figure 4-6: Functions of Preservation Planning



Access

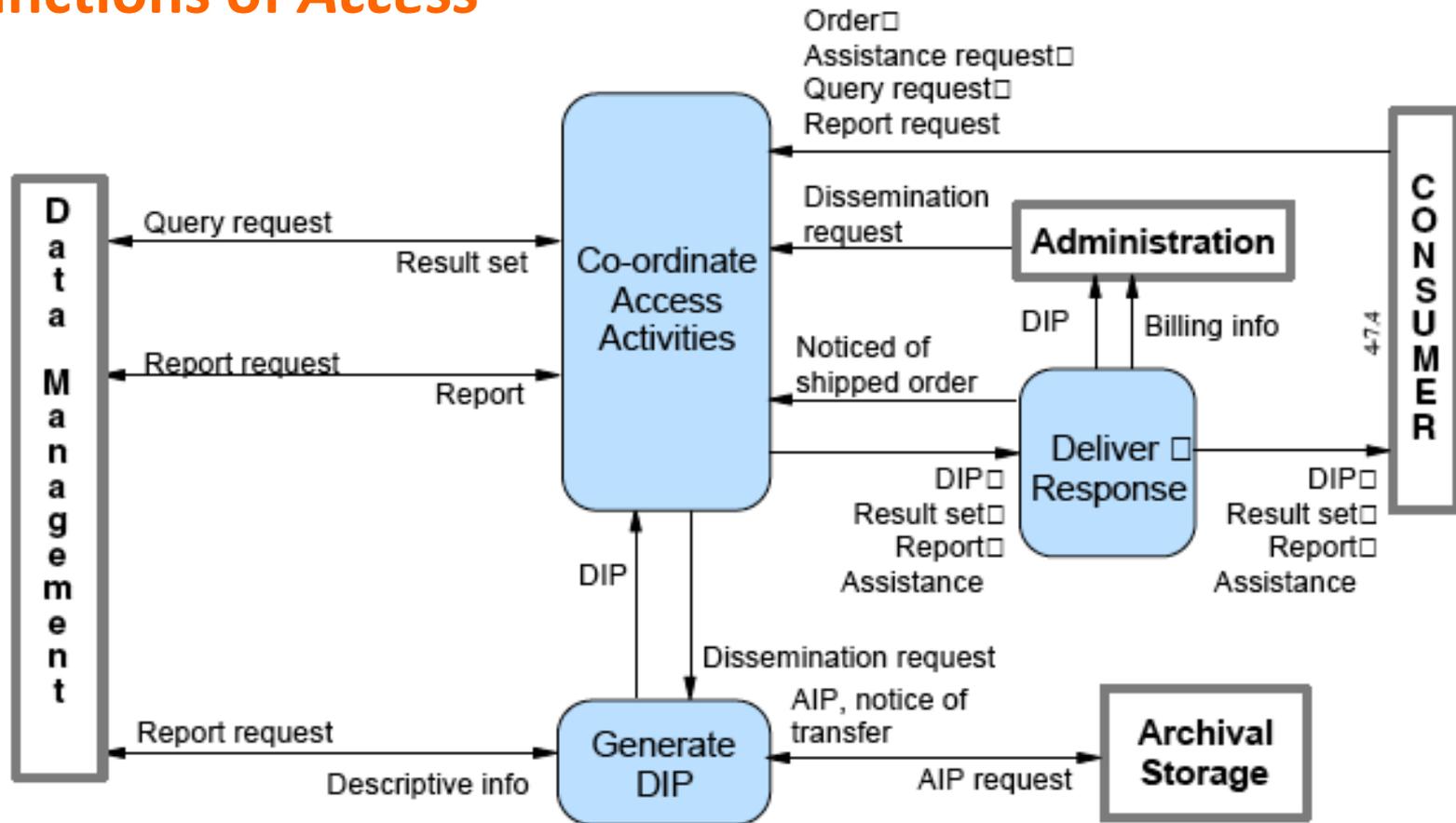
Descripción: Proporciona una interfaz única de usuario para el acceso a la información de los archivos.

Tiene 3 categorías, los *query requests*, los *result sets* y los *report requests*.

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

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

Functions of Access



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

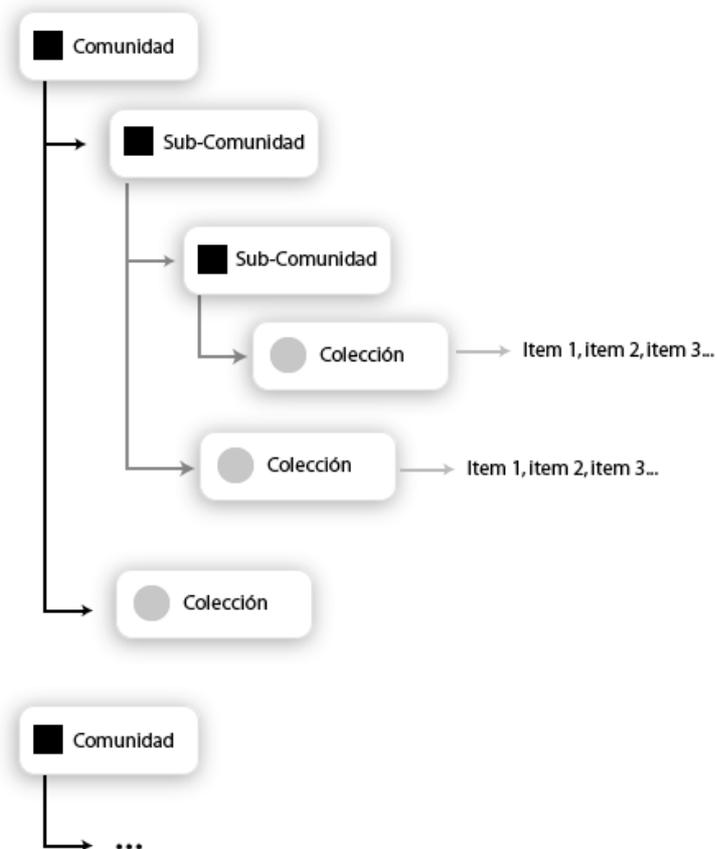
Parte 4- OAIS

Parte 5- DSPACE

Modelo de datos, OAIS en Dspace

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.



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 y 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 ítem, colección y comunidad. Se registran campos adicionales sobre varios elementos: tamaño, suma de comprobación y tipo Mime de cada archivo.



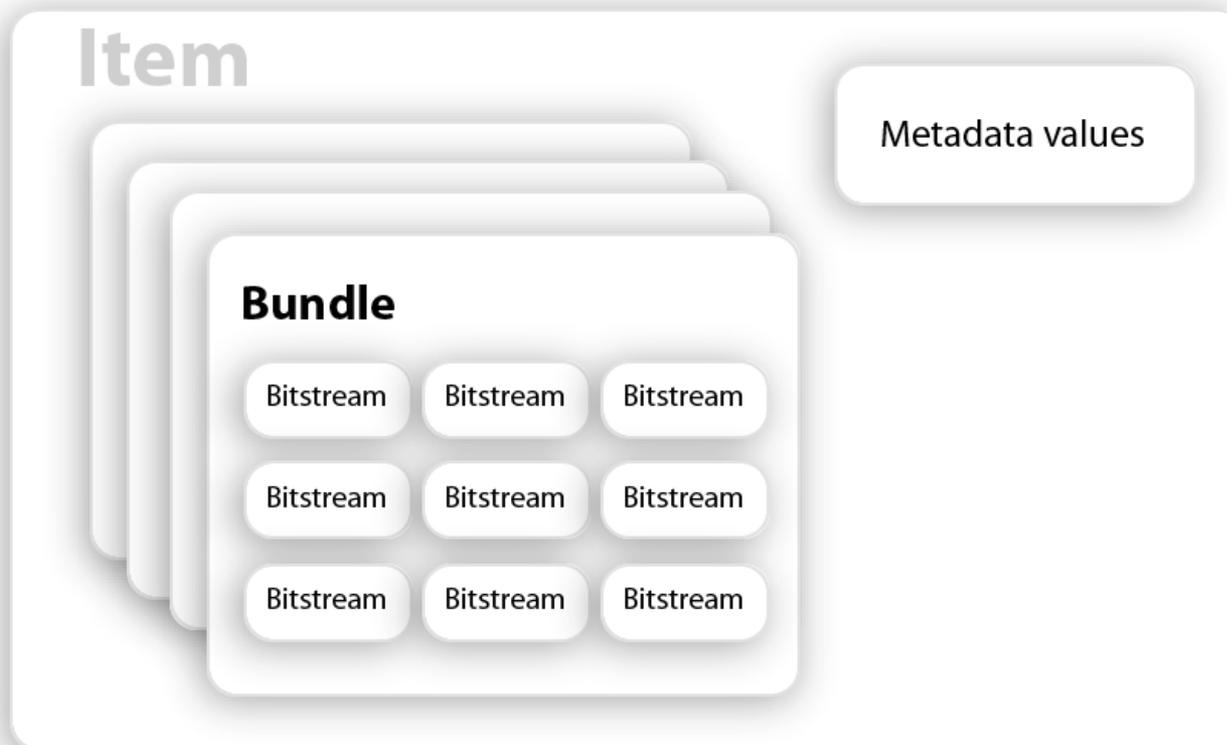
DSpace

3 grupos lógicos de metadatos para el contenido archivado

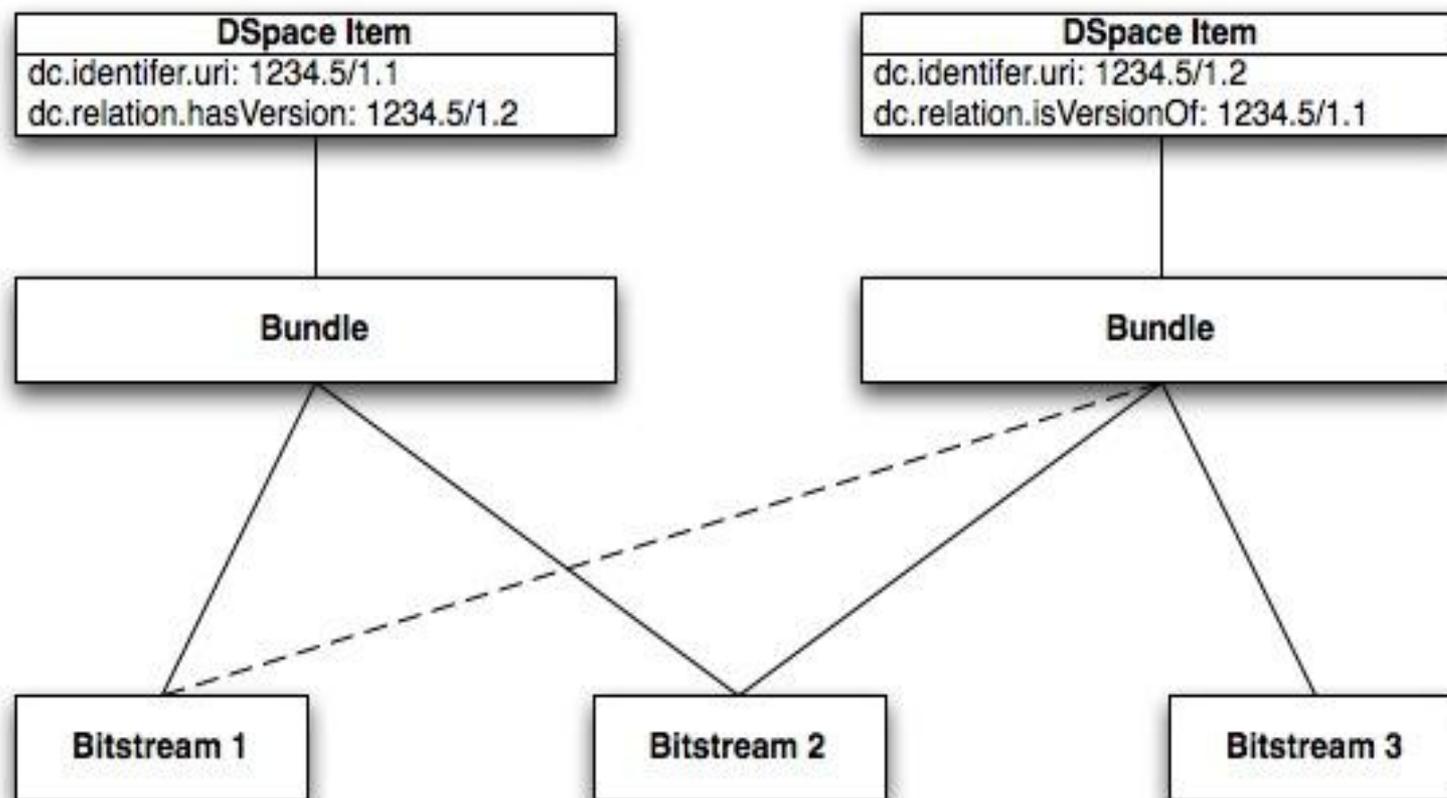
3) Metadatos estructurales: mantiene sólo unos pocos metadatos estructurales sobre los archivos de cada ítem p.e. la relación de pertenencia entre paquetes de archivos (bundles) y orden. Dependiendo cómo se armen las comunidades y colecciones, pueden considerarse las relaciones de pertenencia a colección 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



DSPACE: Modelo de datos



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).

OAIS en DSPACE

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

OAIS en DSPACE

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

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

OAIS en DSPACE

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

Ingest

En DSpace: Se tiene un proceso claramente definido como *Ingest*, que esta compuesto por varios subprocessos. 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 *InProgressSubmission* que, dependiendo de la configuración, pasa por varios estadios durante los cuales se le agregan metadatos descriptivos y administrativos, y se realizan *checksums* a los contenidos. El resultado de esta secuencia es pasado al *item installer*, el cual debe convertir ese objeto en un *Archived Item* de DSpace. El *item installer* 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.

Archival storage

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. Dspace ofrece la posibilidad de definir y ejecutar tareas de *curation* 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.

Archival storage

En DSpace: 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. 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.

Data management

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.

Asimismo, cuando se utiliza el módulo *Discovery* 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.

Administration

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, item, colección y comunidad.

Administration

En DSpace: Por ello, los usuarios encargados de comunidades pueden mantener de control sobre su contenido en el OAI, 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 items 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.

Administration

En DSpace: El módulo de estadísticas provee información de uso y acceso sobre las comunidades, colecciones, items y bitstreams por parte de los visitantes, junto a algunos eventos notables como ser el acceso de usuarios registrados.

Preservation planning

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.

Access

En DSpace: Se proveen dos 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. 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.



Access

En DSpace: 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.

Access

En DSpace: Adicionalmente, es posible extender los mecanismos de exportación e incorporar nuevos ExporterPlugins para generar DIP en otros formatos.

Visualización de información estadística generada por el módulo correspondiente.

¡Muchas gracias!

Marisa R. De Giusti

marisa.degiusti@sedici.unlp.edu.ar

Gonzalo L. Villarreal

gonzalo@prebi.unlp.edu.ar

