



UNLP



## **Bibliotecas y Repositorios Digitales**

Tecnología y Aplicaciones

[sedici.unlp.edu.ar](http://sedici.unlp.edu.ar)



# Bibliotecas y repositorios digitales

## Capítulo 8

La preservación en el repositorio institucional: PREMIS, el Modelo OAIS ISO 14721. Comparación con las facilidades que ofrece DSPACE. Visibilidad institucional y del repositorio (especialmente del repositorio).





# Introducción

En la actualidad, los recursos que se generan como resultado de los conocimientos de las personas y de sus expresiones “nacen”, cada vez más, en formas digitales, sean de carácter cultural, educativo, o engloben información de diferentes áreas del saber, ya sean de naturaleza técnica, artística o administrativa. Los productos de origen digital pueden no contar con un respaldo físico, por ejemplo en papel.

Muchos de estos recursos son valiosos y constituyen un verdadero patrimonio a conservar a futuro para la sociedad.





# 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.



# Metadatos y metadatos de preservación



Los objetos digitales cambian, y dichos cambios deben registrarse y validarse para asegurar la autenticidad del objeto, por lo que también es preciso incorporar metadatos de procedencia y autenticidad. Dado que cualquier actividad de preservación está limitada por los derechos de propiedad intelectual, se hace necesario incluir metadatos para la gestión de los mismos.





# Preservación y estándares

- 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 estándar 14721 (OAIS), los metadatos PREMIS y las directrices para la preservación, en conjunto con el esquema METS, constituyen el marco ideal para la gestión de un repositorio, para asegurar su interoperabilidad y dar preservación a sus contenidos.



# Problemas en la preservación: software



- Muchos problemas en lo relativo a la preservación derivan de una configuración deficiente del software que soporta el repositorio. Es necesario revisar las facilidades del software que soporta el repositorio en comparación con el modelo de preservación OAIS y realizar las personalizaciones necesarias para cumplir con algunos requerimientos del plan de preservación no brindados de forma nativa. Lo mismo con PREMIS.



# El Modelo OAIS



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

ISO 14721: 2003

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







# El Modelo OAIS

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





# El modelo de Referencia OAIS

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





# Sección 1





# Justificación del Modelo de referencia

- Ninguna discusión sobre la conservación de repositorios y flujos de trabajo estaría completa sin al menos una breve introducción al modelo de referencia OAIS.
- Una introducción a este modelo sirve para mostrar cómo implementa muchos de los procesos de flujos de trabajo y cómo se relaciona con la conservación digital.
- Se recomienda como la mejor práctica actual, si bien no se ha establecido aún un consenso firme.





# Antecedentes

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





# Funciones del Modelo de referencia

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





# Propósito y campo de Aplicación

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





# Propósito y campo de Aplicación

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







# Sección 2





# Conceptos en OAIS

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



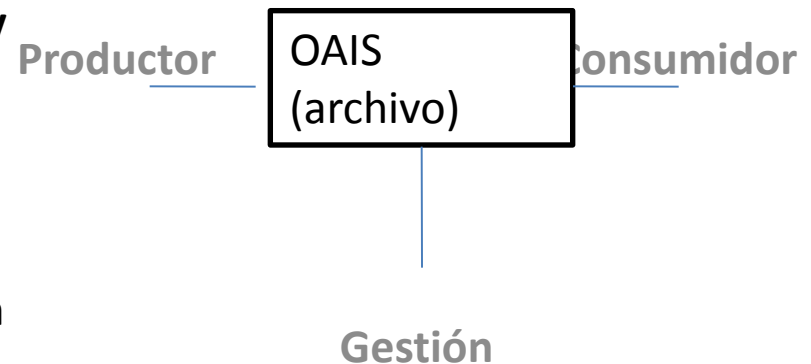


## Medioambiente OAIS

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

## Actores en el modelo

- **Producer-Consumer-Management**



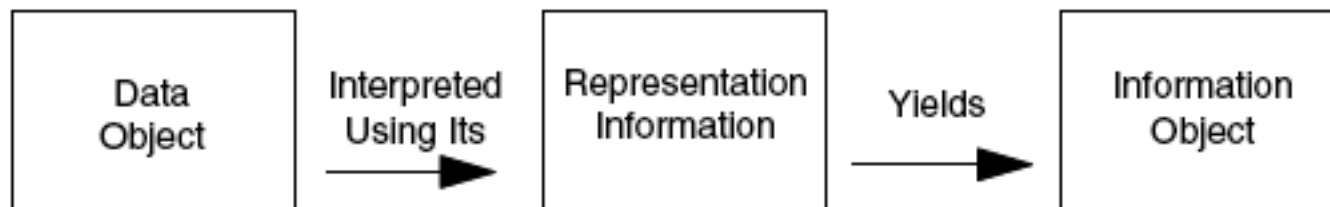


# Conceptos en OAIS

Una definición clara de información es central para la capacidad del OAIS para preservar esa información.

Una persona o un sistema, tienen una base común de conocimientos (KB) que le permite comprender la información.

Se considera información en este campo a cualquier tipo de conocimiento que puede intercambiarse y que se expresa a través de algún tipo de datos: la información en un artículo periodístico, se expresa por caracteres (datos), los cuales bajo el paraguas de un lenguaje (KB), se convierten en información relevante. Si el receptor desconoce la lengua, entonces el artículo tendrá que ser acompañado por información extra, por ejemplo, un diccionario o una gramática.





# Conceptos en OAIS

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





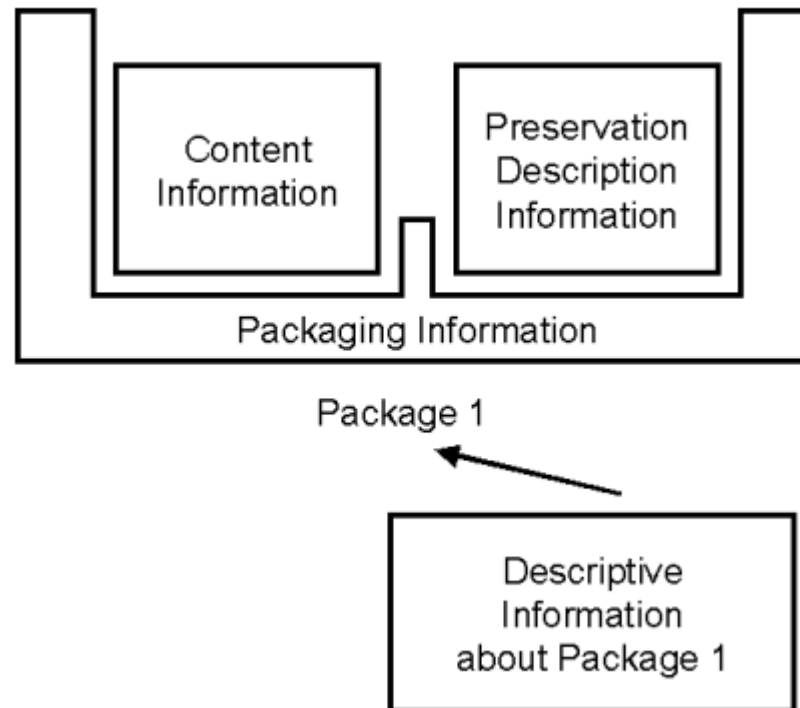
# Conceptos en OAIS

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





# Conceptos en OAIS



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





# Conceptos en OAIS

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





# Estructura del Paquete de Información



**Paquete 1**



Información  
descriptiva sobre  
**Paquete 1**





# El paquete de información -IP

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





# Conceptos en OAIS

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





# Clases de IPs según su función

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





# Clases de IPs según su función

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





# Clases de IPs según su función

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



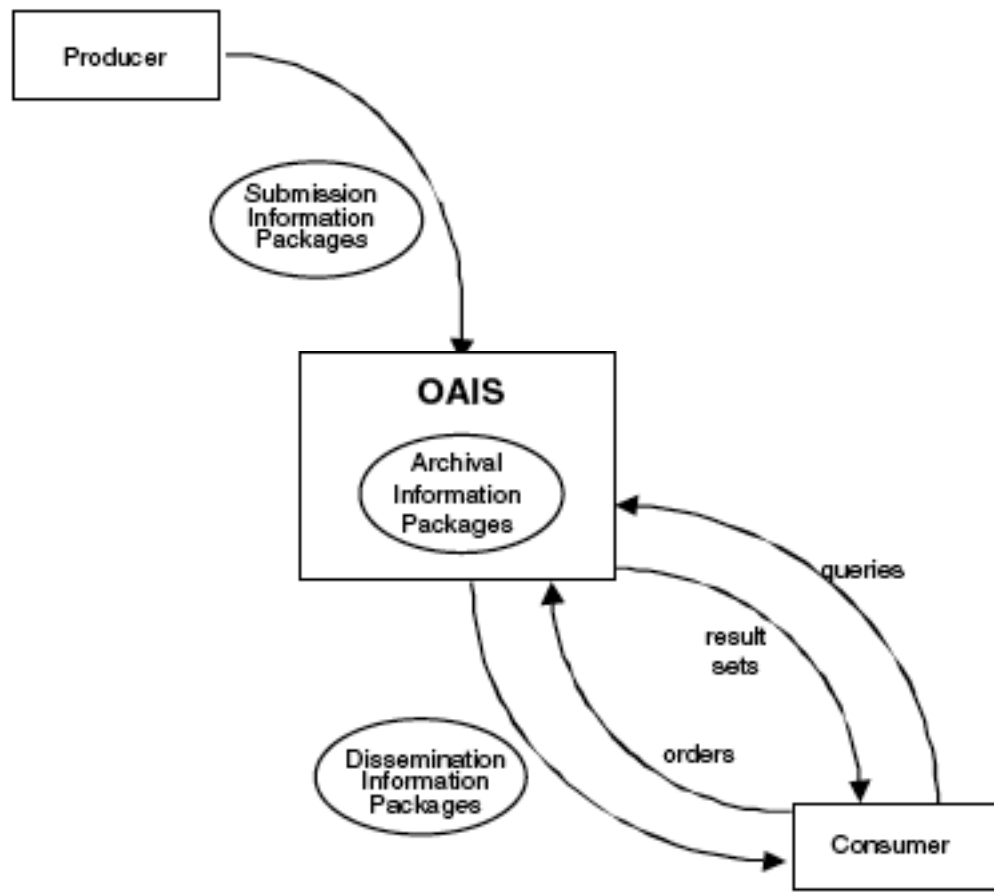
# OAIS interacciones externas de alto nivel



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



# OAIS interacciones externas





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



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





# Sección 3: Responsabilidades del OAIS



# OAIS Responsabilidades obligatorias



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



# Responsabilidades del OAI



- EJEMPLOS





# OAIS Responsabilidades obligatorias

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





# Mecanismos de deslinde de responsabilidades

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



# Sección 4





# OAIS

## Modelo Funcional

### Sección 4.1









# OAIS Modelo funcional

Seis entidades funcionales e interfaces relacionadas:

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





# Modelo OAIS

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





# Modelo OAIS

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





# Modelo OAIS

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





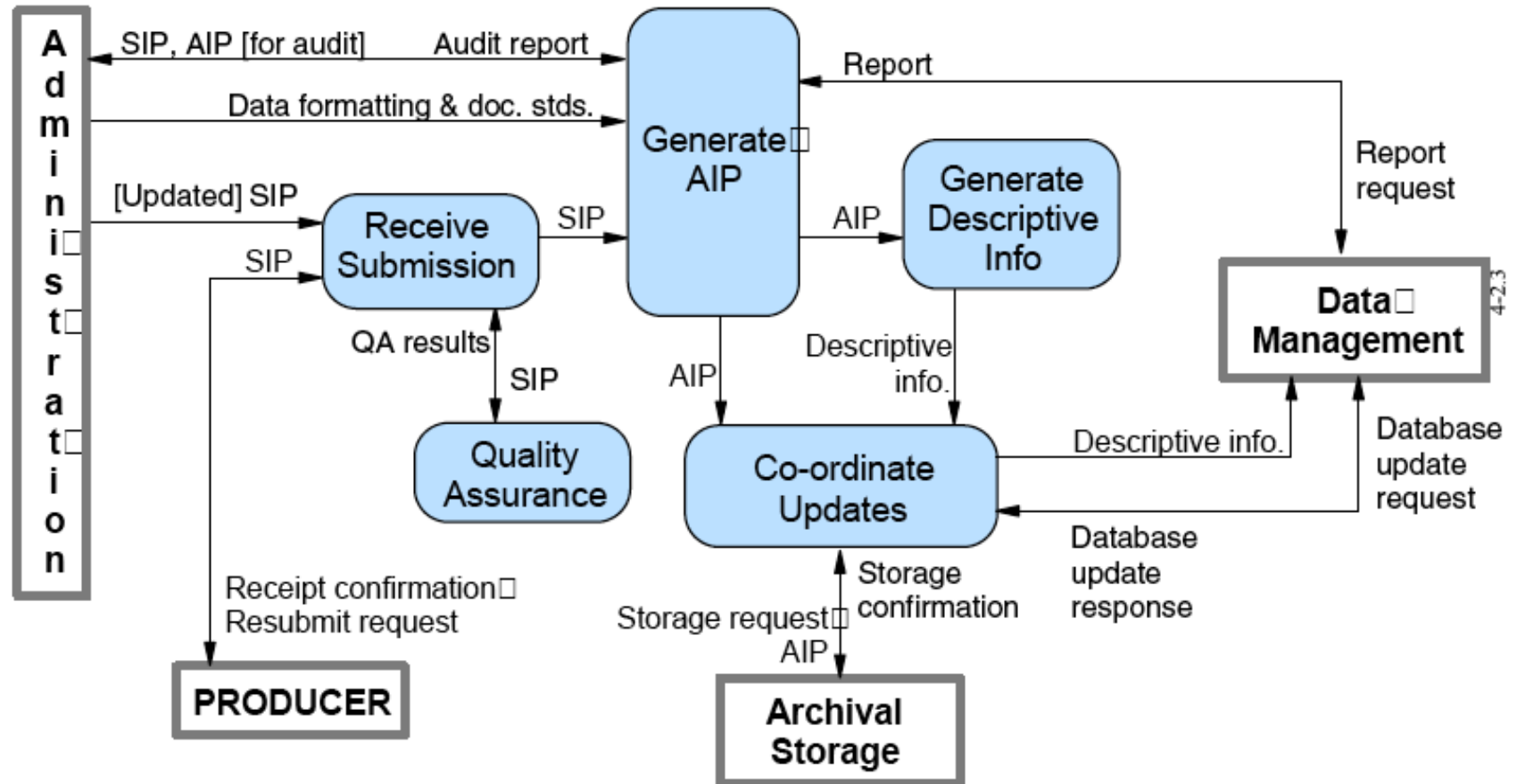
# Ingesta/Ingest/presentación

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





# Functions of Ingest



4-23



# Entidad OAIS Ingest

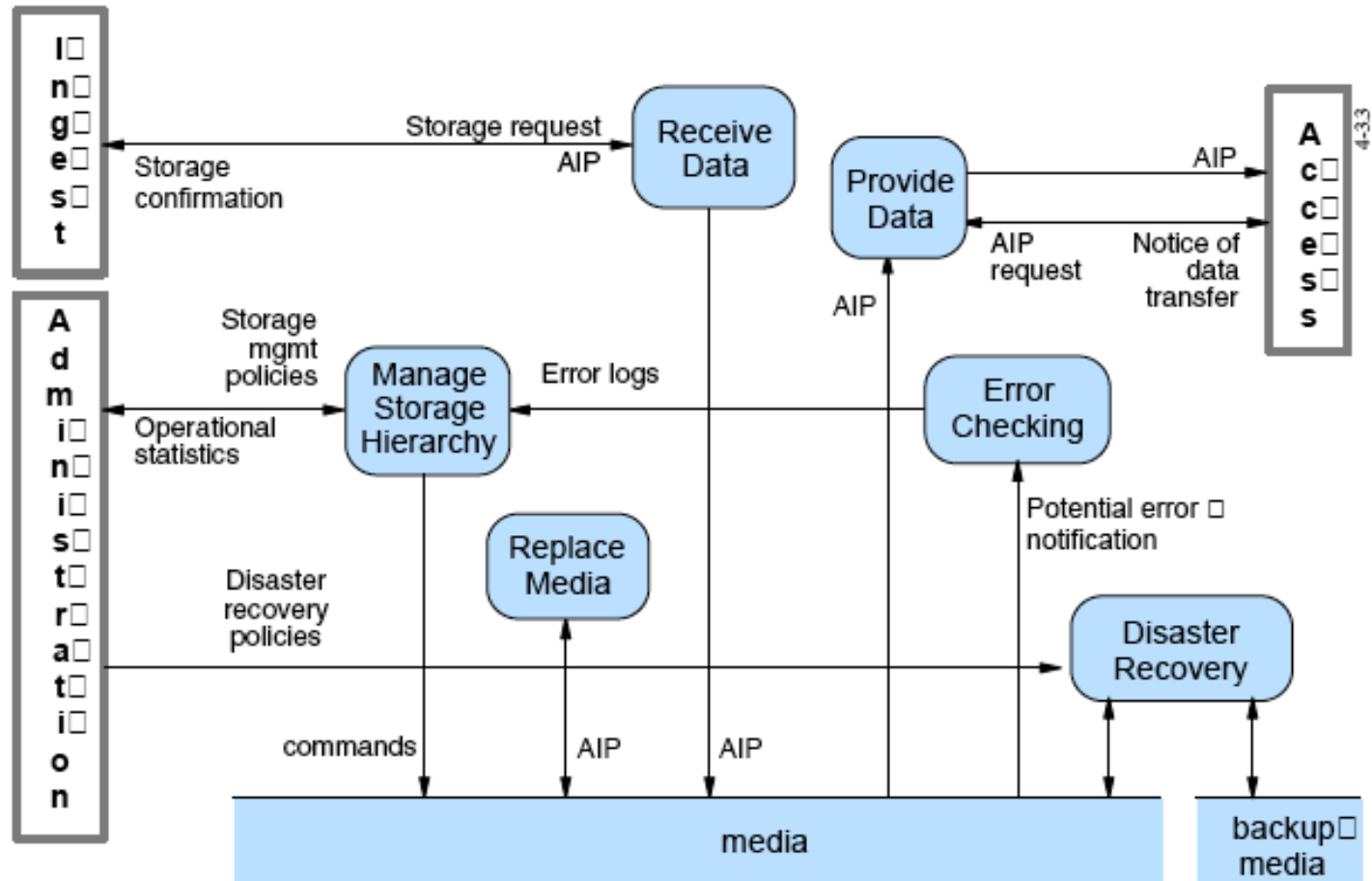
- **Descripción:** Provee los servicios y funciones para aceptar un SIP por parte de los Productores o bajo el control de la Administración.
- Prepara los contenidos para almacenamiento y gestión dentro del archivo.
- Realiza el aseguramiento de calidad/validación de los SIPs.
- Genera el AIP que cumple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







# Functions of Archival Storage





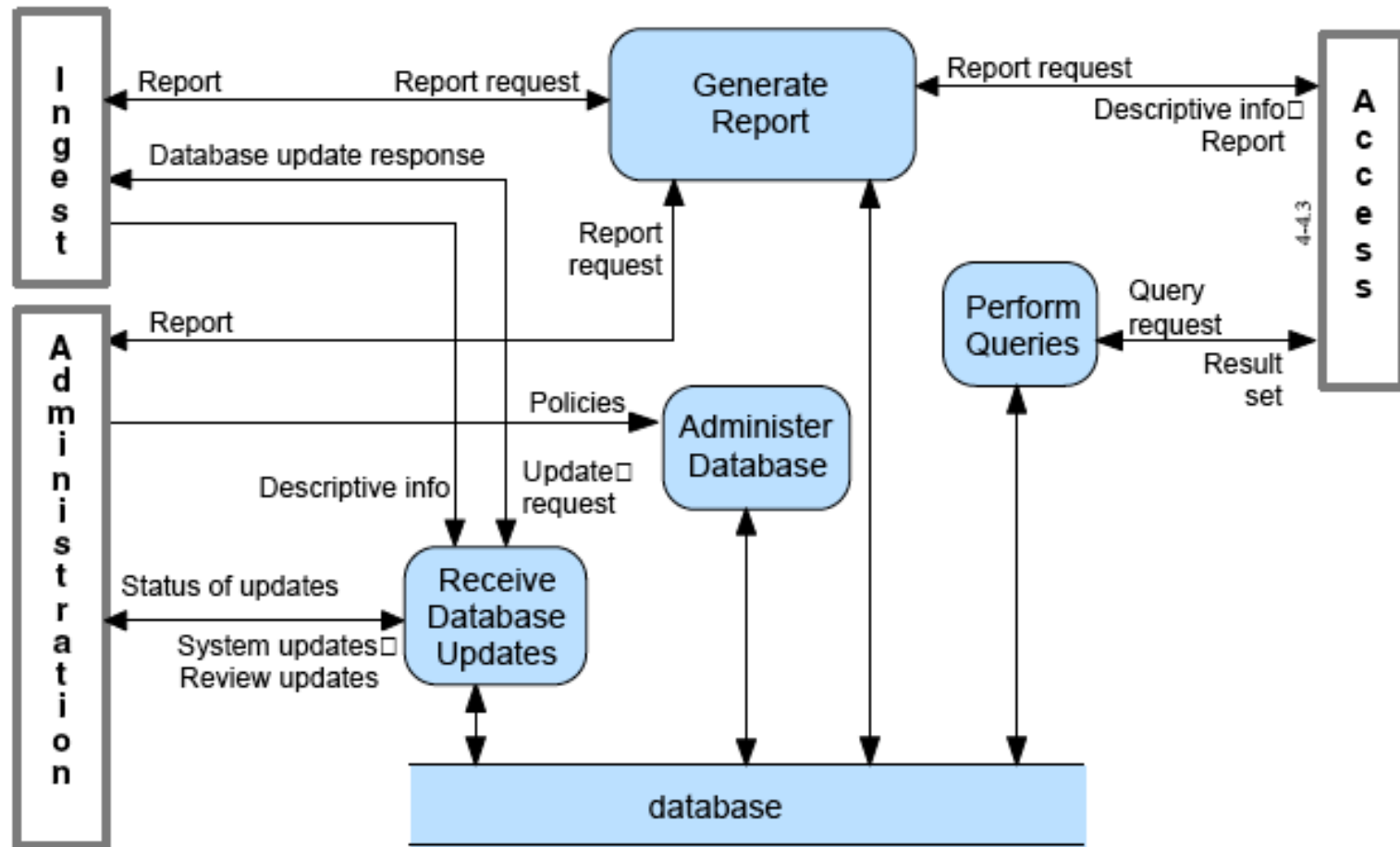
# Entidad OAIS Archival Storage

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





# Functions of Data Management





# Entidad OAIS Data Management

**Descripción:** Provee los servicios y funciones para poblar, mantener y acceder a la información descriptiva que identifica y documenta el contenido del Archivo, y a los datos administrativos usados para gestionarlo.

Es responsable de la administración de la base de datos.

Recibe solicitudes de la entidad *access* y genera un conjunto de resultados.

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

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





# Entidad OAIS Administration

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

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

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





# Entidad OAIS administration

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

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

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

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





# Entidad OAIS administration

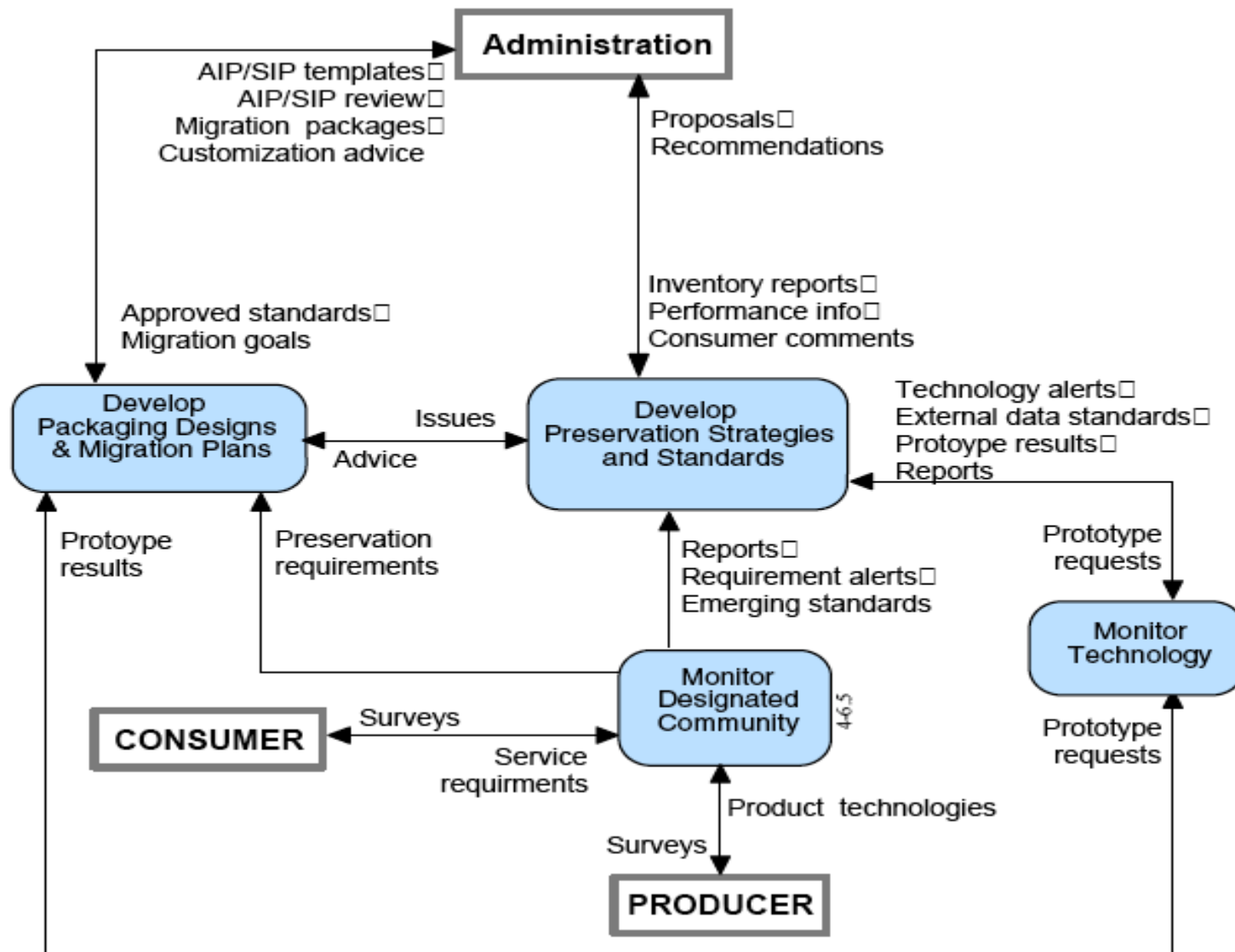
Verifica que los AIP y SIP suministrados sigan las especificaciones. Verifica el PDI según los usos de la comunidad designada.

Revisar periódicamente los contenidos del archivo para determinar si los datos están disponibles.

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



# Functions of Administration







# Entidad OAIS preservation planning

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

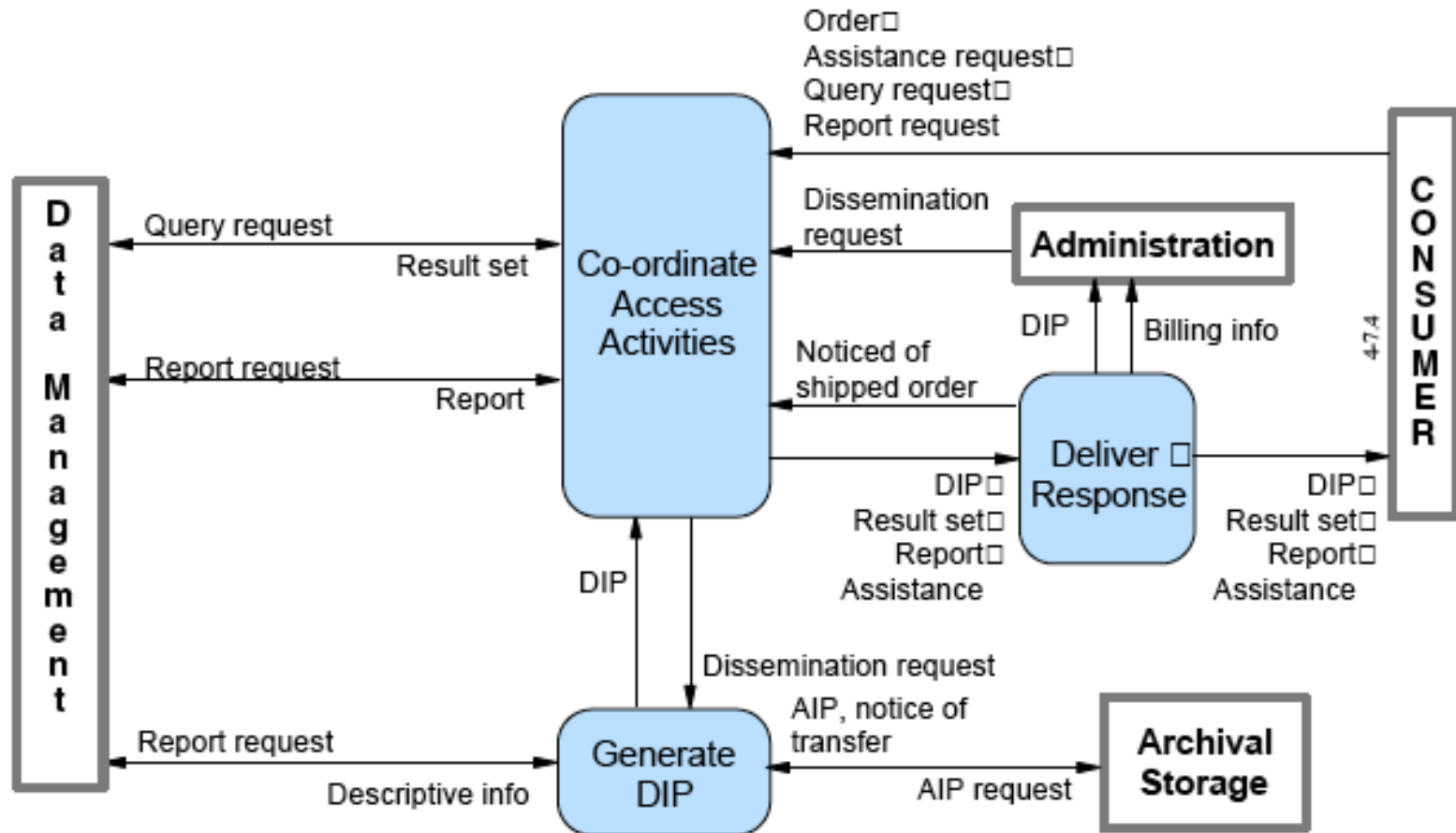
Desarrolla y recomienda estrategias y estándares, que envía a *administration*.

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





# Functions of Access





# Entidad OAIS Access

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

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

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





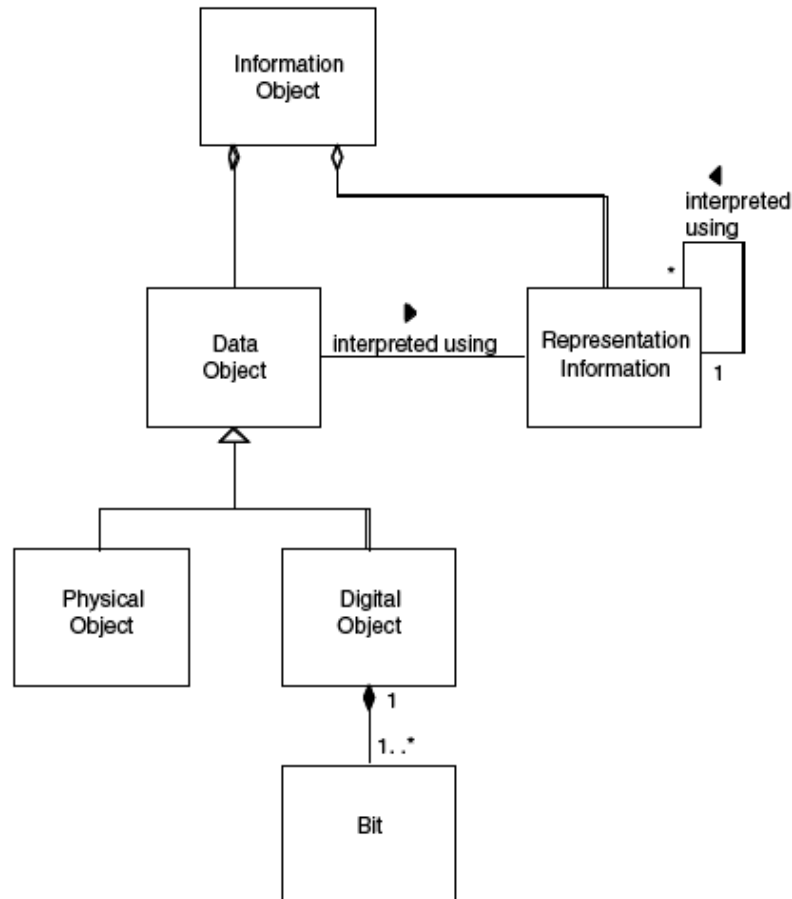
# OAIS

## Modelo de Información Sección 4.2





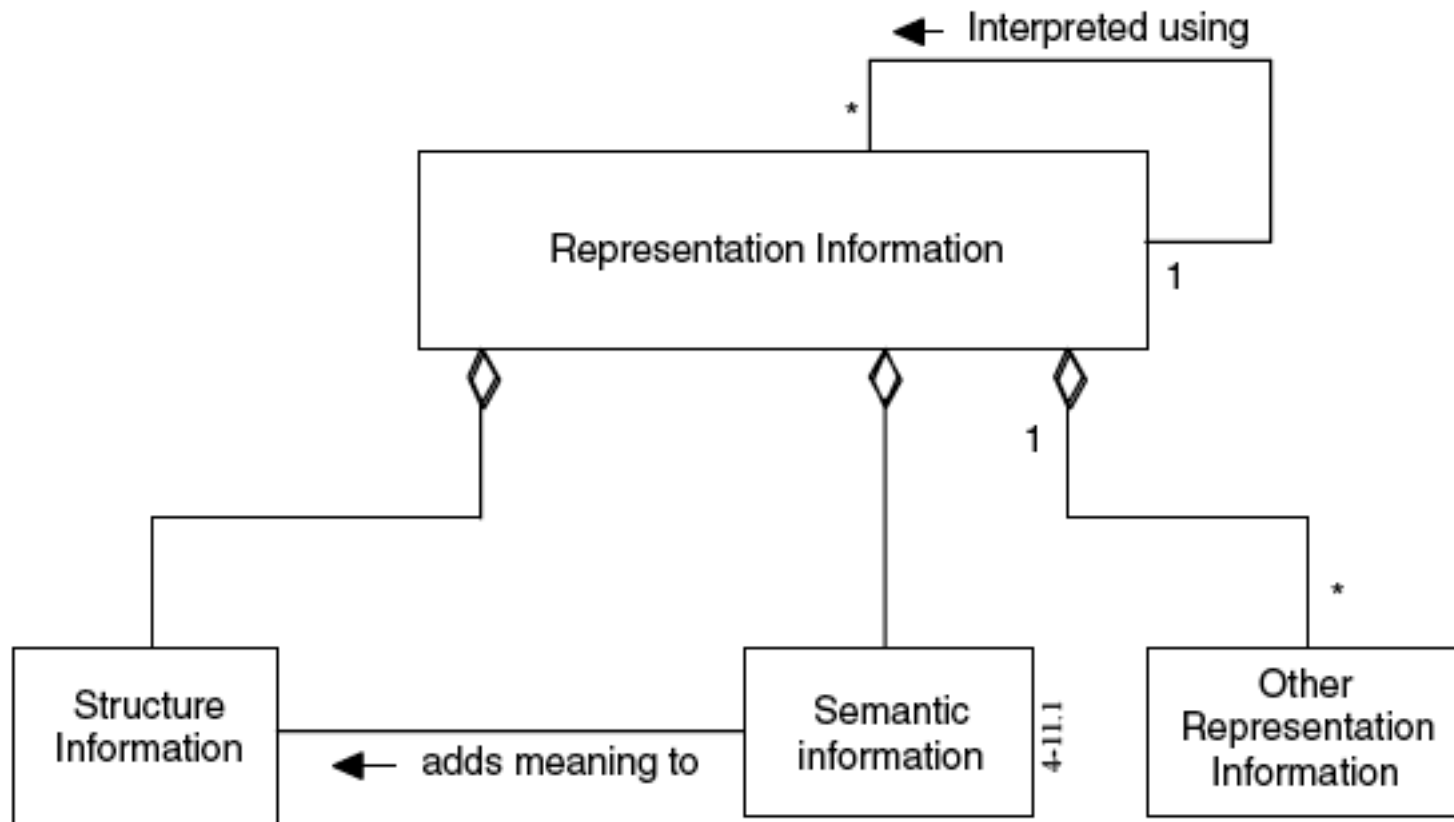
# OAIS Objeto de información



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

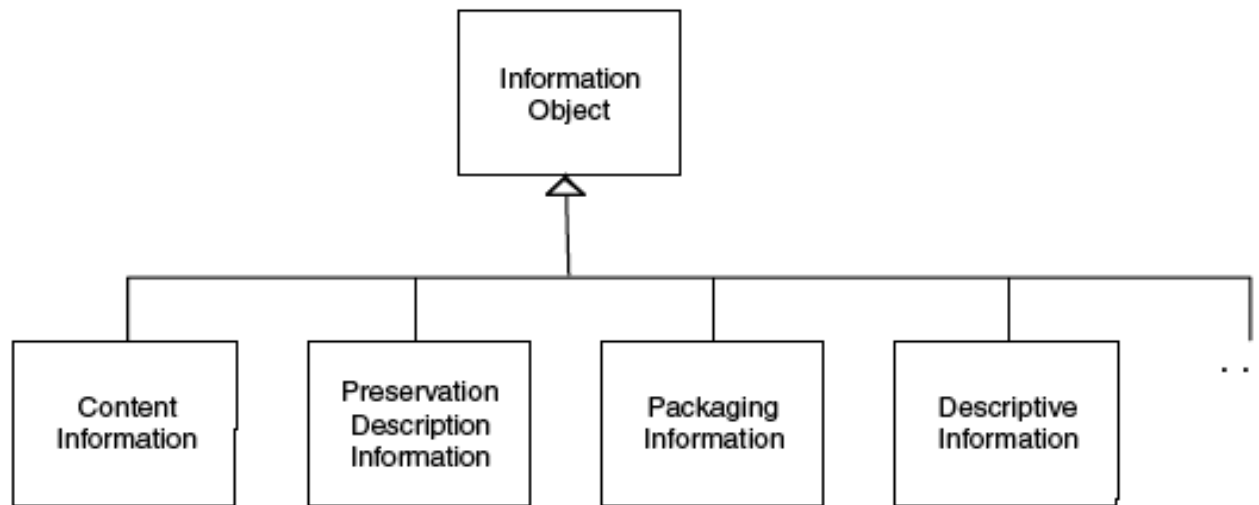


# Representation Information Object





# Tipos de objetos de información



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





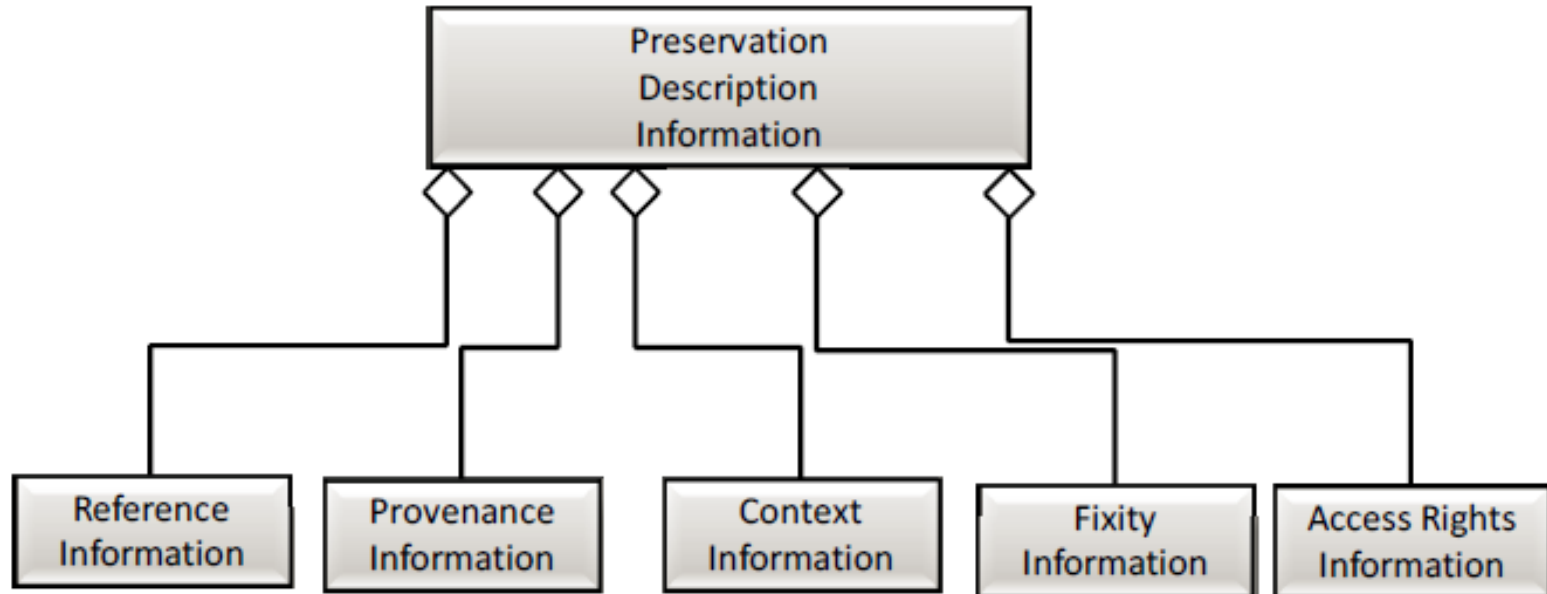
# Información de contenido

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





# Información descriptiva de preservación (PDI)



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

# Información descriptiva de preservación



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

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



# Información descriptiva de preservación



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

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

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





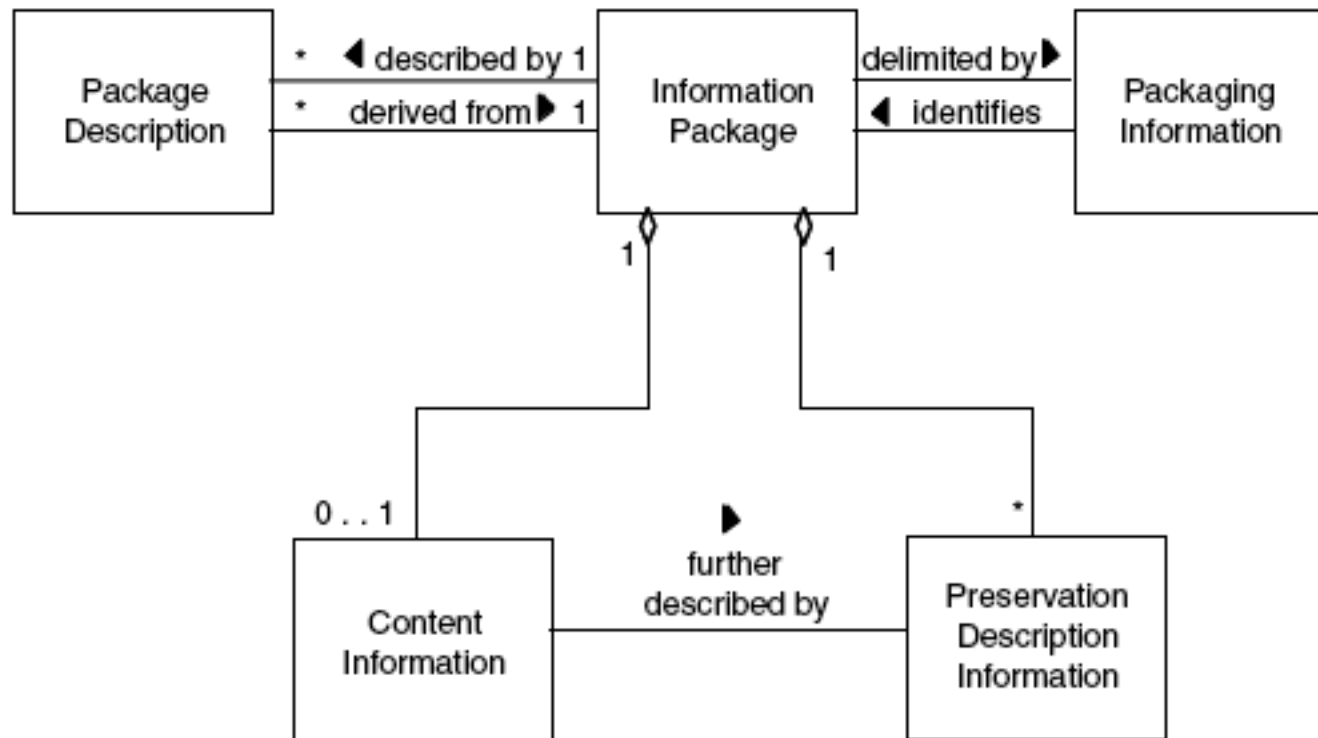
# Paquetes de información en OAIS

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



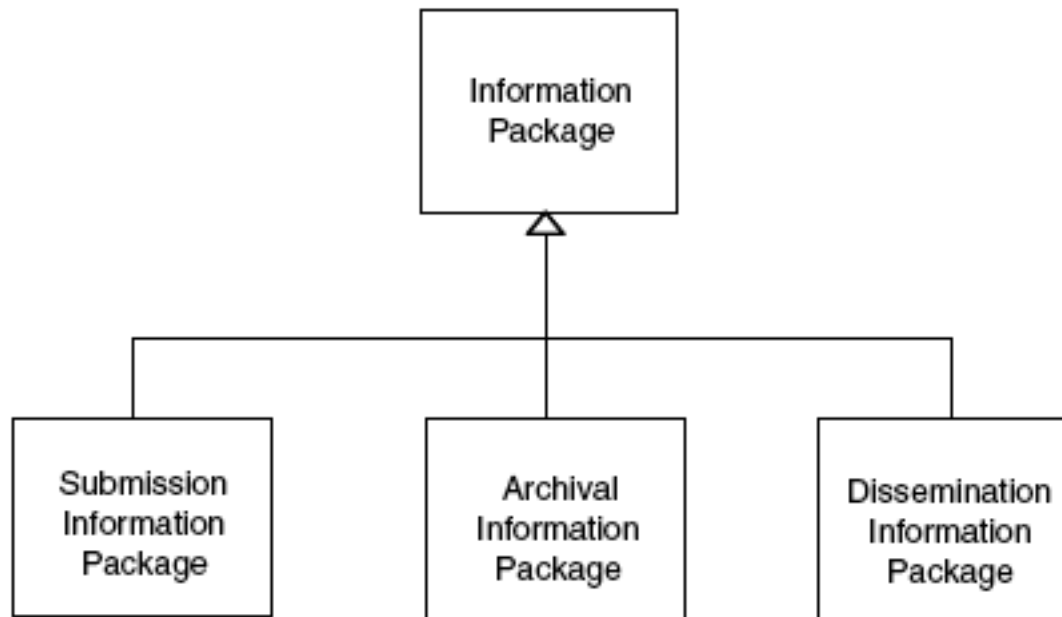


# Paquetes de información en OAIS





# Tipos de paquetes de información



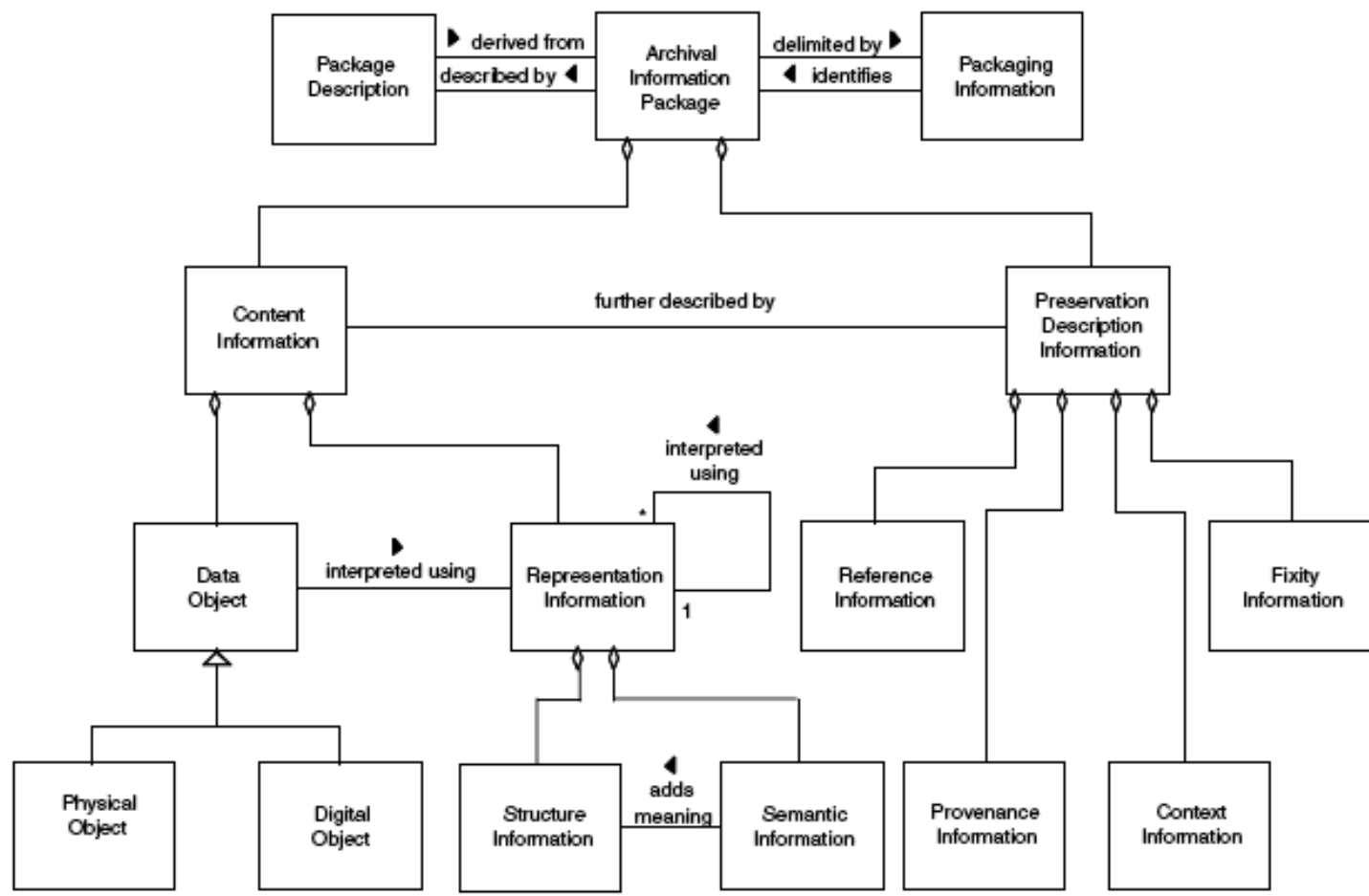


# SIP

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

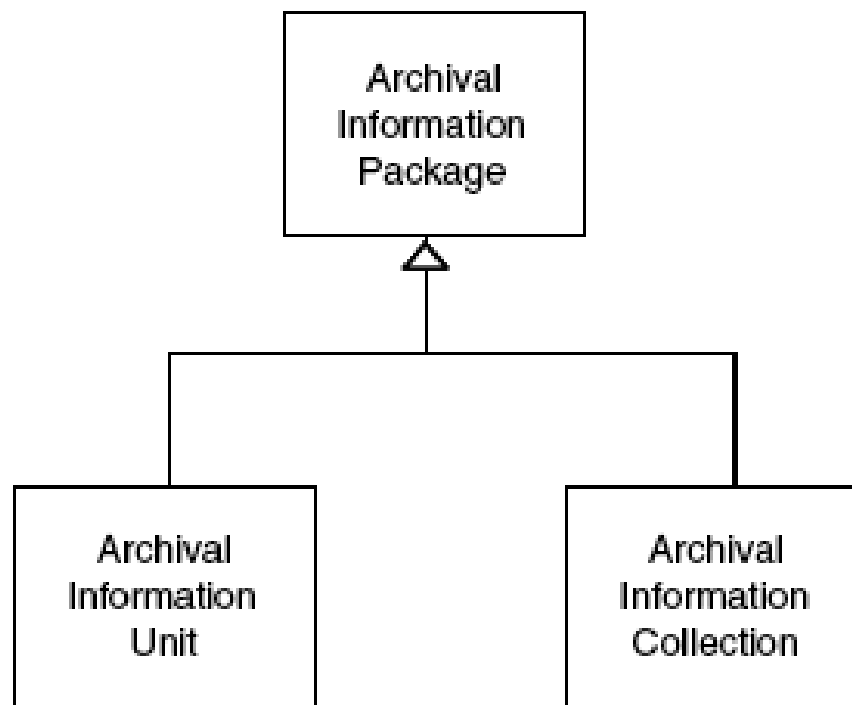


# AIP





# Tipos de AIPs





# DIP

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





# Participantes

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





# Participantes

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



# Participantes



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





# Participantes

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





# Participantes

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





# Participantes

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







# Participantes

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





# Participantes

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





# Participantes

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



# Sección 5: Perspectivas sobre preservación



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



# Saliendo de la 14721





# Aproximaciones a la preservación

Existen numerosas estrategias para asegurar la preservación de la información:

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



# Repository preservation and interoperability(Preserv2)



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





# Metadatos

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







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

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





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





# Metadatos de preservación

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



# PREMIS



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

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





# Diccionario de datos PREMIS

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

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





# Modelo de Datos PREMIS

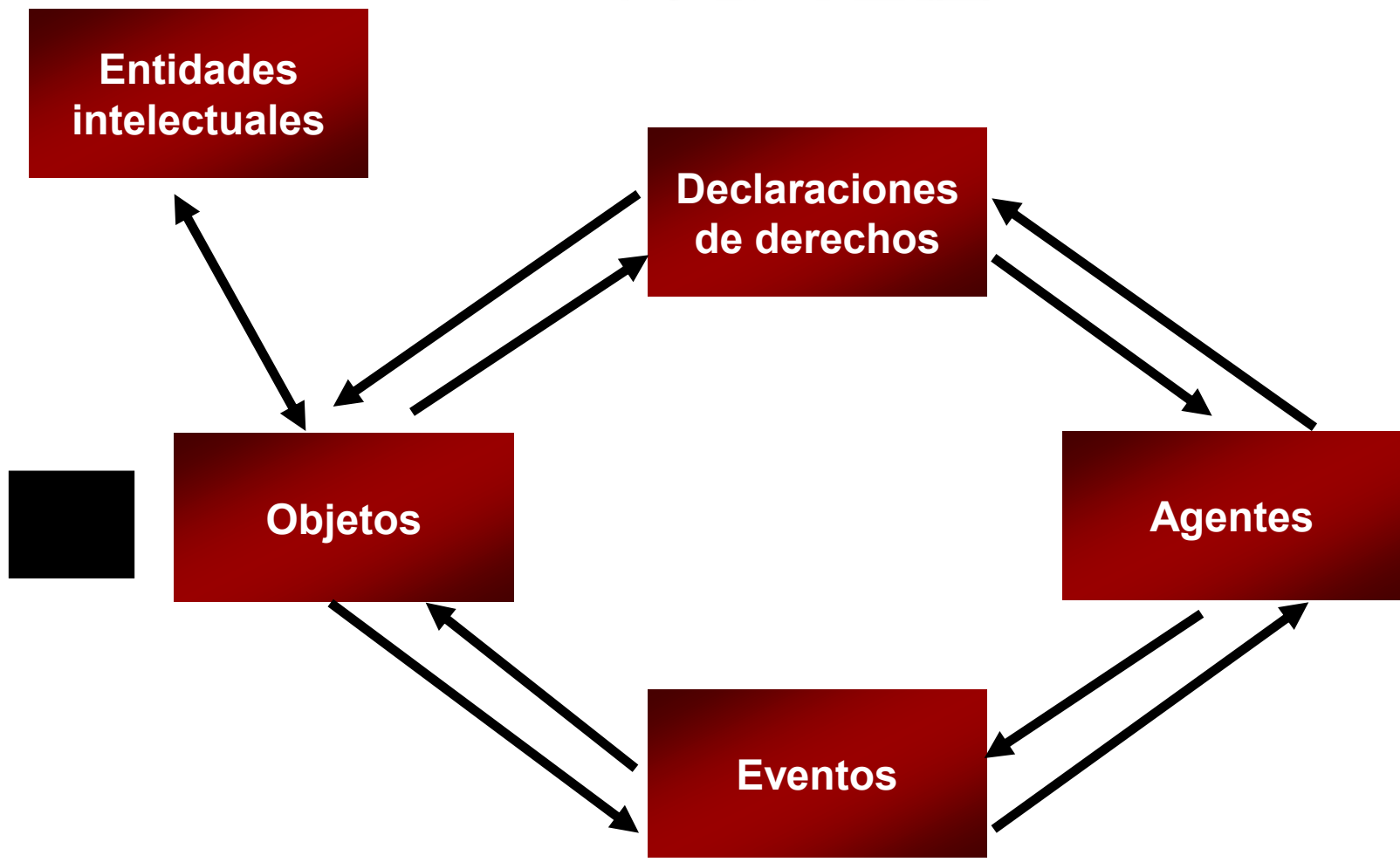
Las entidades que este modelo define se denominan:

- Entidades intelectuales
- Objetos
- Derechos
- Agentes
- Eventos





# Modelo de datos PREMIS





# Entidad intelectual

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





# Objetos



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





# Eventos

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



# Agentes



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





# Derechos

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





VIII Workshop 

*Proyectos digitales*

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

# Metadatos para la preservación digital: PREMIS

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

emendez@bib.uc3m.es

Dpto. Biblioteconomía y Documentación

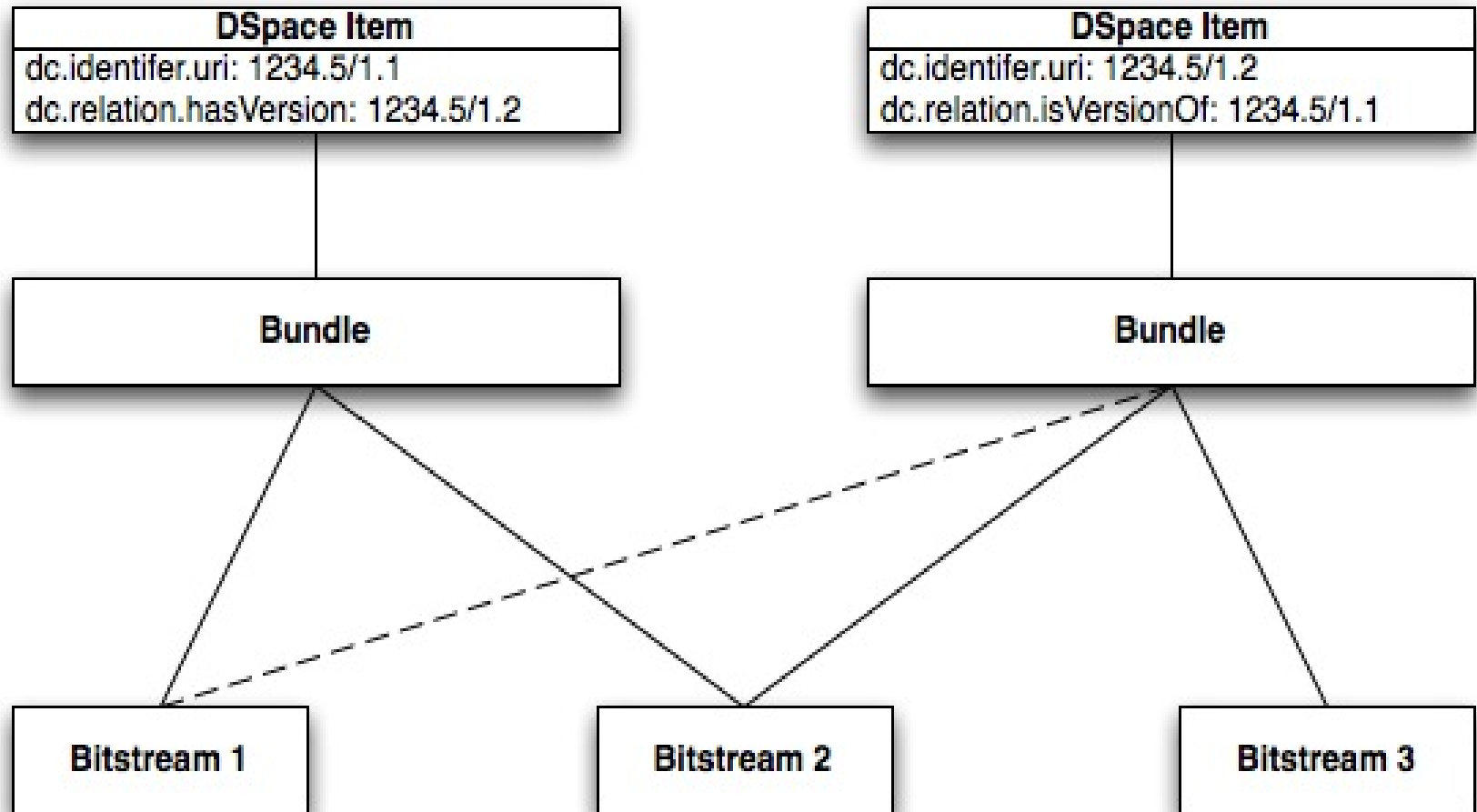


Universidad Carlos III de Madrid





# DSPACE: Modelo de datos



# OAIS y DSPACE



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

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



# OAIS y DSPACE



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







# OAIS y DSPACE

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

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





# OAIS en DSPACE

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





# OAIS en DSPACE

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

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





# OAIS en DSPACE

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





# OAIS en DSPACE

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

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





# OAIS en DSPACE

## Tareas de Curación en Dspace

### Definición

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

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

### Tareas de curación en Dspace

Dspace proporciona herramientas para implementar distintas tareas de curación automáticas.





# OAIS en DSPACE

## Tareas de Curación en Dspace

### ¿Qué se puede realizar?

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





# OAIS en DSPACE

## Tareas de Curación en Dspace

¿Cómo se pueden realizar?

**Dspace provee:**

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







# OAIS en DSPACE

## Tareas de Curación en Dspace

¿Cómo se pueden realizar?

Se debe implementar:

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





# OAIS en DSPACE

## Tareas de Curación en Dspace

### Ejemplos:

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





# OAIS en DSPACE

## Tareas de Curación en Dspace

### Ejemplos de utilización (consola):

• ./dspace curate **-r** **-t** nombreCuration **-i** <handleItem | handleComunidad  
| handleColección | all> **-v**

Donde:

- **-r** define el tipo de reporte para gestionar la salida
- **-l** le indica que la salida será la estándar
- **-t** permite definir el tipo de tarea que se va a ejecutar
- **-i** permite definir los tipos de ítems en donde se aplicará la tarea
- **-v** hace más comunicativa la salida de la tarea





```
/home/nestor/workspace_indigo_2/dspace-sedici/install/bin : sudo
Archivo Editar Ver Marcadores Preferencias Ayuda
terru@pc-nestor ~/Java/workspace/dspace-sedici/install/bin $ sudo ./dspace curate -r - -t evaluatePreservation.HandleValidatio
Adding task: evaluatePreservation.HandleValidation
Starting curation
Curating id: 10915/10672

Procesando Item con id: 9620
El Handle del item: 9620 es válido, el item fue evaluado con 1
La puntuación total del item procesado es: 1.0
El item: 9620 se finalizó de procesar con éxito
Ending curation. Elapsed time: 107
terru@pc-nestor ~/Java/workspace/dspace-sedici/install/bin $
```



# Visibilidad Web

Un repositorio institucional preserva la producción de una institución

El objetivo buscado es que esta producción llegue al mayor público posible

Es necesario trabajar sobre la visibilidad web de la institución y del repositorio

*Conjunto de herramientas, técnicas, servicios, productos y procesos relacionados con incrementar la difusión de la institución en la web*

Definición extraída de <http://www.biblioseo.com/2010/04/que-es-la-visibilidad-web-y-por-que-es.html>





# Visibilidad Web

Abarca gran variedad de contenidos:

- producción científica y académica: artículos en revistas y congresos, tesis y tesinas, libros y ebooks
- revistas y congresos propios
- documentos multimediales: audio, imágenes, videos; objetos de aprendizaje, planes de estudio y materiales de cátedras
- proyectos de investigación, trabajos de extensión, actividades de transferencia
- normativas y resoluciones, planes estratégicos, convenios





# Visibilidad Web

Involucra múltiples actores:

- autoridades, docentes, investigadores, tesistas y becarios, personal de apoyo, personal administrativo
- dependencias, secretarías, proyectos, programas
- unidades de investigación: laboratorios, institutos, centros
- instituciones cooperantes, socios, redes





# Visibilidad Web

A considerar:

- gestión de contenidos en línea
- gestión de portales (organizaciones con muchos portales)
- marketing y diseño web (dos elementos separados, pero muy relacionados)
- posicionamiento web, experiencia de usuario
- foros, blogósfera, wikis
- redes sociales, comunidades virtuales







# Visibilidad Web

- La publicación en OA es el primer paso.
- En el año 2010, Alma Swan\* relevó 31 trabajos que analizan el impacto del OA.
- Midió la cantidad de citas en múltiples disciplinas y sobre muestras de tamaños muy variables (desde cientos de artículos a 1.3 millones).
- Hipótesis: si está en abierto\*\*, incrementará el impacto de la investigación.

\* Swan, Alma (2010) The Open Access citation advantage: Studies and results to date.

\*\* visibilidad, accesibilidad, facilidad de encontrar





# Visibilidad Web

En su investigación, A. Swan realiza ciertas consideraciones:

- existen investigadores que no poseen acceso vía suscripción a todos los trabajos publicados que son relevantes en su campo, y esto puede influenciar su trabajo.
- si estos artículos se hicieran públicamente disponibles, estos autores los leerían
- que algunos de estos artículos serían relevantes y aplicables para el trabajo de estos investigadores (o sea, citables)
- que otros de estos artículos serían irrelevantes o inaplicables y no serían citados





# Visibilidad Web

De lo anterior se desprende que:

- no se espera que un artículo, por el sólo hecho de estar en abierto, tenga mayor impacto. La calidad (originalidad, relevancia, actualidad) sigue siendo un factor clave.
- Sí se espera el que OA aumente la cantidad de lectores, y produzca un aumento en las citas si el artículo posee los méritos necesarios.
- También se espera que el aumento de citas varíe según la disciplina y con el tiempo.





# Visibilidad Web

Resultado obtenido en el estudio: 27 a favor, 4 en contra

Detectó una importante variabilidad por disciplina:

- Física/astronomía, medicina, IT, agricultura, Computer Science: entre el 150 y 600% de incremento de citas con OA
- matemática, ingeniería eléctrica, ciencias políticas: incrementos entre entre 35 y 91%
- biología, filosofía: entre -5% (negativo) y 45% de incremento





# Visibilidad Web

**Estrategia:** entender qué y cómo se expone

- Identificar tipos de contenidos: trabajos de investigación, artículos de difusión general, noticias institucionales
- Relevar sus formatos (HTML, documentos de texto, imágenes, videos, animaciones)
- Identificar el público destinatario (según formato? según tipo?)
- Relevar los espacios de difusión web y herramientas: CMS, soft. de repositorios, de publicaciones periódicas, de congresos, de e-learning





# Visibilidad Web

Cómo se vinculan las áreas? Dependencias, programas, UI's, cátedras, departamentos, secretarías...

Relevar todos estos espacios es complejo: muchos espacios, evolucionan, surgen nuevos...





# Visibilidad Web

Podemos usar un crawler/web spider para mapear los sitios institucionales y estudiar cómo se interconectan

Esto nos permitirá identificar oportunidades de difusión más fácilmente, así como también detectar áreas para fortalecer

Ej: crawler4j, Nutch, Ankhor, Webphinx



The screenshot displays the Webphinx Crawler application. On the left, the 'Crawl' tab is active, showing the starting URL 'http://www.unlp.edu.ar', a depth of 5 hops, and a Breadth first crawling strategy. Below this, a graph visualization shows a central hub of pages connected to several smaller clusters of pages, representing the site's structure. On the right, a terminal window shows the crawler's log, detailing the progress of the crawl with timestamps and the names of the pages being processed, such as 'Escuela Anexa', 'Colegio Nacional', and 'Dirección'.

**Imagen 1:** Análisis del sitio web de la UNLP con Webphinx Crawler, usando BFS con profundidad 5. Grafo de subsitios y conexiones (izquierda), y log de avance (derecha).

<http://www.cs.cmu.edu/~rcm/webphinx/>





# Visibilidad Web

¿Dónde impacta la visibilidad web?

- Sociedad (transparencia, retorno)
- Investigadores (altmetrics)
- Rankings institucionales





# Visibilidad Web - Altmetrics

- El impacto de la producción científica se mide tradicionalmente a partir del Impact Factor
- El IF se aplica sobre las revistas, pero no analiza puntualmente del impacto de cada artículo
- Los índices tradicionales (IF, h-index) son lentos, a veces el impacto se refleja muchos años después
- La revisión por pares sigue siendo válida, pero "está empezando a mostrar sus años" (eventualmente, el paper se publicará en alguna revista)





# Visibilidad Web - Altmetrics

Nuevas herramientas para difusión de la prod. científica:

- revistas digitales de OA: pueden registrar las descargas y los accesos,
- repositorios institucionales: idem anterior, y a la vez interoperar y diseminar,
- gestores de bibliografía (mendeley, zotero), servicios orientados a ciertos documentos (slideshare, figshare, scribd,) y redes de investigadores (RG, academia),
- blogs (científicos) y redes sociales: FB, G+, Twitter, LinkedIn





# Visibilidad Web - Altmetrics

- Las Altmetrics buscan considerar estas herramientas a la hora de medir el impacto de una investigación
- Permiten obtener métricas mucho antes que los índices tradicionales
- No buscan reemplazar otras métricas, sino sumar una nueva métrica

## Algunos enlaces de interés

Manifiesto, Road map, About. <http://altmetrics.org>

Altmetrics 2014 Workshop, an ACM Web Science Conference Workshop. <http://altmetrics.org/altmetrics14>

Altmetrics Workshop 2012 y 2011. <http://altmetrics.org/altmetrics12> ; <http://altmetrics.org/workshop2011>

Altmetric.com: We make article level metrics easy (\$). <http://www.altmetric.com>

PLOS Article Level Metrics. <http://article-level-metrics.plos.org/alt-metrics/>





# Visibilidad Web - Webometrics

Ranking elaborado por el **CISC**

El Ranking de repositorios analiza diversos indicadores:



- Size (Google)
- Visibility (Majestic SEO y ahrefs)
- Altmetrics (Academia, FB, LinkedIn, Mendeley, RG, Slideshare, Twitter, Wikipedia, YouTube)
- Rich files (google pdf, doc/docx, ppt/pptx)
- Scholar (total, 2009-2013, pdf, pdf 2009-2013)



# Visibilidad Web - Webometrics



Size: el tamaño sí importa



**Web** Images News More ▾ Search tools

About 745,000 results (0.50 seconds)

Google promotion

**Try Google Webmaster Tools**

[www.google.com/webmasters/](http://www.google.com/webmasters/)

Do you own **sedici.unlp.edu.ar**? Get indexing and ranking data from Google.

**SEDICI - Repositorio de la Universidad Nacional de La Plata**

[sedici.unlp.edu.ar/](http://sedici.unlp.edu.ar/) ▾ [Translate this page](#)

Repositorio institucional de la Universidad Nacional de La Plata.

**SEDICI - Buscar**

[sedici.unlp.edu.ar/discover](http://sedici.unlp.edu.ar/discover) ▾ [Translate this page](#)





# Visibilidad Web - Webometrics

## Rich files

**Web** Images News More ▾ Search tools

About 20,600 results (0.34 seconds)

Tip: Search for **English** results only. You can specify your search language in Preferences

**[PDF] Documento completo - SeDiCI - Universidad Nacional de ...**

[sedici.unlp.edu.ar/.../Documento\\_completo.pdf?...1](http://sedici.unlp.edu.ar/.../Documento_completo.pdf?...1) ▾ Translate this page

Jul 15, 2014 - Monsalvo Antón, J. M. (Ed.) (2013), Sociedades urbanas y culturas políticas en la Baja Edad Media Castellana, Salamanca: Ediciones ...

**[PDF] 16 th International Conference and Exhibition on ... - SeDiCI**

[sedici.unlp.edu.ar/bitstream/handle/.../Documento\\_completo\\_\\_.pdf?...1](http://sedici.unlp.edu.ar/bitstream/handle/.../Documento_completo__.pdf?...1) ▾





# Visibilidad Web - Webometrics

**Visibility:** enlaces entrantes

- El impacto se mide según cuántos sitios externos nos enlazan
- Existen herramientas específicas para medir sitios y registrar backlinks. Ej. ahrefs, MajesticSEO.
- Son pagas, pero proveen servicios gratuitos básicos que sirven para comenzar

¿Priorizamos enlaces externos o tamaño del dominio?







<https://ahrefs.com/site-explorer/overview/subdomains/?target=sedici.unlp.edu.ar>



SITE EXPLORER    SEO REPORT    BACKLINKS REPORT    APPS    LABS    PLANS & PRICING

- OVERVIEW
  - Dashboard
  - Top Pages
- INBOUND LINKS
  - Links
    - New / Lost
  - Anchors
  - Referring Domains
    - New / Lost
  - Referring IPs
- OUTGOING LINKS
  - Linked Domains
  - Anchors
- EXPORT
  - CSV
  - PDF

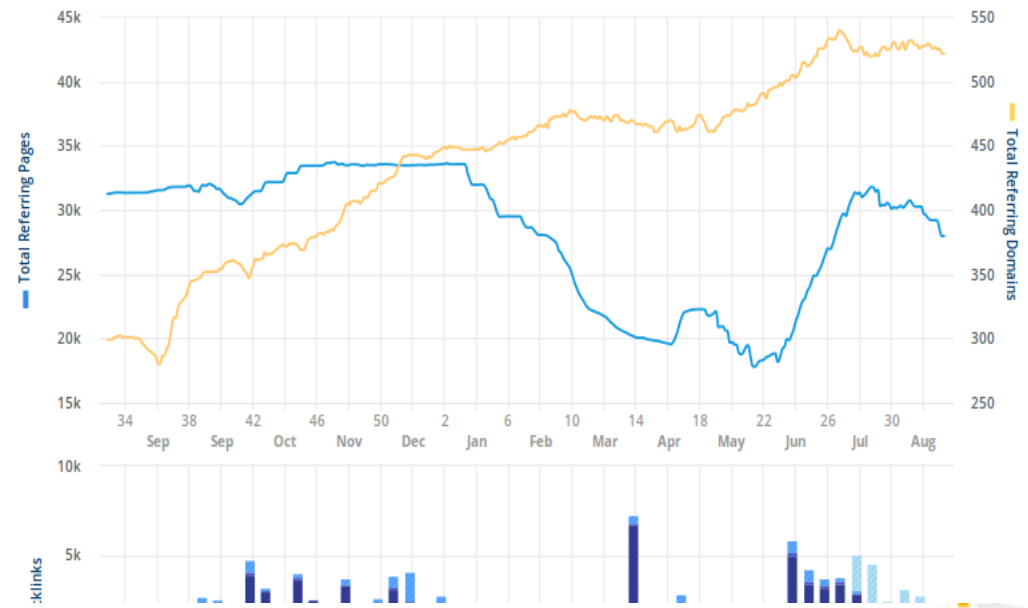
URL Rank **38**    Ahrefs Domain Rank **65**    Backlinks **41 K**    Referring Domains **522**

Referring Pages	27,984
Total Backlinks	41,298
Crawled Pages	133,810
Referring IPs	432
Referring Subnets	390
Referring Domains	522
Governmental	8
Educational	64
.com	149
.net	36
.org	82

Backlink Types:	
text	27,817
dofollow	26,096

### Referring Pages

JUN JUL AUG ALL ONE YEAR LAST 30 DAYS



PREBI



<http://www.majesticseo.com/reports/site-explorer?q=sedici.unlp.edu.ar&oq=sedici.unlp.edu.ar>

backlinks	Domains	IPS	Subnets
<b>45,537</b>	<b>457</b>	<b>396</b>	<b>374</b>

In the last 5 years

<b>106,276</b>	<b>1,540</b>	<b>938</b>	<b>793</b>
----------------	--------------	------------	------------

**Homepage Title** **SEDICI - Repository of National University of La Plata**

**Educational Ref. Backlinks** **2,073**

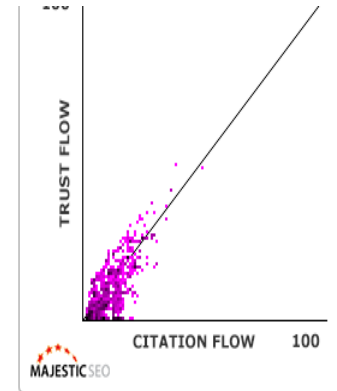
**Governmental Ref. Backlinks** **111**

**Educational Ref. Domains** **74**

**Governmental Ref. Domains** **13**

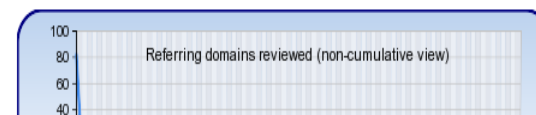
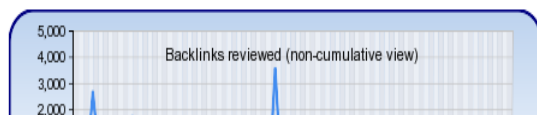
**Indexed URLs** **374,512**

**Position in Majestic Million** This site is NOT in the Majestic Million.



External Backlinks Link Profile

### Backlink History



# Visibilidad Web - Herramientas



Más allá de los rankings, estas herramientas son muy útiles para evaluar nuestra visibilidad y tener un panorama acerca de cómo nos ven, cómo llegan a nosotros, qué contenidos impactan más que otros.

Como todo en internet, hay muchas herramientas más.





# Visibilidad Web - Herramientas

**Google Webmaster Tools** (nuestro sitio visto por el crawler de Google)

- Nos da un panorama más o menos completo de nuestro sitio web (según Google):
- Errores de sus crawlers (sitio caído? servidor colapsado?)
- URLs indexadas
- Búsquedas en Google (impresiones, clics)





sedici.unlp.edu.ar

Help



### Search Queries

Top queries

Top pages

Filters

Web X

Jul 13, 2014

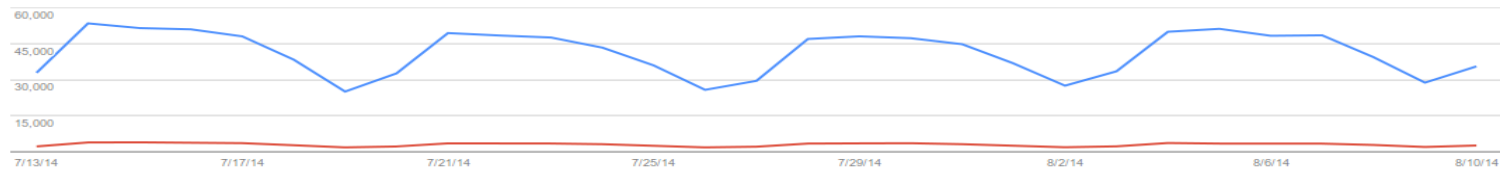
to

Aug 12, 2014

Queries  
15,794

Impressions  
1,197,467 -24%  
Displaying 80,818

Clicks  
86,311 -28%  
Displaying 10,200



Download this table

Download chart data

Basic

With change

Show

25 rows

1-25 of 15,794



Query	Impressions	Change	Clicks <span style="color:red">▲</span>	Change	CTR	Change	Avg. position	Change
☆ sedici	1,141	<span style="color:green">+23%</span>	337	<span style="color:red">-35%</span>	30%	<span style="color:red">-30</span>	5.0	<span style="color:red">-1.0</span>
☆ fisicoquimica pdf	320	<span style="color:green">+31%</span>	139	<span style="color:green">+20%</span>	43%	<span style="color:red">-4.0</span>	2.1	<span style="color:green">+0.9</span>
☆ razonamiento juridico	289	<span style="color:red">-44%</span>	88	<span style="color:red">-67%</span>	30%	<span style="color:red">-20</span>	2.2	<span style="color:red">-0.8</span>
☆ aseguramiento de la calidad del software	190	<span style="color:green">+38%</span>	80	<span style="color:green">+54%</span>	42%	<span style="color:green">+4.0</span>	3.2	<span style="color:red">-0.2</span>
☆ sedici unlp	84	<span style="color:red">-14%</span>	69	<span style="color:red">-9%</span>	82%	<span style="color:green">+5.0</span>	1.0	
☆ libros de fisicoquimica pdf	103	<span style="color:green">+56%</span>	58	<span style="color:green">+32%</span>	56%	<span style="color:red">-10</span>	3.1	<span style="color:red">-0.8</span>

**Imagen 1:** Principales búsquedas en Google donde se muestran enlaces a SEDICI, durante el receso invernal de 2014. Observar la baja generalizada en impresiones y en clics ( $\text{Clickthrough Rate} = \text{Clicks} / \text{Impressions}$ ).  
*Fuente Google Webmaster Tools*





## Links to Your Site

Total links  
**13,441,617**

### Who links the most

<a href="http://unlp.edu.ar">unlp.edu.ar</a>	13,040,261
<a href="http://siu.edu.ar">siu.edu.ar</a>	124,846
<a href="http://gob.ar">gob.ar</a>	68,626
<a href="http://163.10.0.0">163.10.0.0</a>	53,063
<a href="http://scoop.it">scoop.it</a>	19,647

[More »](#)

### How your data is linked

[acceder](#)

[sedici servicio de difusión de la creación intelectual](#)

[libros de cátedra](#)

[haga click aquí para acceder al objeto digital](#)

[eventos](#)

[More »](#)

### Your most linked content

<a href="http://sedici.unlp.edu.ar/">http://sedici.unlp.edu.ar/</a>	363,717
<a href="/handle/10915/27874">/handle/10915/27874</a>	149,141
<a href="/pages/comoAgregarTrabajos">/pages/comoAgregarTrabajos</a>	119,911
<a href="/pages/comoLlegar">/pages/comoLlegar</a>	117,470
<a href="/browse?type=author&amp;rpp=5">/browse?type=author&amp;rpp=5</a>	65,596

[More »](#)

**Imagen 2:** Enlaces entrantes a SEDICI. Observar los contenidos más enlazados: 1) Home, 2) Colección Libros de Cátedra, 3) Instructivo para agregar trabajos, 4) Instrucciones para llegar a la oficina, 5) Búsqueda por autor . Fuente Google Webmaster Tools





# Visibilidad Web - Herramientas

**Analytics:** nos ayudan a entender cómo nuestros visitantes interactúan con nuestros sitios

Google tiene su servicio **Google Analytics**, gratuito y muy completo

Si no queremos depender de Google, existen herramientas libres que podemos instalar. Por ejemplo:

- Piwik: similar a Google Analytics (LAMP)
- AWStats: análisis de logs





# Visibilidad Web - Redes sociales

- Permiten alcanzar un público inmediato muy amplio (amigos, followers, groups, pages)
- Permiten alcanzar un público mucho más amplio a partir de re-comparticiones (retwits, shares, +1)
  - Pero... qué compartimos y dónde?

Ej. El Repositorio Institucional tiene tesis, artículos, publicaciones en congresos

A mis amigos, tíos y abuelos, no les interesa leer una tesis doctoral o un artículo de una revista!

El idioma de los contenidos también importa







# Visibilidad Web - Redes sociales

Adaptarse al entorno: las redes sociales son distintas, LinkedIn "permite" más contenido académico, Facebook no tanto, Research Gate/Academia/Mendeley mucho.

- Algunos contenidos son mucho más llamativos para el público en general: libros, audios, videos, entrevistas
- Otros están más orientados a un público más especializado: papers, ponencias, tesis, reportes





# Visibilidad Web - Redes sociales

Aprovechar eventos circunstanciales, efemérides, noticias

- compartir colecciones y trabajos puntuales
- etiquetar autores si es posible, y posibles interesados directos
- promover debates, involucrar y fidelizar a la audiencia: por ejemplo, un blog
  - <http://sedici.unlp.edu.ar/blog>: noticias institucionales, acceso abierto, derechos de autor, materiales destacados





# Visibilidad Web - Wikipedia

Wikipedia es un excelente lugar para difundir trabajos científicos y académicos

Muy interesante para editores/publicadores de revistas, y para repositorios institucionales

Permite ofertar contenido propio en un espacio muy visible:

- mejorando artículos: más complejo, aporte más interesante
- aportando referencias de calidad: complejo
- aportando enlaces externos: más fácil, más volátil también
  - ¡Cuidado con el copyright! CC-BY o CC-BY-SA



# Visibilidad Web - Wikipedia



Varios trabajos de investigación reconocen en Wikipedia un gran potencial para incrementar la visibilidad y el impacto

El objetivo no es hacer publicidad en Wikipedia, sino mejorar esta enciclopedia

Using Wikipedia to Enhance the Visibility of Digitized Archival Assets, Michael Szajewski  
<http://www.dlib.org/dlib/march13/szajewski/03szajewski.html>

- Wikipedia and institutional repositories: an academic symbiosis? Alastair G. Smith,  
[http://researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/1848/Conference\\_paper.pdf?sequence=1](http://researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/1848/Conference_paper.pdf?sequence=1)
- Increasing the Discoverability of Digital Collections Using Wikipedia - A Review with deep analysis, Ed Galloway, Cassandra DellaCorte <http://palrap.pitt.edu/ojs/index.php/palrap/article/view/60>
- Wikipedia Lover, Not a Hater: Harnessing Wikipedia to Increase the Discoverability of Library Resources, Danielle Eldera, R. Niccole Westbrookb & Michele Reillyb  
<http://www.tandfonline.com/doi/abs/10.1080/19322909.2012.641808#preview>





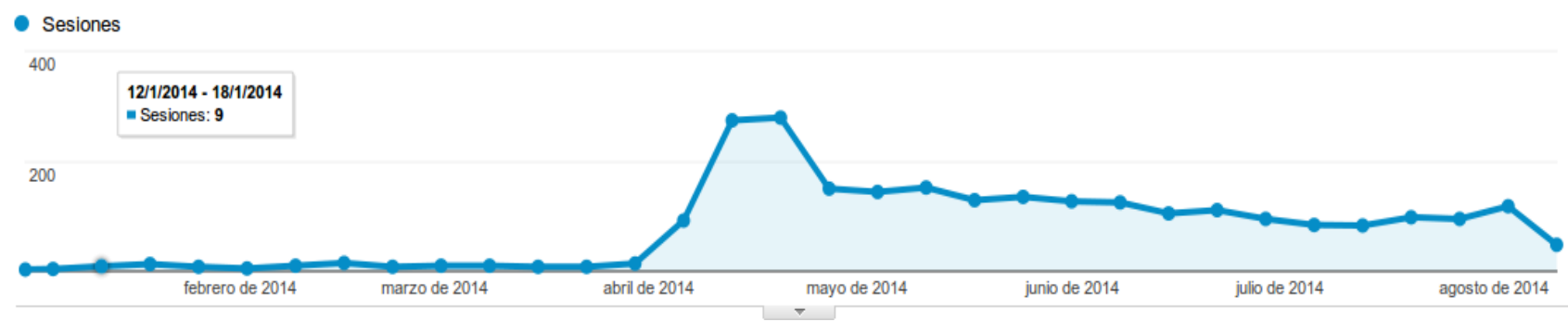
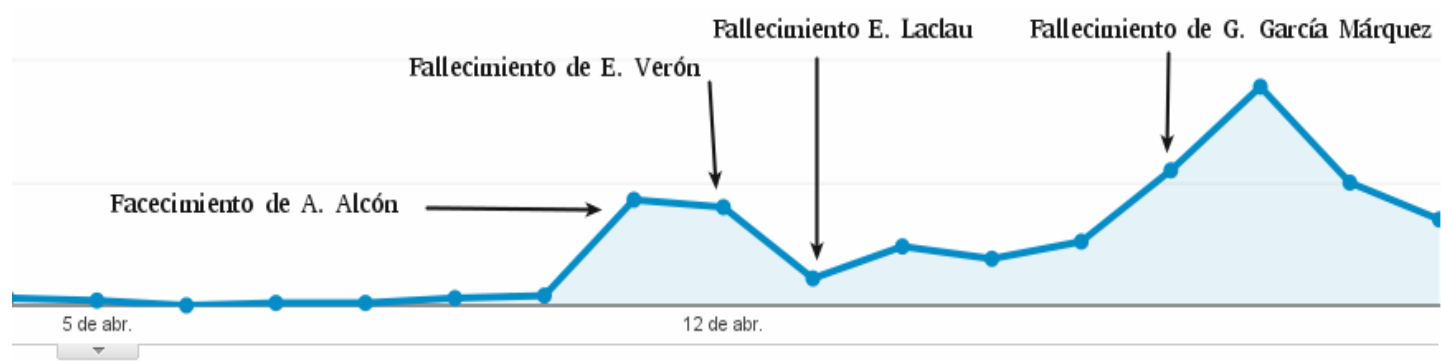
# Visibilidad Web - Wikipedia

Durante 2014, varios artículos en Wikipedia fueron editados, incorporando contenidos relevantes alojados en SEDICI

Muchas ediciones se vincularon con fallecimientos recientes (Eliseo Verón, Alfredo Alcón, Ernesto Laclau, Gabriel García Márquez)

Otras surgían a partir de tesis y publicaciones que se depositaban en el repositorio





Incremento de visitas recibidas desde Wikipedia durante el mes de abril de 2014, acumulado por día (arriba). Acceso sostenido desde abril 2014 a agosto 2014, acumulado por semana (abajo). Fuente: Google Analytics.





Fuente ?	Adquisición		
	Sesiones ? ↓	% de nuevas sesiones ?	Nuevos usuarios ?
	<b>41.711</b> % del total: 9,44% (442.077)	<b>74,47%</b> Promedio del sitio: 84,37% (-11,73%)	<b>31.062</b> % del total: 8,33% (372.968)
<input type="checkbox"/> 1. unlp.edu.ar	<b>4.782</b> (11,46%)	<b>69,91%</b>	<b>3.343</b> (10,76%)
<input type="checkbox"/> 2. facebook.com	<b>3.914</b> (9,38%)	<b>73,20%</b>	<b>2.865</b> (9,22%)
<input type="checkbox"/> 3. biblioteca.universia.net	<b>3.035</b> (7,28%)	<b>82,17%</b>	<b>2.494</b> (8,03%)
<input type="checkbox"/> 4. bdu.siu.edu.ar	<b>1.877</b> (4,50%)	<b>70,54%</b>	<b>1.324</b> (4,26%)
<input type="checkbox"/> 5. ms.gba.gov.ar	<b>1.832</b> (4,39%)	<b>68,50%</b>	<b>1.255</b> (4,04%)
<input type="checkbox"/> 6. scirus.com	<b>1.631</b> (3,91%)	<b>87,68%</b>	<b>1.430</b> (4,60%)
<input type="checkbox"/> 7. dialnet.unirioja.es	<b>1.125</b> (2,70%)	<b>87,11%</b>	<b>980</b> (3,15%)
<input type="checkbox"/> 8. scholar.google.es	<b>758</b> (1,82%)	<b>81,27%</b>	<b>616</b> (1,98%)
<input type="checkbox"/> 9. scholar.google.com	<b>625</b> (1,50%)	<b>83,36%</b>	<b>521</b> (1,68%)
<input type="checkbox"/> 10. revistas.unlp.edu.ar	<b>616</b> (1,48%)	<b>67,69%</b>	<b>417</b> (1,34%)

Fuente ?	Sesiones ? ↓	% de nuevas sesiones ?	Nuevos usuarios ?
	<b>37.792</b> % del total: 10,15% (372.515)	<b>76,85%</b> Promedio del sitio: 83,41% (-7,86%)	<b>29.042</b> % del total: 9,35% (310.700)
1. facebook.com	<b>4.283</b> (11,33%)	<b>73,71%</b>	<b>3.157</b> (10,87%)
2. unlp.edu.ar	<b>3.305</b> (8,75%)	<b>69,20%</b>	<b>2.287</b> (7,87%)
3. es.wikipedia.org	<b>2.569</b> (6,80%)	<b>90,11%</b>	<b>2.315</b> (7,97%)
4. biblioteca.universia.net	<b>1.601</b> (4,24%)	<b>82,20%</b>	<b>1.316</b> (4,53%)
5. ms.gba.gov.ar	<b>1.526</b> (4,04%)	<b>69,59%</b>	<b>1.062</b> (3,66%)
6. bdu.siu.edu.ar	<b>1.337</b> (3,54%)	<b>72,10%</b>	<b>964</b> (3,32%)
7. revistas.unlp.edu.ar	<b>1.143</b> (3,02%)	<b>69,82%</b>	<b>798</b> (2,75%)
8. dialnet.unirioja.es	<b>943</b> (2,50%)	<b>86,43%</b>	<b>815</b> (2,81%)
9. l.facebook.com	<b>931</b> (2,46%)	<b>81,95%</b>	<b>763</b> (2,63%)
10. agro.unlp.edu.ar	<b>925</b> (2,45%)	<b>67,46%</b>	<b>624</b> (2,15%)

**Izquierda:** Accesos desde sitios externos durante todo el año 2013. Observar que Wikipedia no figura en el top 10. **Derecha:** Accesos a SEDICI desde sitios externos entre enero y julio de 2014. La versión en español de Wikipedia ascendió a la posición 3 en 7 meses (en realidad 4: abril-julio). Fuente Google Analytics.





<input type="checkbox"/>	1. <a href="#">/wiki/Gabriel_García_Márquez</a>		<b>287</b> (11,17%)	<input type="checkbox"/>	10. <a href="#">/wiki/Eliseo_Verón</a>		<b>56</b> (2,18%)
<input type="checkbox"/>	2. <a href="#">/wiki/Cien_años_de_soledad</a>		<b>283</b> (11,02%)	<input type="checkbox"/>	11. <a href="#">/wiki/Jacques_Le_Goff</a>		<b>55</b> (2,14%)
<input type="checkbox"/>	3. <a href="#">/wiki/Alfredo_Alcón</a>		<b>139</b> (5,41%)	<input type="checkbox"/>	12. <a href="#">/wiki/Juego</a>		<b>54</b> (2,10%)
<input type="checkbox"/>	4. <a href="#">/wiki/Servicio_de_Difusión_de_la_Creación_Intelectual</a>		<b>120</b> (4,67%)	<input type="checkbox"/>	13. <a href="#">/wiki/Ana_María_Matute</a>		<b>52</b> (2,02%)
<input type="checkbox"/>	5. <a href="#">/wiki/Ernesto_Laclau</a>		<b>86</b> (3,35%)	<input type="checkbox"/>	14. <a href="#">/wiki/Julio_Cortázar</a>		<b>51</b> (1,99%)
<input type="checkbox"/>	6. <a href="#">/wiki/Scrum</a>		<b>82</b> (3,19%)	<input type="checkbox"/>	15. <a href="#">/wiki/Prueba_unitaria</a>		<b>47</b> (1,83%)
<input type="checkbox"/>	7. <a href="#">/wiki/Escherichia_coli</a>		<b>78</b> (3,04%)	<input type="checkbox"/>	16. <a href="#">/wiki/Álgebra_de_Virasoro</a>		<b>43</b> (1,67%)
<input type="checkbox"/>	8. <a href="#">/wiki/Índice_de_vegetación_de_diferencia_normalizada</a>		<b>77</b> (3,00%)	<input type="checkbox"/>	17. <a href="#">/wiki/Repositorio</a>		<b>37</b> (1,44%)
<input type="checkbox"/>	9. <a href="#">/wiki/Sistema_de_posicionamiento_global</a>		<b>65</b> (2,53%)	<input type="checkbox"/>	18. <a href="#">/wiki/Cerámica_técnica</a>		<b>33</b> (1,28%)
				<input type="checkbox"/>	19. <a href="#">/wiki/Nanopartícula</a>		<b>31</b> (1,21%)

Artículos en Wikipedia que mayores accesos generan. Período Enero-Julio 2014. Fuente Google Analytics.







# Visibilidad Web e Interoperabilidad

Mediante interoperabilidad (OpenSearch, RSS/Atom, webservice, OAI), podemos exponer nuestros materiales en múltiples espacios

## Por ejemplo:

- blogs institucionales (que se llevan muy bien con las redes sociales)
  - programas institucionales: educación a distancia, radio, editorial
  - sitios de laboratorios\*, departamentos, cátedras
  - bases de datos, agregadores e índices internacionales
- \*Orduña-Malea, Enrique, Cabezas-Clavijo, Álvaro, Delgado López-Cózar, Emilio. Visibilidad e impacto web de los grupos de investigación de información y documentación en las universidades públicas españolas. 2013. <http://hdl.handle.net/10481/31303> .





# Visibilidad Web

Algunos artículos de interés en la web:

- Alma Swan and Leslie Carr. Institutions, their repositories and the Web. Serials review, 2008. <http://eprints.soton.ac.uk/264965/>
- Kenning Arlitsch and Patrick O'Brien. Improving Institutional Repository Search Engine Visibility in Google and Google Scholar. <http://www.slideshare.net/nomadusg/improving-institutional-repository-search-engine-visibility-in-google-and-google-scholar>
- Google Scholar and Dspace. Bram Luyten, Enero 2014. <http://atmire.com/website/?q=content/google-scholar-and-dspace>



# Visibilidad Web



Hay mucho trabajo por hacer: medir, analizar, vincular, repetir.

La interoperabilidad simplifica algunas tareas

pero...

El **acceso abierto** asegura la disponibilidad de los contenidos, maximiza su diseminación y su impacto

