

ESCUELA COMPLUTENSE LATINOAMERICANA

*preservación
digital de
documentos*
archivos, bibliotecas y museos

Acceso abierto, repositorios, preservación, estándares

Dra. Marisa R. De Giusti

Universidad Nacional de La Plata

marisa.degiusti@sedici.unlp.edu.ar

Miércoles, 25 Febrero de 2015

OBJETIVO GENERAL del curso

- Conocer los múltiples conceptos, procesos y actitudes que conlleva la creación, gestión y funcionamiento de las colecciones digitales en forma de bibliotecas o repositorios. Para ello se revisará la definición e identificación de la estructura externa e interna de los objetos digitales; sus problemas de longevidad en el tiempo; su obsolescencia tecnológica; los mecanismos y técnicas que permiten y aseguran su perdurabilidad; y el concepto de acceso a la información como realidad que mejor motiva la elaboración de políticas específicas de conservación digital permanente.

OBJETIVOS ESPECÍFICOS de esta clase

- Definir los orígenes de los RIs, definir los repositorios e informar sobre los alcances de los repositorios como herramientas de difusión de los documentos digitales
- Crear conciencia en los profesionales, usuarios, funcionarios de bibliotecas y archivos, políticos, investigadores, etc., sobre los riesgos que conlleva mantener en el tiempo los objetos digitales y dar accesibilidad permanente a los mismos.
- Analizar formatos y estándares para cumplir con el objetivo de preservación.

CONTENIDOS de esta clase

1. Movimiento de acceso abierto y conceptos básicos de repositorios digitales
2. Formatos estandarizados para la conservación y difusión de objetos digitales
3. Qué es la preservación de objetos digitales. El modelo de referencia OAIS, Norma ISO 14721.

Sección 1



Movimiento de Acceso Abierto



Tiene como fin asegurar el acceso libre y abierto más amplio posible a la producción científica. Una de las formas de lograr ese objetivo es por medio de la creación de **repositorios institucionales** donde se deposita esa producción científica para hacerla accesible sin restricciones y preservarla digitalmente como un bien común para la sociedad de hoy y del futuro.

Open access: empezó hace mucho!



1966 dos hechos fundacionales:

- El lanzamiento de ERIC (Educational Resources Information Center) por el Departamento de Educación en Estados Unidos y
- El lanzamiento de **Medline** (disponible en la red en 1997) por la Biblioteca Nacional de Medicina de ese mismo país.
 - Los inicios del movimiento por el acceso libre a la literatura científica revisada por pares se encarnan con la historia misma de Internet.

Open access: empezó hace mucho!



- **Hasta los 90's** la historia recoge numerosos hitos, de los cuales destaca en Julio de 1987 el Proyecto Perseo, destinado a reunir materiales para el estudio de la Antigua Grecia, lanzado en CDs y libre en la web recién en 1994.
- **1991 – ArXiv** <http://arxiv.org/>

Una fecha relevante la constituye el lanzamiento por Paul Ginsparg en Agosto de 1991 de ArXiv que, definido de una forma sencilla por su creador no es otra cosa que "*un sistema de distribución automática de artículos de investigación, sin las operaciones editoriales asociadas a la revisión por pares*".

Stevan Harnad (1990-1992)

Una diferencia crucial entre los *preprints* en el mundo digital y los del mundo impreso es que los archivos digitales podían tener mecanismos que permitirían, todavía dentro de la fase informal de pre-publicación, una retroalimentación inmediata por parte de colegas a través de un sistema de comentarios. Este sistema abierto de revisión por pares enriquecería el artículo y a su vez permitiría que el proceso de investigación fuese más abierto. Harnad llamó a esto “Scholarly Skywriting”: lo que permitiría que los colegas enviaran *preprints* de sus manuscritos a un servidor central, el cual contaría con herramientas tecnológicas para la retroalimentación inmediata a través de comentarios , antes de la publicación en una revista académica.

Stevan Harnad (1990-1992)

- Harnad no cuestionaba la utilidad del proceso de publicación formal, sino su capacidad y rapidez comunicacional.
- La publicación académica tiene dos funciones que vale la pena distinguir: la publicación como práctica comunicativa en dónde la principal prioridad es ser leído por colegas, y la publicación que tiene una perspectiva funcionalista y sirve para dar estatus y autoría, asignar recursos y formar un registro (Kling and McKim 1999).
- La distribución de *preprints* obedecía a la primera.

1993 - WWW

El CERN anuncia la posibilidad de libre uso de la tecnología WWW sin cargo alguno. Esta institución juega un papel fundamental en el entramado de apoyo al movimiento desde todos los ángulos: poniendo a disposición del mismo software libre, cumpliendo tareas de repositorio con sus propios servidores para los artículos de investigación y alojando las iniciativas europeas en el ámbito.

Stevan Harnad 1994

- En 1994 Stevan Harnad, lanza la iniciativa por el *autoarchivo* (self-archiving): el propio autor deposita su trabajo en un lugar confiable. Recoge la iniciativa ArXiv y la desarrolla para su aplicación en otros campos y la piensa ya no sólo para pre prints sino para trabajos con revisión.
- Analiza el impacto para la comunidad científica del depósito de artículos científicos en un archivo de acceso abierto (en aquel entonces, un sitio FTP). Este trabajo suscitó una discusión sobre todo el sistema de comunicación científica y constituyó uno de los primeros reclamos por hacer más abierta la ciencia.



School of Electronics and
Computer Science versity of
Southampton.

1997

- La Asociación de Bibliotecas de Investigación de los Estados Unidos (Association of Research Libraries-ARL) pone en funcionamiento la iniciativa denominada **Scholarly Publishing & Academic Resources Coalition (SPARC)** que es una alianza internacional que trabaja para corregir el desequilibrio del sistema de edición científica y estimular la emergencia de nuevos modelos de comunicación académica.
- Se lanza **CogPrints** el primer depósito de artículos de investigación en las áreas de psicología, neurociencias, lingüística, filosofía y ciencias de la computación.
- Se inicia el acceso libre al **MEDLINE** a través del Pubmed, iniciativa de los decanos universitarios en Estados Unidos que aboga por el acceso libre a los resultados de la investigación científica en todos los campos.

Problemas

- A medida que crecieron los acervos a partir de diferentes iniciativas, surgieron dificultades de interoperabilidad entre servidores y algunos problemas tales como el de realizar una búsqueda en más de un acervo. La necesidad de infraestructura que apoyase la iniciativa en los aspectos precedentes y otros similares hizo avanzar la rueda...

Convención de Santa Fe, Albuquerque, USA, 1999

Destinada a proponer una serie de principios organizativos y especificaciones técnicas para permitir que los diversos sistemas de archivo y publicación fueran interoperables.

En la convención de Santa Fe un resultado adicional fue desarrollar software para de este modo facilitar la puesta en marcha de repositorios, se enunciaron a la vez las características o las funcionalidades que debía reunir el software a desarrollar:

- Un mecanismo de depósito;
- Un sistema de almacenamiento a largo plazo;
- Un sistema de gestión política con respecto a la presentación de documentos y su conservación;
- Una interfaz simple que permitiera a terceros recopilar metadatos de recursos provenientes de distintas fuentes (OAI-PMH).

2000

- Se crea un archivo central de la literatura biomédica similar al Pubmed, se plasma con la creación por parte de la Biblioteca Nacional de los Estados Unidos de **Pubmed Central**.
- Se destaca el lanzamiento de la primera iniciativa de un editor privado: **BiomedCentral** que en la actualidad edita un importante número de revistas accesibles a texto completo y que pone en marcha un sistema de financiación que, parece va a generalizarse: son los propios autores los que pagan una cuota por artículo publicado en la revista.

2000 - PLoS

La carta planteaba que *“los editores de nuestras revistas científicas tienen legítimo derecho a obtener una recompensa financiera justa por su papel en la comunicación científica. Creemos, sin embargo, que el registro de las ideas y la investigación científica no deben pertenecer ni ser controladas por los editores, sino que deben pertenecer al público y deben estar disponibles libremente a través de una biblioteca pública online”*.

La carta amenazaba a las revistas con dejar de publicar en ellas o de arbitrar los artículos- a menos que en septiembre del 2001 comenzaran a hacer disponibles sus contenidos (6 meses después de publicados) en *PubMed Central* u otro sitio web similar. La carta la firmaron más de 30.000 científicos de todo el mundo.

PloS que se convirtió en una editorial de acceso abierto

2000

La Universidad de Southampton lanza **Eprints**, un sistema de publicación y depósito de archivos digitales, de código abierto y libre, para la creación de depósitos de literatura científica siguiendo los postulados establecidos en Santa Fe.



Por qué el AA

- Pocas empresas editoriales monopolizan alrededor de 60% de la edición y el acceso a las revistas más importantes en el mundo.
- Aumento de precios de suscripciones a bases de datos comerciales y disminución de los presupuestos para la compra.
- Los científicos tienen que pagar por su propia producción.
- Poca visibilidad de los resultados de investigación.
- Los avances tecnológicos (internet, protocolos, software.) que facilitan y favorecen el desarrollo de iniciativas y proyectos.
- Las diferentes declaraciones, donde se exhorta a que se hagan del “Acceso Abierto” una prioridad en las políticas de desarrollo científico.
- Buena parte de la investigación se realiza con fondos públicos, por lo que debería estar a disposición de toda la comunidad académica, sin pago alguno.

Declaraciones: base de OA las tres B's

En diciembre del 2001, el *Open Society Institute* organizó una reunión en Budapest donde participaron actores que tuvieron gran influencia en el movimiento a favor del acceso abierto. Resultado de esta reunión fue la

- **Budapest Open Access Initiative (2/2002)**

www.soros.org/openaccess/read.shtml

- **Bethesda Statement on Open Access Publishing (6/2003)**

http://ictlogy.net/articles/bethesda_es.html

- **Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (10/2003)**

http://www.geotropico.org/1_2_documentos_berlin.html

Budapest OA Initiative- Open Access

En ella se recomiendan las modalidades de publicación en revistas de acceso abierto o bien a través del autoarchivo en archivos electrónicos abiertos.



Budapest

Esta declaración señala que “por acceso abierto a la literatura científica, entendemos su disponibilidad gratuita en Internet, para que cualquier usuario la pueda leer, descargar, copiar, distribuir o imprimir, con la posibilidad de buscar o enlazar al texto completo, recolectar los artículos para su indexación, pasarlos como datos para software o utilizarlos para cualquier otro propósito legítimo, sin más barreras financieras, legales o técnicas que aquellas que supongan acceder a Internet. El único límite a la reproducción y distribución de los artículos publicados y la única función del copyright en este marco, no puede ser otra que garantizar a los autores el control sobre la integridad de su trabajo y el derecho a ser reconocido y citado”.



Bethesda

La **Declaración de Bethesda sobre Publicación de Acceso Abierto**

añade que “para que un trabajo sea de acceso abierto, tiene que cumplir dos condiciones:

1. El/los autor/es y el/los propietario/s de los derechos de propiedad intelectual otorgan a los usuarios un derecho libre, irrevocable, universal y perpetuo de acceso y licencia para copiar, utilizar, distribuir, transmitir y presentar el trabajo públicamente y hacer y distribuir obras derivadas, en cualquier soporte digital para cualquier finalidad responsable, sujeto a la apropiada atribución de la autoría, así como el derecho de hacer una pequeña cantidad de copias impresas para su uso personal.

Bethesda

2. Una versión completa de la obra y todos los materiales suplementarios, incluyendo una copia de los permisos citados anteriormente, en un formato electrónico estándar apropiado se depositará de forma inmediata a la publicación inicial en al menos un repositorio en línea apoyado por una institución académica, una sociedad de intelectuales, una agencia gubernamental, o cualquier otra organización debidamente establecida que persiga facilitar el acceso abierto, la distribución sin restricciones, la interoperabilidad y el archivado a largo plazo (para las ciencias biomédicas, este repositorio es PubMed Central).

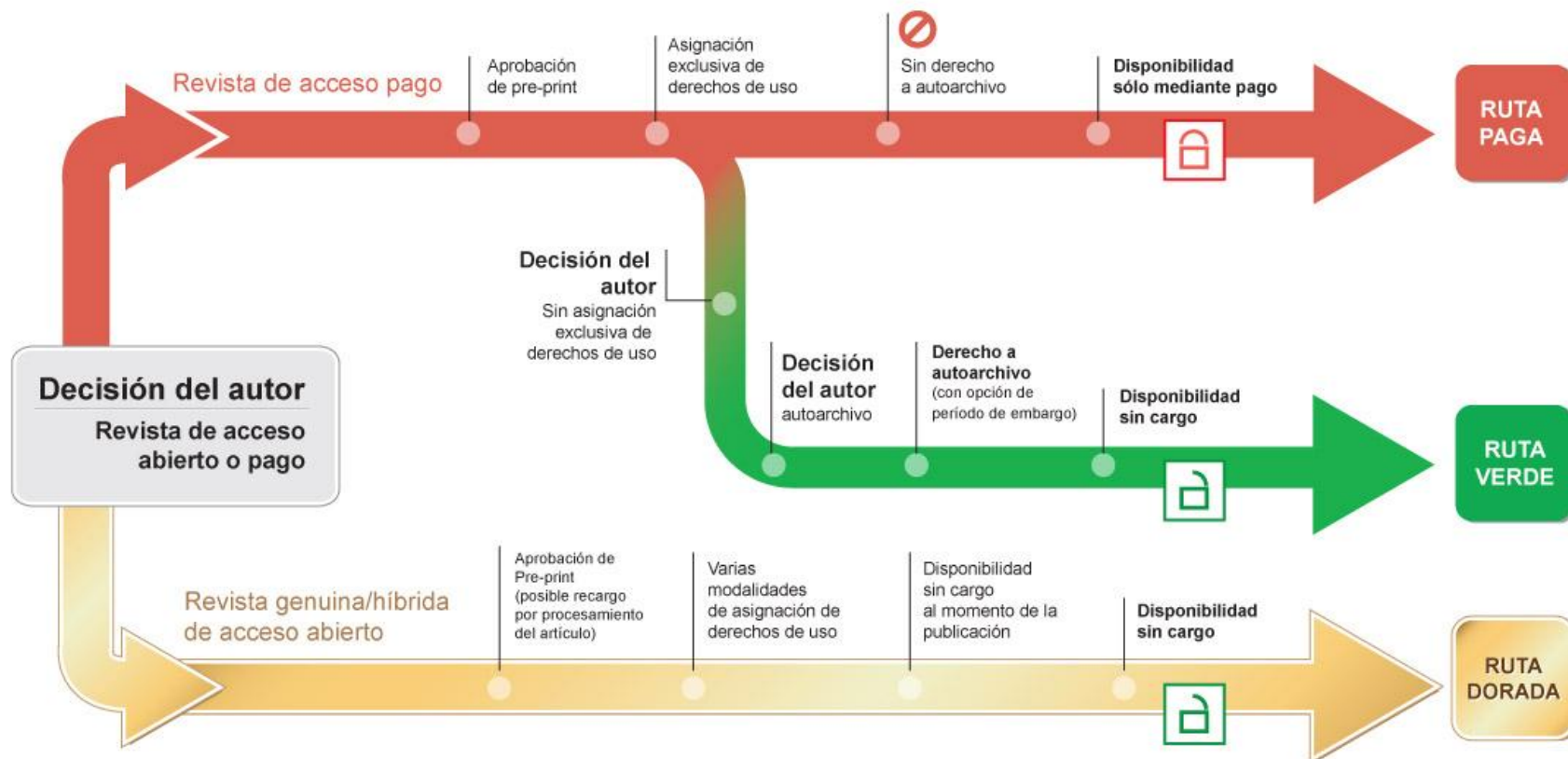
Berlín

- La Declaración de Berlín sobre el Acceso Abierto al Conocimiento en Ciencias y Humanidades, de octubre de 2003, ratifica todo lo anterior y aporta la perspectiva más teórica afirmando que “nuestra misión de diseminar el conocimiento será incompleta si la información no es puesta a disposición de la sociedad de manera rápida y amplia. Es necesario apoyar nuevas posibilidades de diseminación del conocimiento, no sólo a través de la manera clásica, sino también utilizando el paradigma de acceso abierto por medio de Internet”.

Berlín

- “Definimos el acceso abierto como una amplia fuente de conocimiento humano y patrimonio cultural aprobada por la comunidad científica. Para que se pueda alcanzar la visión de una representación del conocimiento global y accesible, la Web del futuro tiene que ser sustentable, interactiva y transparente. El contenido y las herramientas de software deben ser libremente accesibles y compatibles”.

Rutas de publicación



Fuente: Max Planck Society

Vía verde

Alude al autoarchivo o depósito de recursos digitales en repositorios institucionales o temáticos.



Materializan los principios de Acceso Abierto a través del depósito, difusión y visibilidad de los resultados de la actividad científica de una institución

Rol de los repositorios



Una institución, a través de su repositorio, se convierte en un proveedor de contenidos que ofrece la comunicación de sus resultados, su difusión y el acceso a los documentos que alberga, lo cual genera un mayor impacto y visibilidad de los mismos, y permite su uso compartido y reutilización, así como garantiza su preservación.

Melero (2008)

Características del RI



Pertenece a una institución



Es de ámbito académico



Es acumulativo y perpetuo



Es abierto e interactivo

SPARC, 2002

Definiciones del RI

Un archivo digital destinado a capturar y preservar la producción intelectual de una sola o de varias universidades para dar una respuesta convincente a dos cuestiones estratégicas que enfrentan las instituciones académicas: cambio de la comunicación científica y medición de impacto de la calidad con el fin de aumentar su visibilidad.



Raym Crow, 2002

SPARC

Definiciones del RI

Un repositorio institucional, es un conjunto de **servicios** ofrecidos a la comunidad científica para la gestión y diseminación de los materiales en forma digital generados por su comunidad.

Importante la consideración de **servicio** que se da al repositorio, porque demuestra su importancia más allá de servir de la plataforma de almacenamiento y de preservación.



Clifford Lynch, 2003

Clifford A. Lynch (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. <http://www.arl.org/newsltr/226/ir.html>

Definiciones del RI

Es un organismo vivo de la actividad institucional, inseparable de las prácticas de investigación cotidianas, por lo que los encargados de administrarlo deben hacerlo con cuidado y orgullo



Peter Suber (2007)

Peter Suber (2007). Panorama sobre el Acceso Abierto (AA).

<http://legacy.earlham.edu/~peters/fos/overview-spanish.pdf>

Definiciones del RI

Es una base de datos que las universidades e instituciones de investigación ponen a disposición de los autores. Es compatible con la Iniciativa OAI ([Open Access Initiative](#)), gracias a esto todos los metadatos de los artículos depositados son aprovechables y se pueden buscar como si hubieran sido depositados en único archivo virtual global.



Harnad, Stevan (2008)

Harnad, Stevan (2008) El movimiento mundial para el acceso libre: Maximizar el acceso al conocimiento para maximizar el progreso de los conocimientos. <http://eprints.soton.ac.uk/265700/>

Tipos de repositorios

Repositorios institucionales: son aquellos creados por las instituciones con el fin de depositar, preservar y dar visibilidad a la producción científica y académica propia.

Repositorios temáticos: son aquellos que reúnen documentos relacionados con un área temática específica.

Repositorios de datos: repositorios que almacenan, conservan y comparten los datos de las investigaciones.

Repositorios huérfanos: Repositorios, generalmente de orden nacional (ejemplo. Portugal), creados para depositar los trabajos de autores y/o instituciones que no tienen acceso o no pueden crear su propio R.I.

Repositorios Agregadores y Portales: Proveedores de servicios nacionales, regionales, etc., recolectan los metadatos de otros repositorios y ofrecen acceso y servicios asociados con la información recolectada.

BENEFICIOS con los RIs

Beneficios para los autores

- Difusión y mejora de visibilidad de las investigaciones realizadas.
- Disminución de las barreras de acceso a la literatura de frontera.
- Aumento de la visibilidad institucional e individual.
- Cuanto mayor sea la difusión y visibilidad de las publicaciones, mayores probabilidades de obtener impacto.
- Aunque esto no es directamente proporcional, ya existen estudios que demuestran que los artículos publicados en acceso abierto tienen más probabilidades de ser citados.
- Participar de la filosofía del acceso abierto permite a los autores realizar un uso más consciente de sus derechos de autor.
- Facilitan la integración en las redes de conocimiento.



**Los repositorios aumentan la
accesibilidad y visibilidad de la
producción científica del autor.**

Beneficios para las instituciones

- Servir de plataforma de almacenamiento e intercambio de la información producida en la Universidad
- Recopilar y preservar toda la producción científica, académica e institucional
- Maximizar la visibilidad de la producción científica, académica y corporativa
- Maximizar el impacto de esta producción en la sociedad y en la comunidad científica internacional mejorando el Ranking Web.
- Promover la cooperación internacional.



Los repositorios terminan convirtiéndose en un fiel reflejo de la actividad investigadora de la Institución, lo que influye positivamente en la reputación de la misma y en su posicionamiento en la Sociedad.

Beneficios para la sociedad

- Devuelve a la sociedad lo invertido en forma de resultados de investigación; la ciencia al fin y al cabo, no deja de estar financiada por todos y cada uno de nosotros.
- La investigación alcanza una mayor percepción social. Se analiza el impacto verdadero.
- Eliminación de barreras económicas, los países en vías de desarrollo tienen a su alcance la investigación financiada con fondos públicos de manera gratuita.
- Dado que la publicación en abierto es inmediata, se agiliza la comunicación científica.



Ahorros financieros

- Buena parte de ese ahorro se traduce en tiempo, que se ahorra cuando un artículo esta libre.
- Disminución del uso de los recursos por suscripción, las estadísticas de uso de los recursos electrónicos están bajando en las instituciones.



El asesoramiento de J. Wales (creador de la Wikipedia) al gobierno del Reino Unido para liberar artículos académicos, concluye que gracias al acceso abierto el sector público del Reino Unido se ahorrará 28 millones de libras.

El acceso abierto como modelo alternativo

- El acceso abierto ha traído consigo más datos, más tipos de resultados de investigación (no sólo artículos) y mayor granularidad para medir el impacto de la ciencia, poniendo en cuestión los modelos que la analizaban en un medio no digital y/o en sistemas cerrados
- Los editores ya no son los únicos que manejan datos relativos al impacto (repositorios, plataformas abiertas, interactivas, agregadores, web2.0...)
- Iniciativas para estándares internacionales abiertos y nuevos modelos de medición que superen el modelo dominante

Reformas en los sistemas de evaluación

- La principal razón de los repositorios es maximizar la visibilidad de la producción científica de la institución (Swan, 2008)
- Esto lo logran en gran medida gracias a la capacidad de interoperar con otros sistemas
 - Recolección de metadatos desde otros repositorios institucionales y agregadores nacionales e internacionales (SNRD, LAReferencia...)
 - Redes científicas (Research Gate, Academia.edu)
 - Google Scholar, Microsoft Academic
 - Compartición en redes sociales (Facebook, Twitter, Google Plus, LinkedIn...)
 - Exposición desde otros sitios institucionales (RSS, OpenSearch)

Medición de Impacto y Visibilidad de la producción científica con el RI

La medición del impacto y visibilidad abre un nuevo campo de estudio

- parámetros de visibilidad web: size, rich files, scholar, enlaces entrantes (Aguillo, 2006)
 - Ranking de repositorios
 - Ranking de Universidades (en muchos casos fuertemente influenciadas por sus repositorios)
- altmetrics (Galligan, 2013 ; Roemer, 2012) :
 - se suman a métodos tradicionales (ej. Factor de Impacto) como métricas más inmediatas y más sociales.
 - se calculan métricas alternativas basadas en
 - redes sociales
 - blogs científicos
 - video en línea (youtube, vimeo...)
 - Gestores de citas (Mendeley, Zotero)

Reformas en los sistemas de evaluación

- España
- Australia
- Inglaterra
- Argentina misma: Ley 26899

Sección 2



Preservación del contenido de los RIs



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.

Además del acceso abierto al material de investigación la preservación digital es una motivación importante para crear RIs y para asegurar que los materiales de investigación digitales estén disponibles y sean accesibles a largo plazo.

La preservación de los contenidos

- Criterios tradicionales para documentos en papel: en los documentos tradicionales en papel se habla de “negligencia benigna”: el olvido de un manuscrito en un arcón, puede que lo preserve.
- En los documentos digitales, la negligencia benigna no sirve: un disco olvidado 5 años probablemente al encontrarlo no se puede recuperar su información.
 - No a la negligencia benigna.
 - No a la preservación basada en las condiciones ambientales.
 - No se conserva para cualquier usuario futuro sino para una comunidad designada: el conjunto de los consumidores que tienen que entender la información almacenada.
 - No necesariamente se conserva la integridad externa del documento sino las propiedades significativas.
 - Se debe asegurar la integridad y autenticidad del recurso

Problemas en la preservación de OD

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 digital

- La preservación digital supone, en relación con la conservación de los documentos en papel, un importante reto tecnológico, pero también de otros tipos:
- legal, permisos de los autores para realizar las transformaciones necesarias
- económico, ¿quién financia el personal y las acciones para la preservación?,
- organizativo ¿de quién es la responsabilidad de cada acción? ¿cómo se asegura la continuidad de las decisiones?)

(Keefer; Gallart, 2007).

Etapas en la preservación

1. Archivar los documentos digitales
 - gestión documental
2. Preservar el *bitstream*
3. Garantizar el acceso a largo plazo

La preservación supone que:

1. Los datos se mantendrán en el repositorio sin sufrir daños, sin perderse o sin ser alterados de forma malintencionada/o no.
2. Los datos podrán ser localizados y entregados al usuario.
3. Los datos podrán ser interpretados y comprendidos por el usuario.
4. Las metas 1, 2 y 3 serán realizables a largo plazo.

Preservación digital

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

Noción de preservación de UNESCO



“La preservación digital puede definirse como el conjunto de los procesos destinados a garantizar la continuidad de los elementos del patrimonio digital durante todo el tiempo que se consideren necesarios”.

*“La mayor amenaza para la continuidad digital es la desaparición de los medios de acceso. No puede decirse que se han conservado los objetos digitales si, al haber dejado de existir los medios de acceso a ellos, resulta imposible utilizarlos. **El objetivo de la preservación de los objetos digitales es mantener su accesibilidad**, es decir, la capacidad de tener acceso a su mensaje o propósito esencial y auténtico”. (UNESCO, 2003: p. 37)*

Objeto digital

Acciones en su ciclo de vida para mantener el acceso

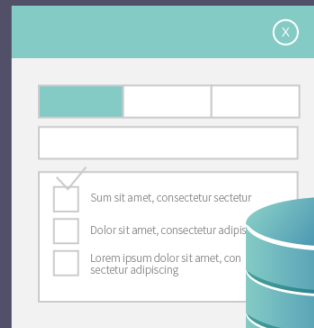
OD Y METADATOS DE PRESERVACIÓN

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

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



Autenticidad

Mediante la documentación de su procedencia

Debe conocerse su **creador**

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

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

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



Preservación de los contenidos de un RI

Criterios nuevos para los recursos digitales:

- que la institución tenga pleno derecho a manipular los datos para asegurar su acceso en entornos informáticos del futuro;
- que el recurso sea de un formato legible actualmente y previsiblemente en el futuro;
- que el recurso esté en un soporte gestionable para su transferencia y/o almacenamiento;
- que el recurso disponga de documentación, incluyendo los metadatos.

DPE Handbook

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 del contenido de los RIs

¿Qué materiales hay en los RIs?

resultados de la investigación (tesis, e-prints);
objetos de docencia y aprendizaje;
datos no elaborados;
fondos digitalizados;
material administrativo.

¿Qué materiales se tienen que preservar a largo plazo?

Criterios tradicionales para tomar la decisión sobre la preservación a largo plazo, principalmente los factores de

- valor
- pertinencia
- uso
- Otros condicionantes: misión, disponibilidad de recursos humanos, económicos, materiales, obligaciones legales o contractuales.

Preservación del contenido de los RIs

Selección de recursos para su preservación

¿Qué formatos? ¿qué versiones? ¿qué material adicional incluir?

¿Qué atributos se quieren preservar?

- datos, funcionalidad
- apariencia, esencia

La decisión dependerá de la misión institucional, las necesidades de la comunidad de usuarios, la capacidad técnica/ tecnológica institucional y los recursos disponibles.

Un avance: estándares

- 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.

Preservación de contenido

Hay una muy importante necesidad de **preservar el contenido digital** en el tiempo, con el objetivo de conservarlo accesible frente a riesgos como

Incendios, Inundaciones, etc

Robos

Problemas de hardware (rotura de discos, etc.)

Cambios tecnológicos constantes

Es un proceso continuo

Además de lo técnico, los esfuerzos de preservación incluyen retos legales, económicos e institucionales.

Preservación de contenido

Digital obsolescence

Es el resultado de la evolución de las tecnologías: a medida que surgen nuevas tecnologías, las viejas van quedando en desuso y se vuelven obsoletas.



Mantener tecnologías obsoletas en funcionamiento puede ser justificado en casos particulares, pero no en la mayoría.

Cornell University Library creó la "Cámara de los horrores"
<http://dpworkshop.org/dpm-eng/oldmedia/chamber.html>

Preservación de contenido

Digital obsolescence

Mantener tecnologías obsoletas requiere conservar

- Hardware
- Software (aplicaciones, librerías, sistema operativo, etc)
- Documentación (manuales, instructivos, etc)
- Personal con la capacitación y las habilidades necesarias para trabajar en ese entorno obsoleto

Suelen ser opciones muy difíciles de mantener y muy costosas.

En general no suele ser la mejor opción

Preservación de contenido

Estrategias

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

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

Preservación de contenido

Migración continua

Migrar la información de una tecnología a la siguiente de forma continua, evitando así la obsolescencia.

- Es una de las opciones de mayor uso
- Asegura el acceso en todo momento (los datos son siempre accesibles mediante una tecnología actual)
- Requiere transformación de los datos originales
- Decisiones sobre qué se desea preservar

Preservación de contenido

Adhesión a estándares internacionales

Es una estrategia que busca apoyarse en la afirmación de que los estándares internacionales son relativamente estables en el tiempo.

- En la actualidad, los estándares evolucionan casi tan rápido como las tecnologías
- Es una estrategia que debería usarse en combinación con otras
- Según la National Initiative for Networked Cultural Heritage, los formatos que no serán declarados obsoletos (al menos en un futuro cercano) son: TIFF y PDF sin compresión, y ASCII y RTF sin compresión, para imágenes y texto respectivamente.

Preservación de contenido

Emulación

Se trata de imitar las características y capacidades de un software y/o hardware, de modo que los procesos "crean" que están funcionando en la plataforma original.

- No hay necesidad de modificar los datos originales (como en la migración), manteniendo la integridad de la información.
- Una vez que se archivaron los datos, solo hay que asegurarse que el soporte físico utilizado siga siendo accesible.
- Se puede usar un mismo emulador para múltiples objetos del mismo tipo.

Preservación de contenido

Encapsulamiento

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

Como elementos a encapsular podemos tener:

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

Preservación de contenido

Metadatos de preservación

Generalmente considerados como metadatos administrativos

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

Varias iniciativas:

- PREMIS: PREservation Metadata: Implementation Strategies
- OAIS: Open Archival Information System
- NEDLIB: Networked European Deposit Library

Preservación de contenido

Políticas de backup

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

- Incendios
- Inundaciones
- Robos
- Fallas de hardware

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

- Incremental
- Espejo

Factores que gravitan para elegir formatos

1. **Divulgación.** Grado en que existen las especificaciones completas y herramientas para validar la integridad técnica y son accesibles para aquellos que crean y mantienen los contenidos digitales. Hay diferentes niveles de divulgación, lo más importante no es la aprobación por un organismo de normalización reconocido, pero sí la existencia de la documentación completa.
2. **Aprobación.** Grado en que el formato ya es utilizada por los principales creadores, difusores, o usuarios de los recursos de información.
3. **Transparencia.** Grado en que la representación digital está abierta al análisis directo con herramientas básicas.
4. **Auto-documentación.** Que permita agregar metadatos administrativos descriptivos, técnicos y otros porque esto es útil para la preservación.
5. **Dependencias externas.** Grado en que un formato en particular depende de hardware en particular, el sistema operativo o el software para la representación.
6. **Impacto de Patentes.** Grado en que la capacidad de las instituciones archivísticas para sostener contenido en un formato que puede ser inhibido por patentes.
7. **Mecanismos de protección técnica.** Implementación de mecanismos como el cifrado que impiden la conservación de los contenidos.

Selección de formatos: generalidades

Para asegurar la preservación de la información, el formato elegido debe ser legible por una aplicación durante el mayor tiempo posible. Esto implica evitar formatos propietarios cerrados, como documentos en Word (en cualquiera de las versiones). Lo recomendable es utilizar formatos propietarios pero abiertos, como el formato [TIFF](#), o mejor aún, formatos no propietarios como el ASCII. La razón es evidente: los documentos Word, por ejemplo, es necesario migrarlos a otros formatos Word más modernos no pudiéndose garantizar que estas migraciones no ocasionen pérdidas de información o cambios en la presentación de los documentos. Además existe otra razón importante para evitar el uso de formatos propietarios ¿Quién puede asegurar la supervivencia en un futuro de empresas por importantes que éstas sean en la actualidad?

Selección de formatos: generalidades

La utilización de un formato de codificación simple y universal como [XML](#) permite perpetuar los documentos electrónicos. XML es el formato ideal ya que además de ser un formato no propietario, y por tanto ofrecer garantía de preservación de la información (ASCII), permite estructurar la información y el intercambio de información a todos los medios.

Selección de formatos : generalidades

Para asegurar la integridad de los documentos que contienen objetos electrónicos (imágenes, sonidos, modelos, fórmulas, hiperenlaces..) se debe emplear la misma filosofía que con la información textual. Los formatos imagen considerados mejores para la conservación son el [TIFF \(Tagged Image File Format\)](#) que su compresión no experimenta ninguna pérdida de calidad, el [PNG \(Portable Network Graphics\)](#), cuya compresión experimenta apenas pérdidas en la resolución y además es muy ligero y el [JPEG](#).

Selección de formatos : generalidades

En cuanto a los Formatos mixtos los mejores son el [Postscript](#), que puede ser enviado a cualquier periférico que soporte este lenguaje, sin tener en cuenta su resolución, produciendo un resultado adaptado a cada tipo de periférico y el [PDF \(Portable Document Format\)](#), basado en el Postscript, propietario pero abierto de la casa Adobe y que facilita un programa gratuito para poder leer este tipo de documentos. Para la preservación, se recomienda especialmente el [PDF/A](#)

Sobre PDF/A

- PDF/A es un estándar para codificar documentos en un formato “impreso”, que es portable entre sistemas y ampliamente usado para distribución y archivado de documentos. Sin embargo, la pertinencia de un archivo PDF para preservación depende de las opciones elegidas cuando el PDF fue creado: en particular, si se embebieron las fuentes necesarias para renderizar el documento, si se usa encriptación y si se preserva información adicional del documento original, más allá de lo que se precisa para imprimirlo.

Sobre PDF/A

- El estándar PDF/A no define una estrategia de archivado o los objetivos de un sistema de archivado. Sí identifica un “perfil” para documentos electrónicos que asegura que los documentos pueden ser reproducidos exactamente de la misma manera durante años. Un elemento clave para esta reproductibilidad es que los documentos PDF/A deben ser 100% auto-contenidos: esto significa que toda la información necesaria para mostrar el documento de la misma manera cada vez, debe embeberse dentro del archivo. Esto incluye (pero no se limita a) todo el contenido (texto, imágenes rasterizadas, gráficos vectorizados), fuentes, información de color, etc. Un documento PDF/A no puede jamás depender de información de fuentes externas (por ejemplo, programas fuente o *streams* de datos), aunque se permite que tengan anotaciones (como hipertextos) que enlacen a documentos externos.

Otros elementos de la compatibilidad con PDF/A

- El contenido de audio y video está prohibido.
- Java script y enlaces a archivos ejecutables están prohibidos.
- Todas las fuentes deben estar embebidas, y también deben ser legalmente embebibles para renderización ilimitada y universal. Esto significa para un usuario poder abrir el documento y que los caracteres se muestren de manera correcta (de aquí a X años) aunque no tenga esa tipografía en su computadora.
- Los espacios de colores deben ser especificados de una manera independiente del dispositivo.
- Se prohíbe la encriptación.
- El uso de metadatos basados en estándares se mantiene.

Otros elementos de la compatibilidad con PDF/A

- Las referencias a contenidos externos están prohibidas.
- La compresión de imágenes LZW y JPEG2000 están prohibidas en PDF/A1, pero JPEG 2000 se permite en PDF/A2.
- Capas y objetos transparentes están prohibidos en PDF/A1 pero no en PDF/A2.
- Firmas digitales provisionales se permiten en PDF/A2.
- Los archivos embebidos están prohibidos en PDF/A1, pero PDF/A2 permite embeber archivos PDF/A.
- PDF/A3 permite embeber cualquier formato como XML, CSV, CAD, archivos de Word, planilla de cálculo, otros PDF/A, etc. como objetos archivados completos.

Niveles de cumplimiento

- PDF/A1 posee dos niveles de cumplimiento:
- **PDF/A-1a** aplica corrección semántica y estructura. Cada carácter debe tener su equivalente Unicode. La estructura se expresa por medio de etiquetas.
- **PDF/A-1b** aplica integridad visual.

National Archives of Australia

The screenshot displays the National Archives of Australia website. At the top left, there is the Australian Government logo and the National Archives of Australia logo. The tagline "Your story, our history" is prominently displayed. A navigation menu includes "Home", "The collection", "Records management", "Visit us", and "About us". Below the menu, a breadcrumb trail reads: "Home > About us > Our organisation > Accountability and reporting > Operations and Preservation policies".

The main content area is titled "Digital Preservation Policy" and includes the following information:

- Preserving Archival Digital Records Transferred from Commonwealth Agencies**
- July 2009
- Version**
1,2, July 2011

A sidebar on the left lists various organizational sections, with "Accountability and reporting" selected. The "Contents" section lists 11 numbered items:

1. Policy statement
2. What is digital preservation?
3. Objectives of digital preservation at the National Archives
4. Challenges to the preservation of digital records
5. Scope
6. Digital preservation principles
7. Digital preservation approach
8. Preservation formats
9. Records which cannot be converted to an open format
10. Essential characteristics of digital records
11. Cooperation / collaboration

Preservation file formats

Australia

The digital preservation file formats we have selected are listed in the following table.

The preferred preservation file formats are the targets for file format conversions. Both the preferred and acceptable preservation file formats are based on open standards and any digital records received in these formats are preserved as they are.

Format Category	Preferred Preservation open file format ¹ (preserved as they are)	Files in these formats are converted to our preferred preservation open file format	Acceptable open file format ² (Not converted, preserved as they are)
Archive	An index of the contents is created as XML. The content of the archive is converted according to the appropriate preservation file format.	<ul style="list-style-type: none"> • Compressed archives (gzip, bzip2, war, zip) • Uncompressed archives (jar, tar, zip) 	<ul style="list-style-type: none"> • Drawing Exchange Format (dxf)
Audio	Free Lossless Audio Codec (flac).	<ul style="list-style-type: none"> • Audio Interchange File Format (aiff) • Broadcast Wave File (bwf) • MPEG-2 audio layer 3 (mp3) • Speex (spx) • Vorbis (ogg, oga) 	
Computer aided design	Not yet decided.	<ul style="list-style-type: none"> • Wave Audio File (wav) • Drawing (dws, dwt, dwg) • Design Web Format (dwt) 	
Email	XML and XSL files are created for each email. Any attachments are converted according to the appropriate preservation file format.	<ul style="list-style-type: none"> • Mailbox (mbx, mbox) • Outlook Mail Message (msg) • Outlook Personal Information Store (pst) 	<ul style="list-style-type: none"> • Email (eml)

1 Preferred preservation file formats are based on open standards and are the targets for file format conversions.

2 Acceptable preservation file formats are those based on open standards. Any digital records received in these formats are preserved as they are.

Format Category	Preferred Preservation open file format ¹ (preserved as they are)	Files in these formats are converted to our preferred preservation open file format	Acceptable open file format ² (Not converted, preserved as they are)
Geospatial data	Not yet decided.	<ul style="list-style-type: none"> • Spatial Data File (sdf) 	<ul style="list-style-type: none"> • Geography Markup Language (gml)
Image - raster	Portable Network Graphics (png).	<ul style="list-style-type: none"> • Bitmap (bmp, gif, pcx, pnm, ras, xbm) • Photoshop (psd) • Tagged Image File Format (tiff) • Windows Cursor (cur) 	<ul style="list-style-type: none"> • Open Document Graphics (odg) • Joint Photographic Experts Group (jpeg) • Portable Document Format (pdf)
Image - vector	Not yet implemented.	<ul style="list-style-type: none"> • Adobe Illustrator (ai) • Encapsulated PostScript (eps) 	<ul style="list-style-type: none"> • Scalable Vector Graphics (svg)
Office documents	Open Document Format (odf).	<ul style="list-style-type: none"> • Excel (xls, xlsx, xlt) • PowerPoint (pot, pps, ppt, pptx) • Rich Text Format (rtf) • Symbolic Link (slk) • StarOffice (sdd, sdc, sdw, sxc, sxi, sxw) • Word (doc, docx, dot) • Word Perfect (wpd) 	<ul style="list-style-type: none"> • Open Document XML (fodt) • OpenOffice.org XML (stw, stc, std, sti, sxg, sxm)
Project	XML	<ul style="list-style-type: none"> • Project (mpp) 	

Format Category	Preferred Preservation open file format ¹ (preserved as they are)	Files in these formats are converted to our preferred preservation open file format	Acceptable open file format ² (Not converted, preserved as they are)
Plain text	Plain text in Unicode or ASCII.		<ul style="list-style-type: none"> • Style sheets (css, xsl/xslt) • Database tables, such as comma and tab-separated files (csv, tsv) • Scripting files (such as Python, Javascript, Perl, PHP) • Structured Query Language (SQL)
Video – video stream	<p>Not yet decided.</p> <p>Video files are currently preserved as is.</p>	<ul style="list-style-type: none"> • Flash Movie file (swf) • Motion JPEG/JPEG 2000 • RealVideo (rv, rmvb) • Windows Media Video (wmv) 	<ul style="list-style-type: none"> • Theora (ogg, ogv) • "Raw" video

Format Category	Preferred Preservation open file format ¹ (preserved as they are)	Files in these formats are converted to our preferred preservation open file format	Acceptable open file format ² (Not converted, preserved as they are)
Video - container	Not yet decided.	<ul style="list-style-type: none"> • Audio Video Interleave (avi) • Advanced Systems Format (asf) • Flash Video File (flv) • MPEG video (mpeg-2, mpeg-4) • QuickTime Movie (mov) • RealMedia (rm) 	<ul style="list-style-type: none"> • Ogg (ogv) • Matroska (mkv)
Video – audio stream	Free Lossless Audio Codec (flac).	<ul style="list-style-type: none"> • Advanced Audio Coding (aac) • MPEG-2 audio layer 3 (mp3) • RealMedia audio (ra, ram) 	
Website	XHTML	<ul style="list-style-type: none"> • HTML (htm, html) • Active Server Page (asp, aspx) 	
Website archive	Web ARChive (warc).	<ul style="list-style-type: none"> • MIME HTML (mht) 	<ul style="list-style-type: none"> • ARC file format (arc)

Preservation Formats for Electronic Records Smithsonian

In accordance with best practices, SI Archives prefers to preserve transferred electronic records in the formats described in the table below. This table lists the original/creating application by its native format(s), the corresponding primary preservation format (preferred) and the secondary format. The secondary format will be used only when a record of sufficient quality cannot be created in the primary preservation format. A glossary of acronyms used is appended to the end of this document.

Where contributing SI units have prepared their electronic records in a preservation format, SIA will preserve those records in the received format unless extenuating circumstances apply. Records originally transferred to SIA in the secondary preservation format may be later converted to the appropriate primary preservation format by SIA staff as part of the long term maintenance of the record.

This document does not address related considerations and procedures required in the conversion from original formats to preservation formats. It is essential that individuals responsible for these activities refer to forthcoming format-specific Electronic Record publications as appropriate.

Original/ Creating Application	Primary Preservation Format	Secondary Preservation Format	TRM References
<i>(various)</i> .txt	ASCII (keep original extension)		
<i>Corel WordPerfect</i> .wpd, .wpx, .doc, .rtf	PDF 4.0 or higher	RTF (text)	Pg 4-202 (PDF) Pg 4-93 (RTF)
<i>Microsoft Word</i> .doc, .dot, .rtf	PDF 4.0 or higher	RTF (text)	Pg 4-202 (PDF) Pg 4-93 (RTF)
<i>Lotus 1-2-3</i>	<i>Same as original</i>	CSV	No reference
<i>Microsoft Excel</i> *.xls	PDF	CSV	Pg 4-202 (PDF)

Original/ Creating Application	Primary Preservation Format	Secondary Preservation Format	TRM References
<i>Microsoft Powerpoint</i> *.ppt	PDF		Pg 4-202 (PDF)
<i>Graphics Applications</i>			
<i>(various graphics applications)</i> *.bmp, *.tif, *.jpg, *.gif	TIFF (24-bit RGB) (minimum 3000 pixels along long dimension or 400 dpi, whichever is greater)	JPEG (uncompressed) (for JPG and GIF formats only.)	Pg 4-98 (TIFF) Pg 4-101 (JPEG)
<i>Adobe Photoshop, Illustrator</i> .psd, .eps	TIFF (24-bit RGB) (minimum 3000 pixels along long dimension or 400 dpi, whichever is greater)		Pg 4-98 (TIFF)
<i>Macromedia Fireworks/ Freehand</i> .png	TIFF (24-bit RGB) (minimum 3000 pixels along long dimension or 400 dpi, whichever is greater)	TIFF (8-bit, grayscale or b/w for black and white images only) (minimum 3000 pixels along long dimension or 400 dpi, whichever is greater)	Pg 4-98 (TIFF)

Original/ Creating Application	Primary Preservation Format	Secondary Preservation Format	TRM References
<i>Microsoft Image Composer</i> .mic	TIFF (24-bit RGB) (minimum 3000 pixels along long dimension or 400 dpi, whichever is greater)	TIFF (8-bit, grayscale or b/w for black and white images only) (minimum 3000 pixels along long dimension or 400 dpi, whichever is greater)	Pg 4-98 (TIFF)
<i>Database Management Systems (DBMS)</i>			
<i>Dbase III+</i> .dbf, .dbt	<i>Same as original</i>	CSV	CSV not listed Dbase not listed
<i>Lotus Notes (database)</i>	<i>Same as original</i>	CSV	CSV not listed Lotus Notes not listed
<i>Microsoft Access (database)</i> *.mdb	<i>Same as original</i>	CSV	Pg 4-42, 4-74, 10-7
<i>Microsoft SQL Server</i> *.bak	<i>Same as original, full backup made with MS SQL</i>	CSV	Pg 10-10 (MS SQL Server)
<i>Sybase SQL Server, Adaptive Enterprise Server, SQL Anywhere</i>	<i>Same as original, full backup made with Sybase dbms. Also requires full backup of Sybase master database.</i>	CSV	Pg 4-74 (Sybase)

Original/ Creating Application	Primary Preservation Format	Secondary Preservation Format	TRM References
<i>WebPage Developers and HTML Editors</i>			
<i>Adobe PageMill, Microsoft Frontpage, Macromedia Dreamweaver (other various)</i> .htm, .html, .shtml	XHTML 1.0	PDF (Isolated, single webpage only)	Pg 4-86 (XHTML) Pg 4-89
<i>Active Server Page(various)</i> .asp	XHTML 1.0	ASP (depending on parameter complexity)	Pg 4-86 (XHTML) Pg 4-25 (ASP)
<i>Macromedia Cold Fusion</i> .cfm	XHTML 1.0		Pg 4-86 (XHTML) Pg 4-33 (CFM)
<i>Macromedia Flash/Shockwave</i> .swf	SWF (No change)	<i>Analyze carefully for embedded URLs or other file dependencies.</i>	Pg 4-27, 4-30, 4-105, 10-15, 10-23
<i>Real Media</i> .rm, .ram	<i>Same as original</i>		Pg 4-27, 10-6
<i>Architectural Design/Engineering Graphics</i>			
<i>AutoCAD</i> *.cad	PDF (v6.0 with layer retention)		Pg 4-202 (PDF) Pg 10-15 (CAD)
<i>Microsoft Visio</i> *.vsd	PDF (v6.0 with layer retention)		Pg 4-202 (PDF) Pg 10-16 (Visio)



Recommended File Formats

Recommended File Formats for Preservation

The UT Digital Repository accepts work in any digital format; however, submission in a file format listed below is strongly encouraged. The formats listed below are considered relatively stable and therefore facilitate long-term preservation efforts. These formats exhibit all or many of the following characteristics: open documentation; support across a range of software platforms; wide adoption; no compression (or lossless compression); no embedded files or embedded programs/scripts; and non-proprietary format. For additional information on the UT Digital Repository's preservation strategy, please see the [Preservation Policy](#).

Textual Formats

Acrobat PDF/A
Comma-Separated Values
Open Office Formats
Plain Text (US-ASCII, UTF-8)
XML

File Extensions

.pdf
.csv
.odt, .ods, .odp
.txt
.xml

Image/Graphic Formats

JPEG
JPEG2000
PNG
SVG 1.1 (no Java binding)
TIFF

File Extensions

.jpg
.jp2
.png
.svg
.tif, .tiff

Audio Formats

AIFF
WAVE

File Extensions

.aif, .aiff
.wav

https://repositories.lib.utexas.edu/recommended_file_formats



Algunas recomendaciones en digitalización

Como guía en procesos de digitalización pueden usarse:

- "Technical Guidelines for Digitizing Cultural Heritage Materials" generado en 2010 por la Federal Agencies Digitization Guidelines Initiative (FADGI).
- "Directrices para proyectos de digitalización de colecciones y fondos de dominio público" , IFLA (2002).
- "Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files - Raster Images" , NARA (2004).
- “Recomendaciones para la digitalización de los documentos en archivos”. Junta de Castilla y León (2011).

Algunas recomendaciones en digitalización

- De cada obra digitalizada es recomendable tener al menos 3 versiones: a) imagen maestra resultante del escaneo, b) documento maestro con imagen derivada de la anterior y el OCR generado sobre ella y c) otra(s) derivada(s) a los fines de uso y distribución en la web.

Enlaces de interés

<http://www.digitizationguidelines.gov/>

[Directrices IFLA para digitalización](#)

[Recomendaciones para la digitalización de archivos. Junta de Castilla y León](#)

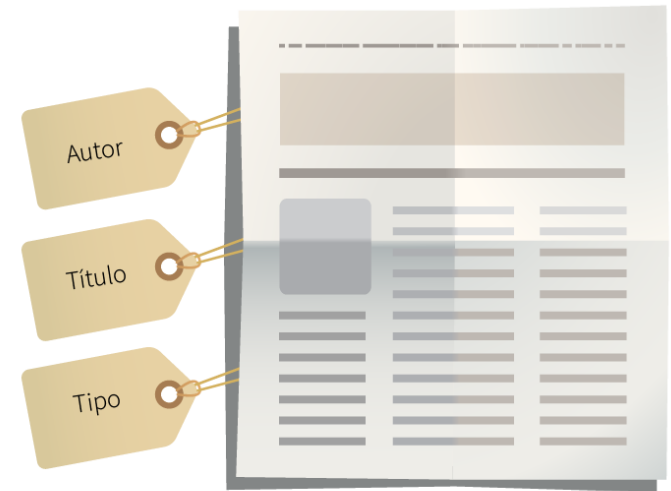


Formatos. ¿Cómo conocer lo que tiene un RI?

- 1. Perfilamiento automatizado de los objetos del repositorio:** esto involucra al objeto de contenido (CDO) con sus propiedades significativas y a la información de representación de ese objeto (RI). Realizar el perfil con DROID que contrasta con el registro PRONOM y brinda un reporte.
- El punto 1 es una de las 3 partes que se consideran importantes a la luz de cumplir con la ISO 14721 y realizar una evaluación del repositorio en los aspectos de preservación y accesibilidad

¿Qué acciones se proponen?

Nombre	Descripción	Formato	Ver	Orden
Bloque: TEXT				
<input type="checkbox"/> Tesina de Licen ... mazan Maria Belen.pdf.txt	Extracted text	Text	[Ver]	1 (Anterior:1)
<input type="checkbox"/> presentación.xps).pdf.txt	Extracted text	Text	[Ver]	2 (Anterior:2)
Bloque: ORIGINAL				
<input type="checkbox"/> Tesina de Licenciatura - Almazan Maria Belen.pdf (principal)	Documento completo	Adobe PDF	[Ver]	1 (Anterior:1)
<input type="checkbox"/> ...	Presentación	Adobe	[Ver]	2



Información descriptiva (DI)

De Giusti, Mrarisa R. (2014). Tesis doctoral: “UNA METODOLOGÍA DE EVALUACIÓN DE REPOSITARIOS DIGITALES PARA ASEGURAR LA PRESERVACIÓN EN EL TIEMPO Y EL ACCESO A LOS CONTENIDOS”. Disponible en: <http://hdl.handle.net/10915/43157>

Resumen PD en RIs

- Regulación de todos los procedimientos.
- Regulación de los derechos de preservación digital sobre los documentos.
- Regulación de los formatos admisibles.
- Control de formatos en la ingestión.
- Formatos de visualización y de preservación
- Almacenaje de metadatos técnicos.
- Copias sistemáticas externas.
- Creación de procedimientos de contingencia ante desastres.
- Auditoría interna/externa de seguridad.
- Plan de preservación...

Sección 3



El Modelo OAIS

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

ISO 14721: 2012

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.

ISO 14721

- El modelo *OAIS*, *totalmente aceptado internacionalmente* y estandarizado como norma ISO 14721, establece las seis funciones que han de configurar un sistema integral de preservación, así como los flujos de gestión de los archivos y sus metadatos.

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.

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 descrito 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.

Conceptos en OAIS

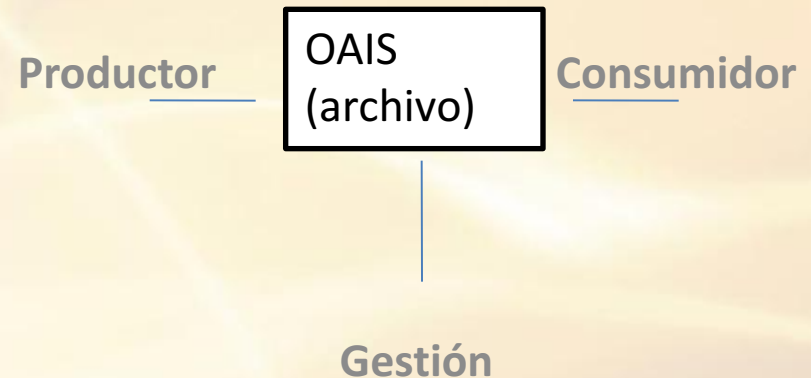
El propósito de esta parte de la norma es motivar y describir varios conceptos clave, de alto nivel del OAIS.

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**

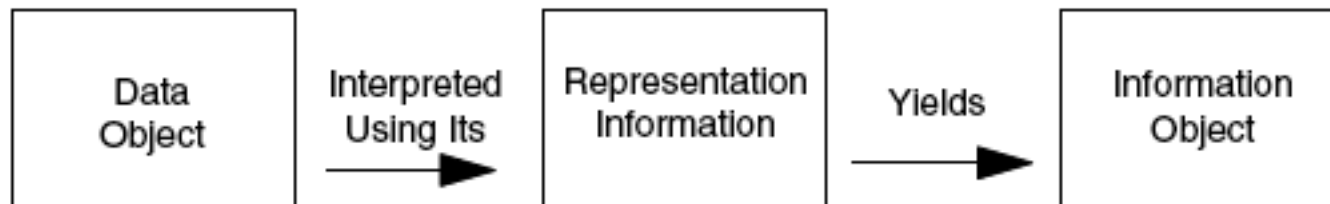


Conceptos en OAIS

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.



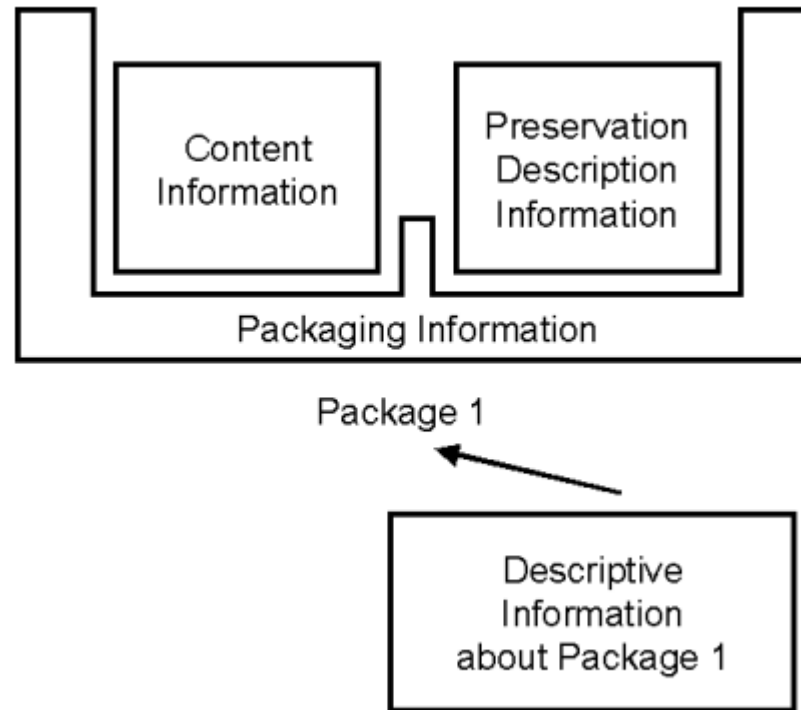
Conceptos en OAIS

- 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

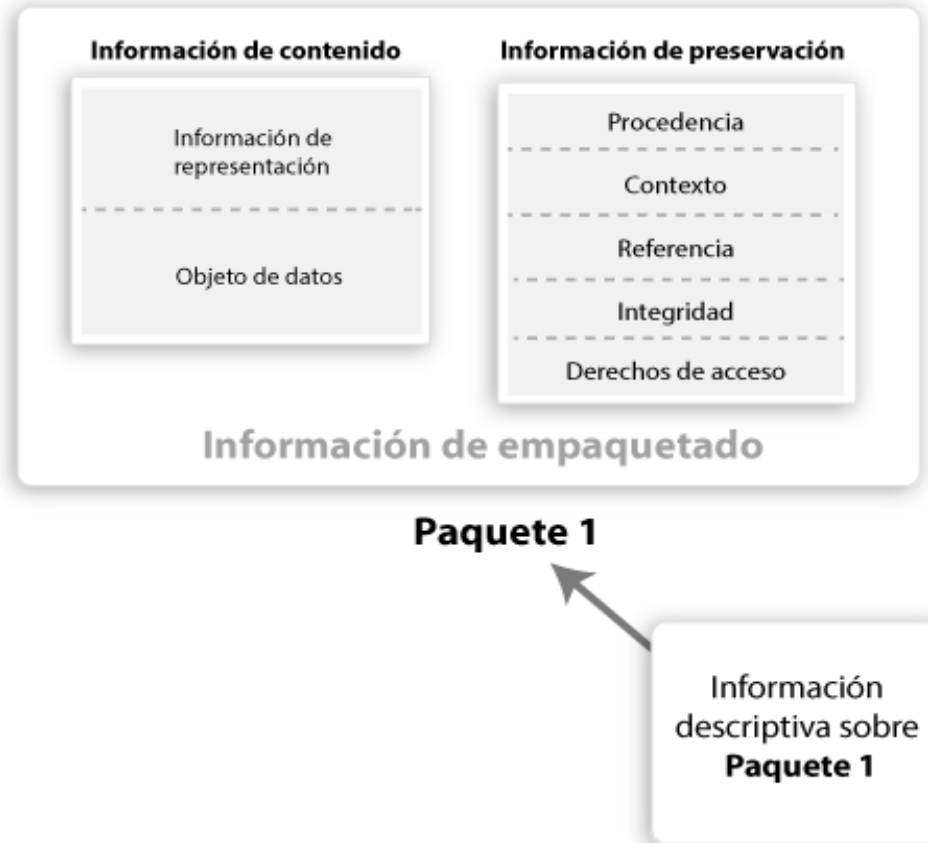


ISO 14721: Fig 2-3: Paquete de información: conceptos y relaciones

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



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**, la **integridad** y los **derechos de acceso**.

Elementos de la PDI

- 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 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.
 - Cuestiones de copyright, propiedad intelectual y restricciones legales para el uso.
 - Autoridad para modificar el modo de representación de la información.
 - Acuerdos con organizaciones externas.
3. Determinación de la comunidad designada de consumidores.

OAIS

Modelo Funcional Sección 4.1 de la norma



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

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.

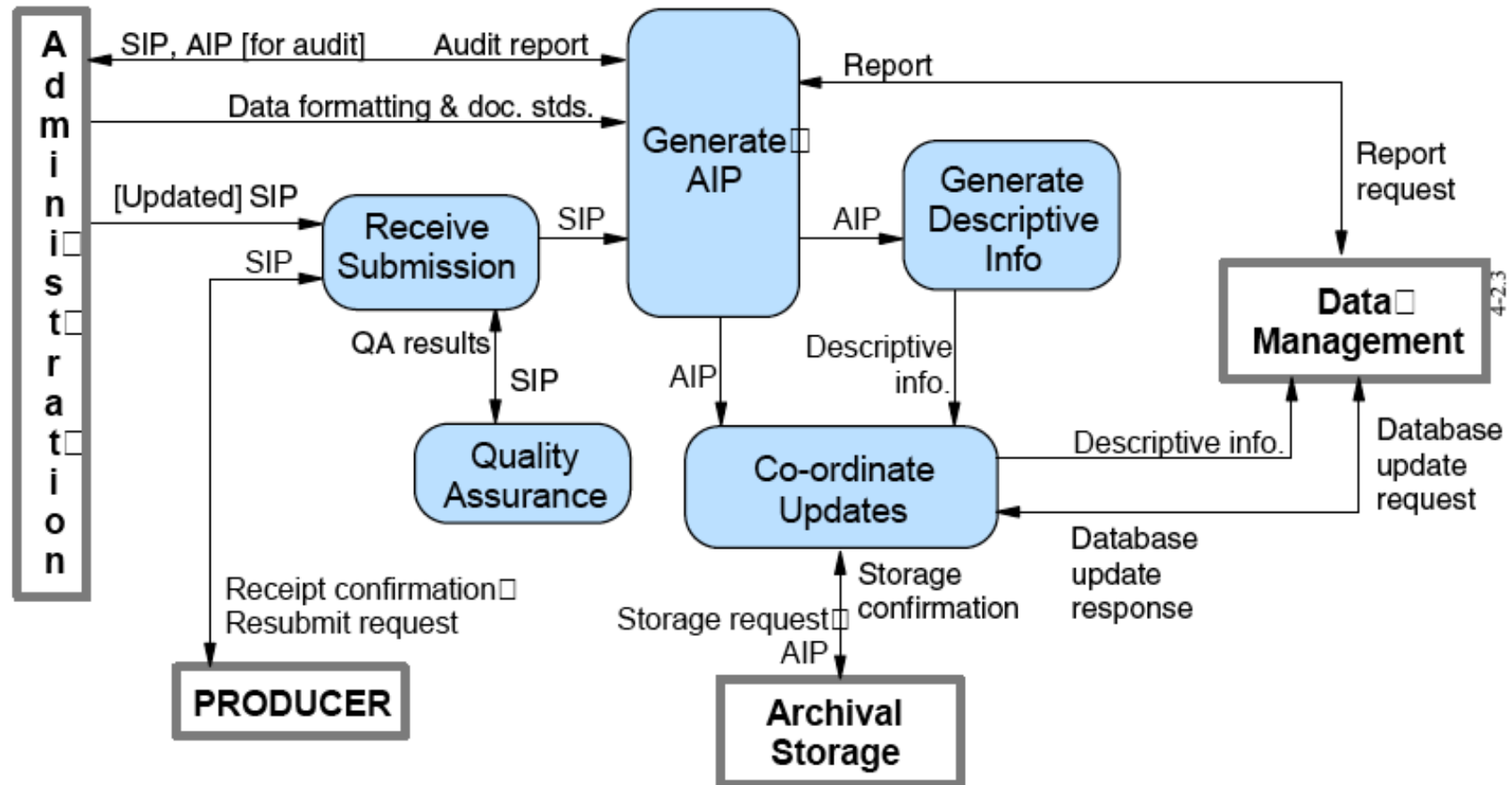
Modelo OAIS

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.

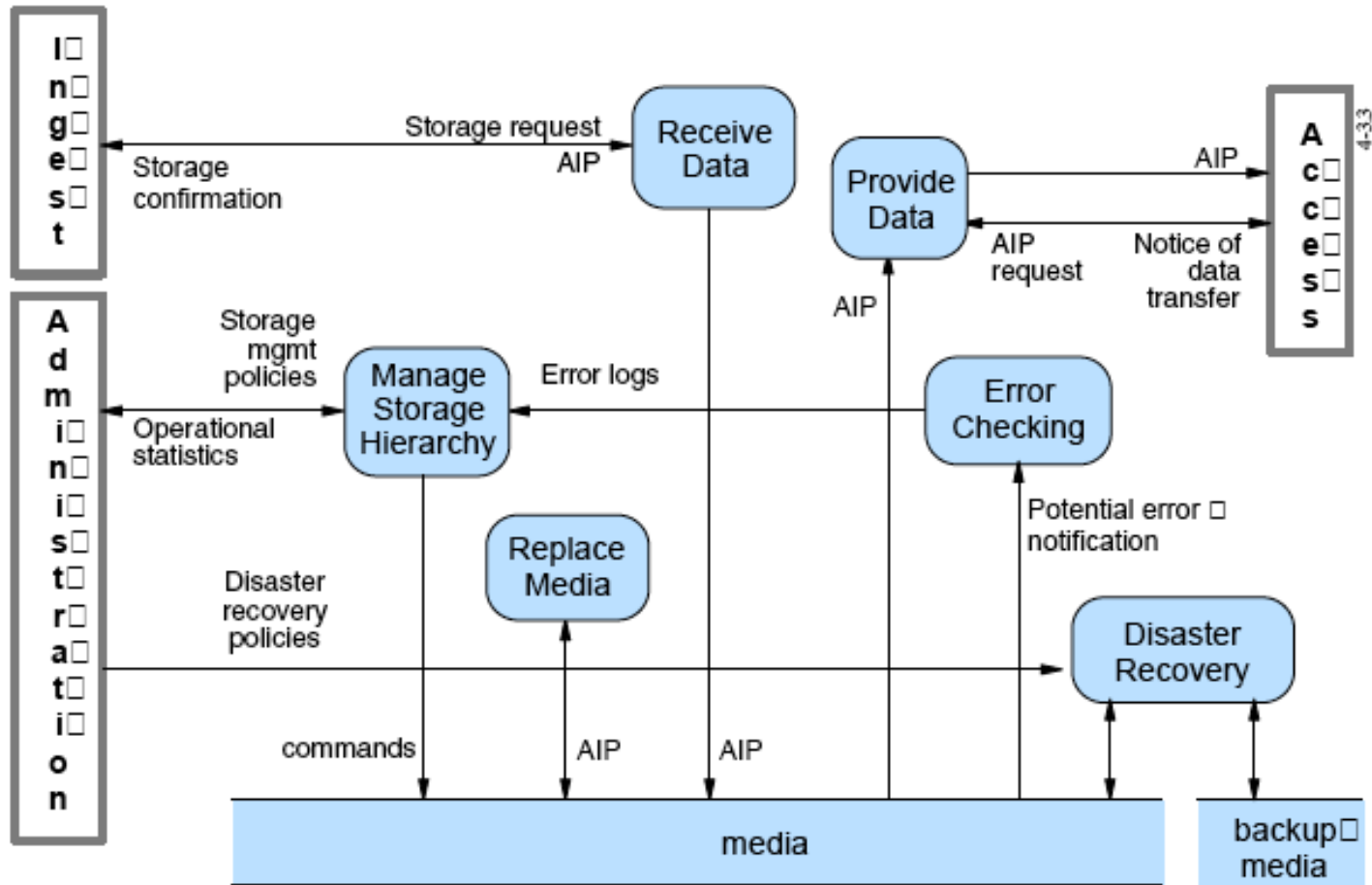
Functions of Ingest



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.

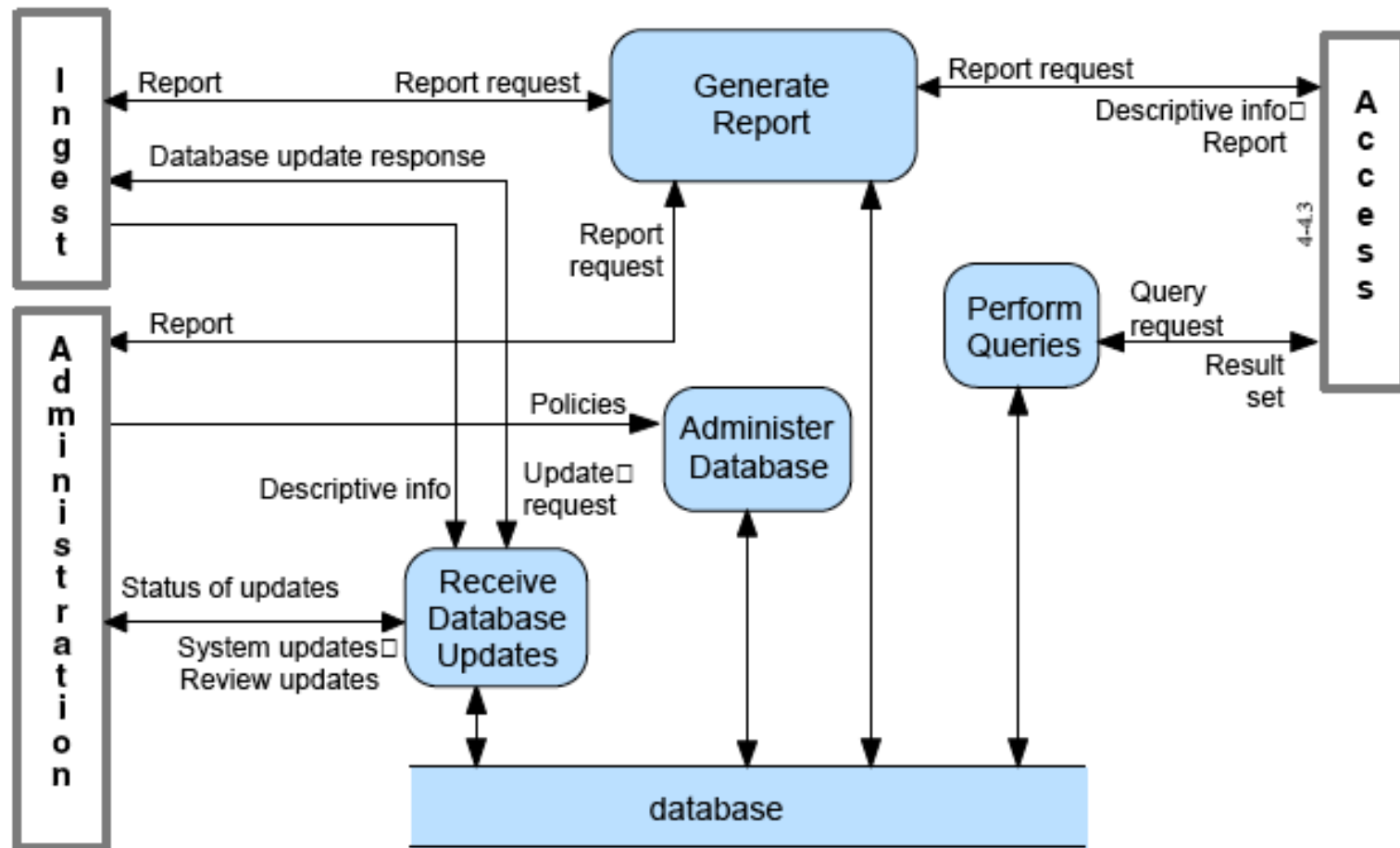
Functions of Archival Storage



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*.

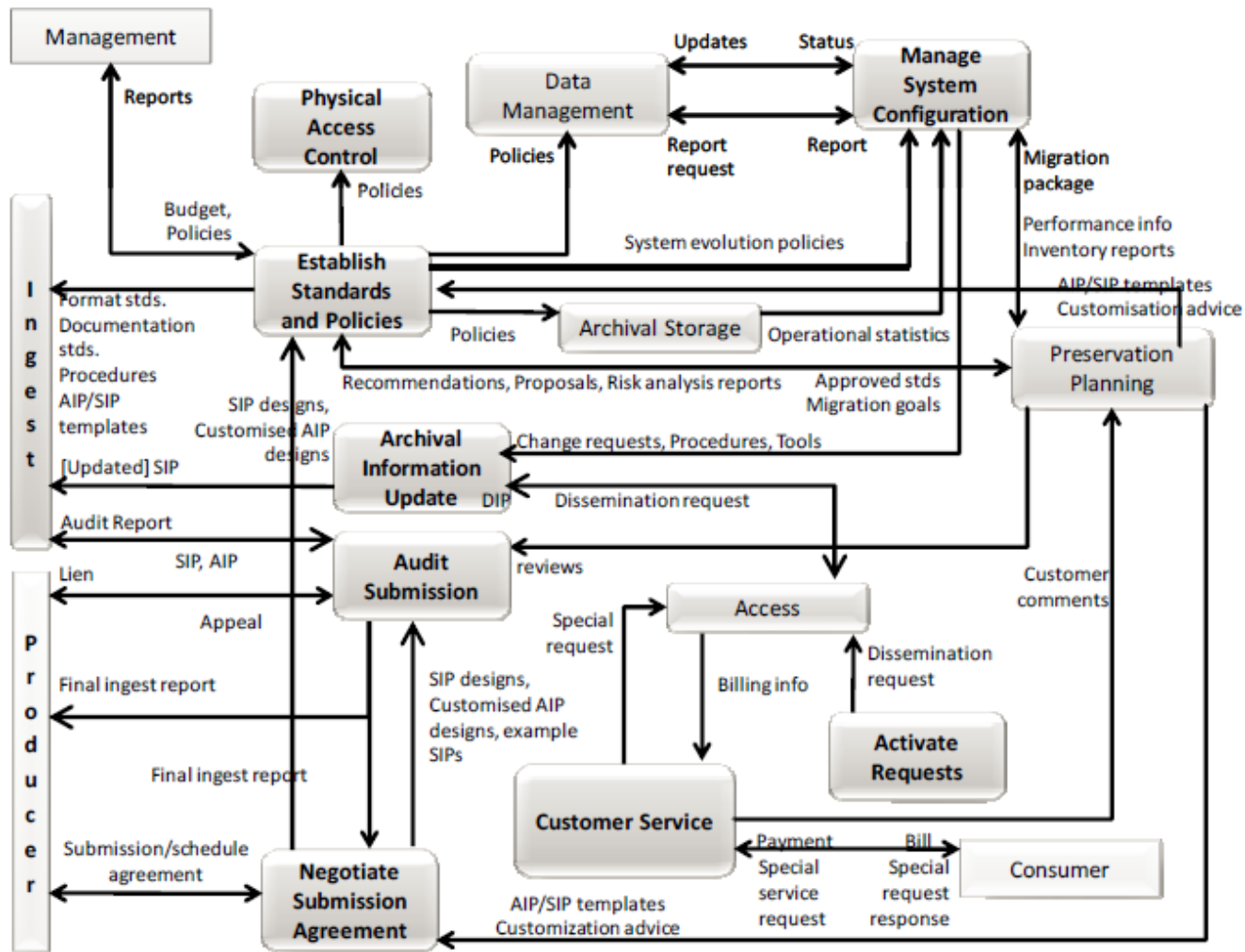
Functions of Data Management



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*.

Functions of 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.

Envía los pedidos 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.

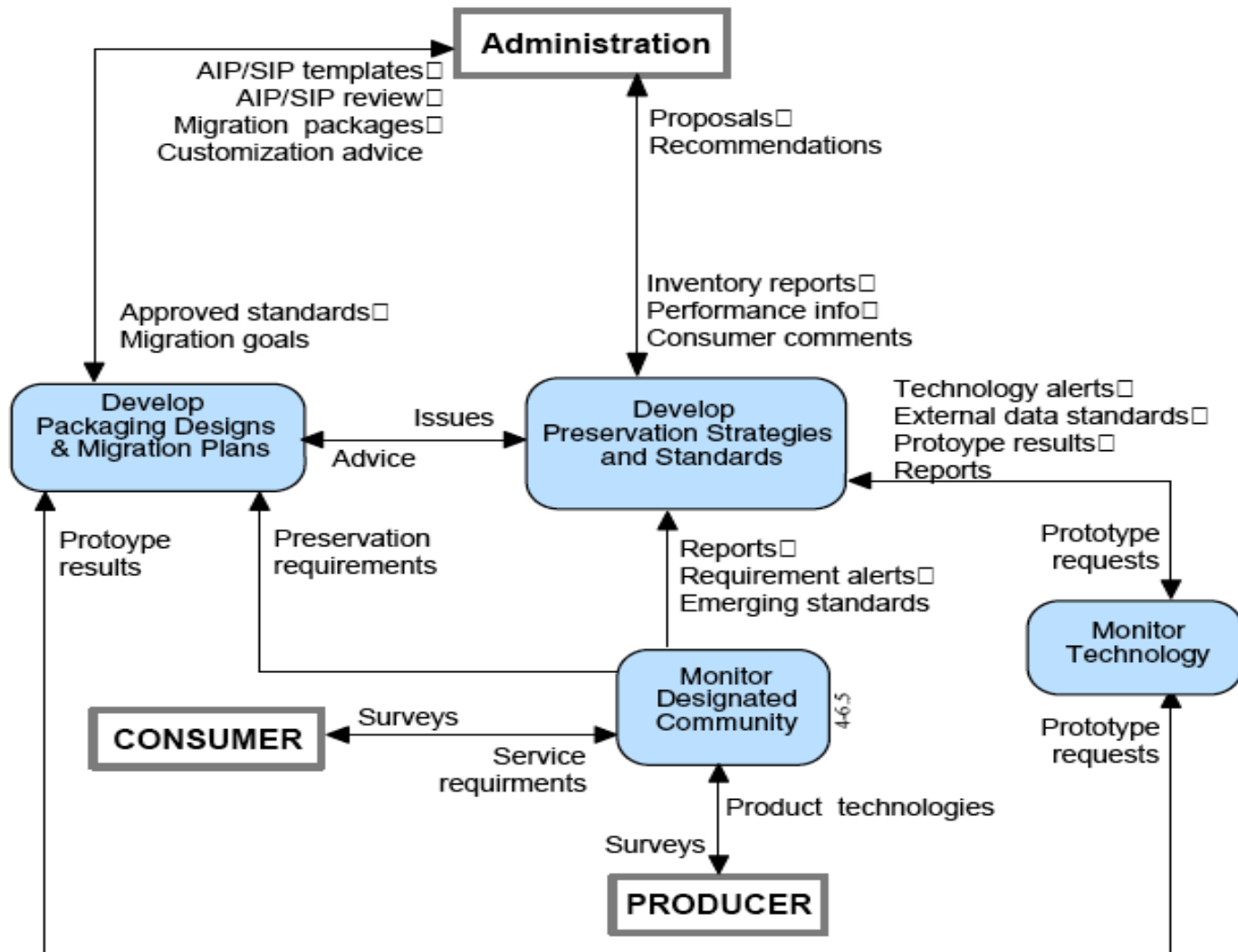
Entidad OAIS administration

Verifica que los AIP y SIP suministrados sigan las especificaciones. En el caso de SIP y de AIP verifica la comprensión por parte de la comunidad designada. Verifica que la Información de representación y la PDI sean adecuadas y comprensibles para la comunidad designada.

Mantiene un registro de de solicitudes y revisa periódicamente los contenidos del archivo para determinar si los datos están disponibles.

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

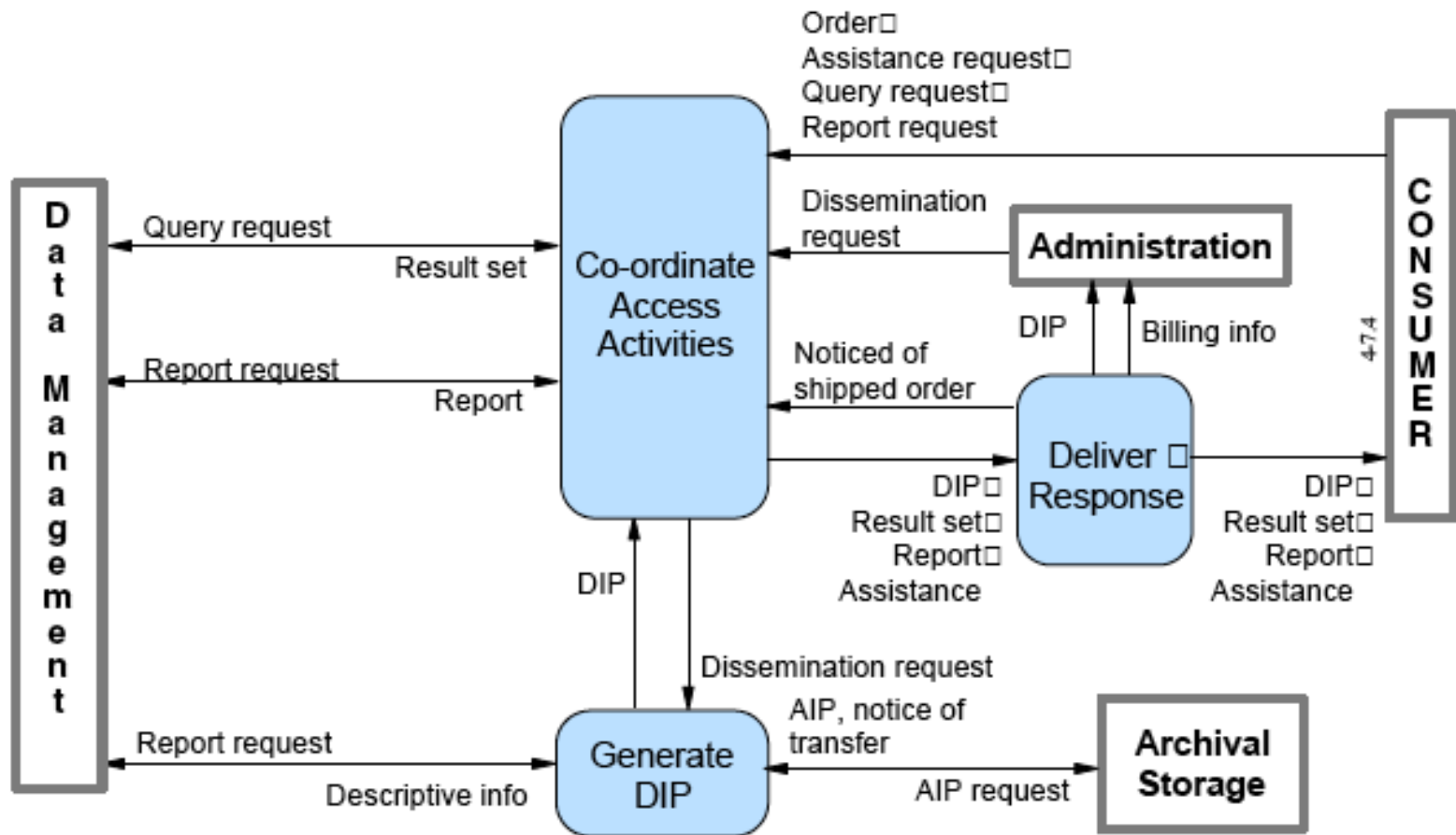
Functions of Preservation Planning



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.

Recibe las solicitudes de información.

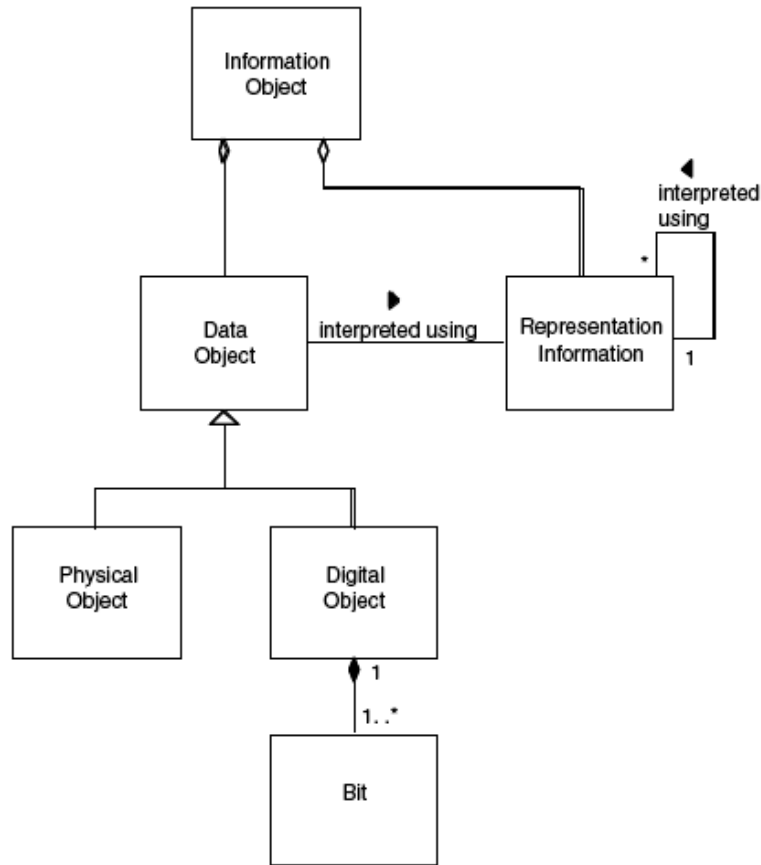
Acepta los requerimientos de los paquetes de diseminación recuperados de los AIP de la entidad *archival storage* y transmite un *report request* al *Data Management* generando un DIP.

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

OAIS

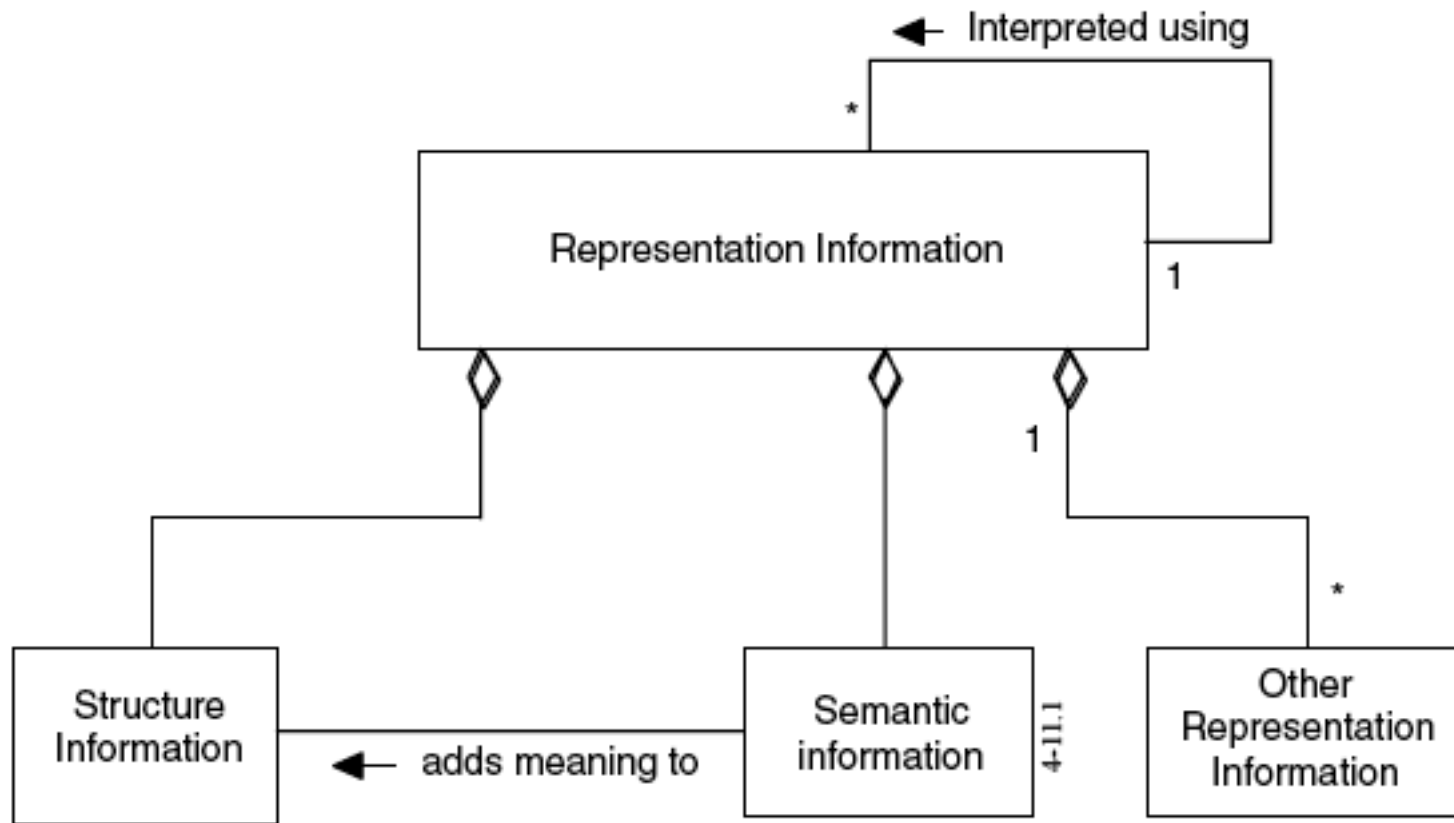
Modelo de Información
Sección 4.2 de la norma

OAIS Objeto de información

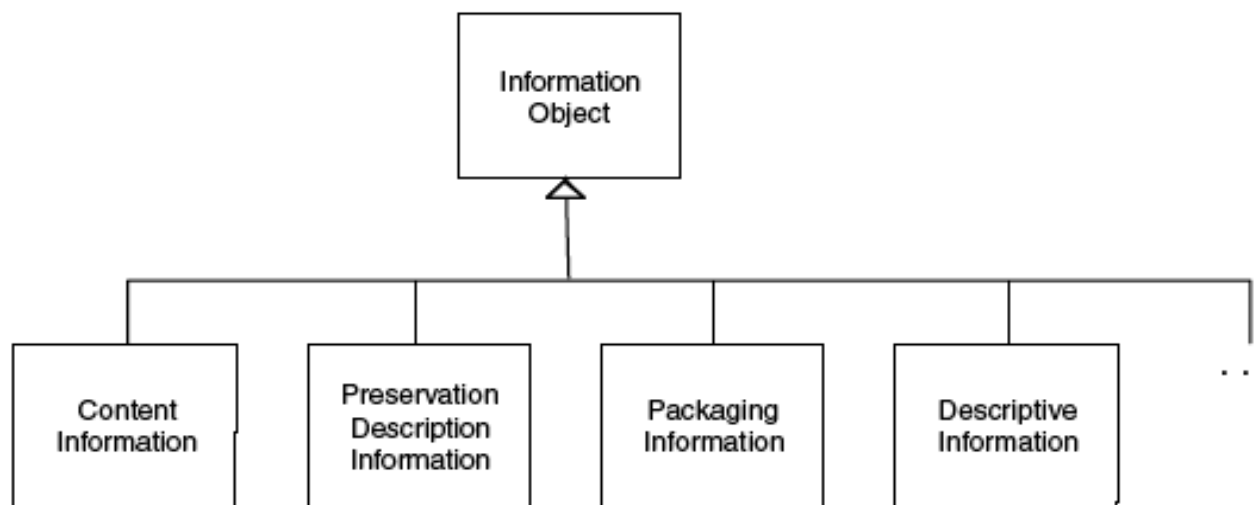


- El **Objeto de Información** está compuesto de un Objeto de Datos, que puede ser físico o digital, e Información de Representación que permite la interpretación completa de los datos.

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

Información descriptiva de preservación (PDI)

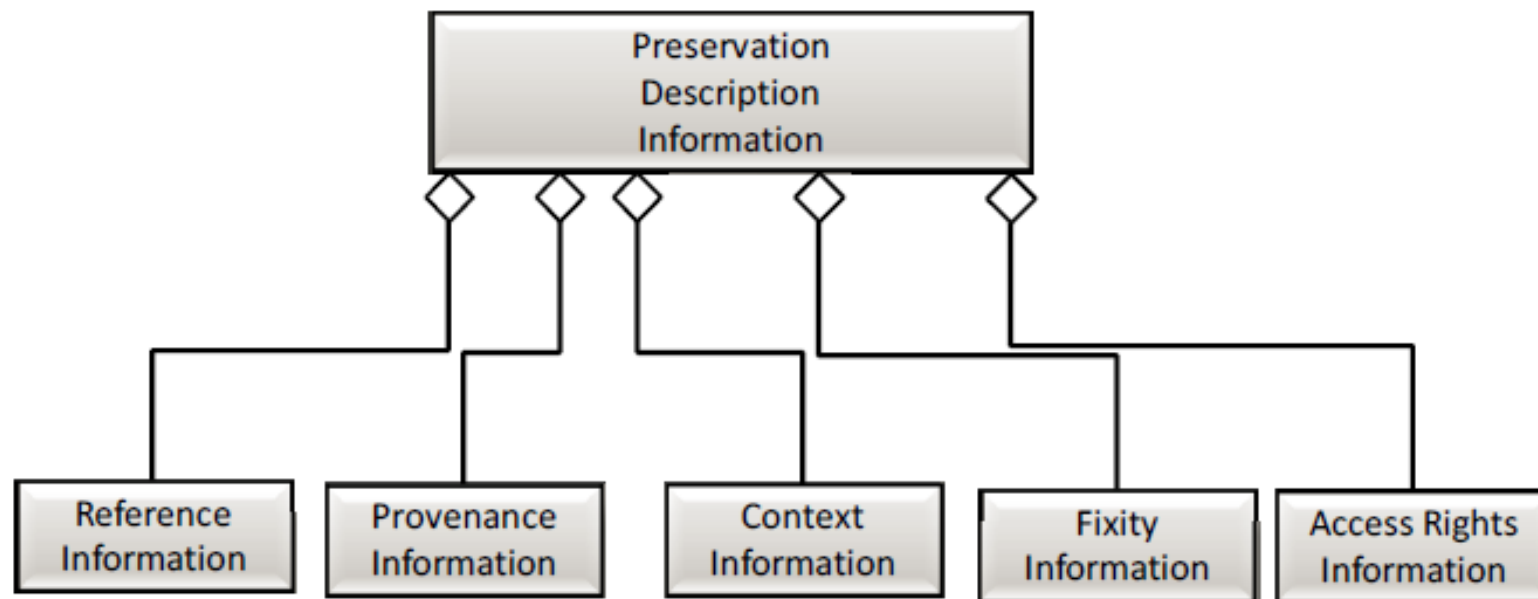


Figure 4-16: Preservation Description Information

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.

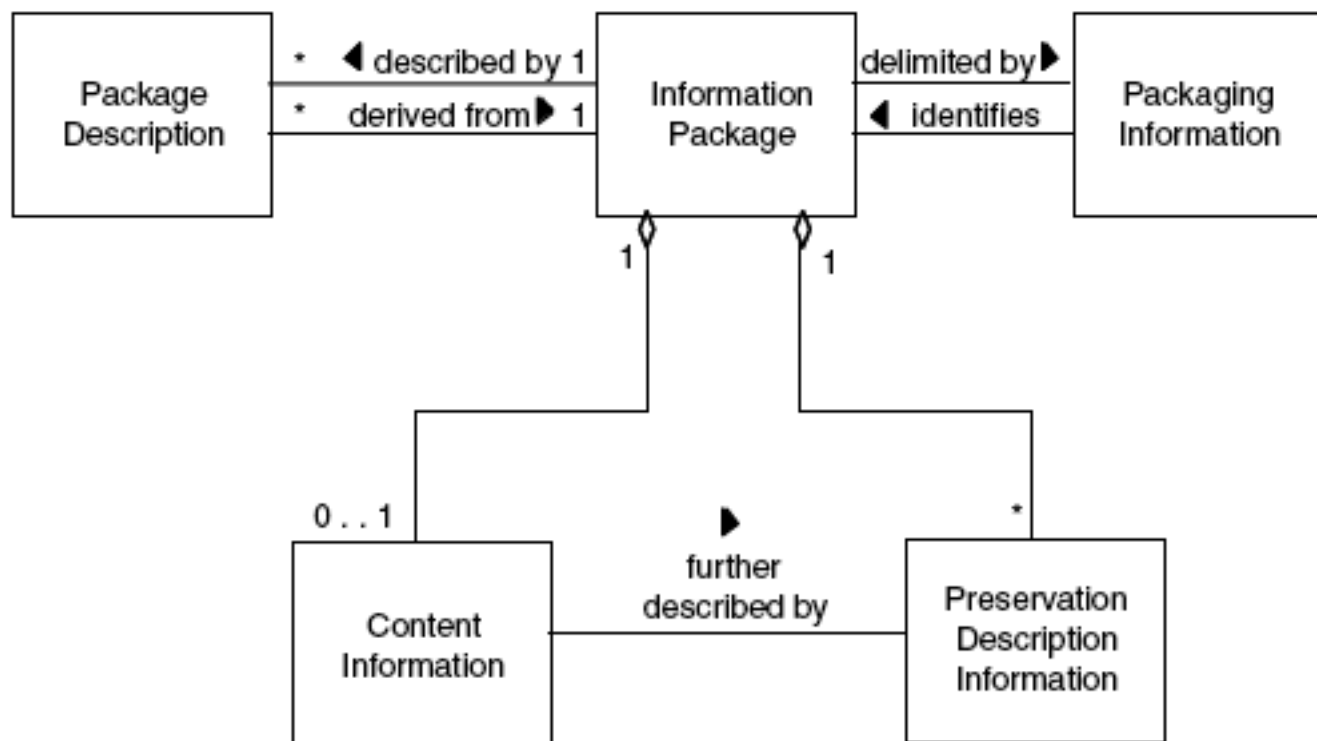
Información de la fijeza: proporciona los controles de integridad de los datos o claves de validación usados para asegurar que la información de contenido no ha sido alterada.

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

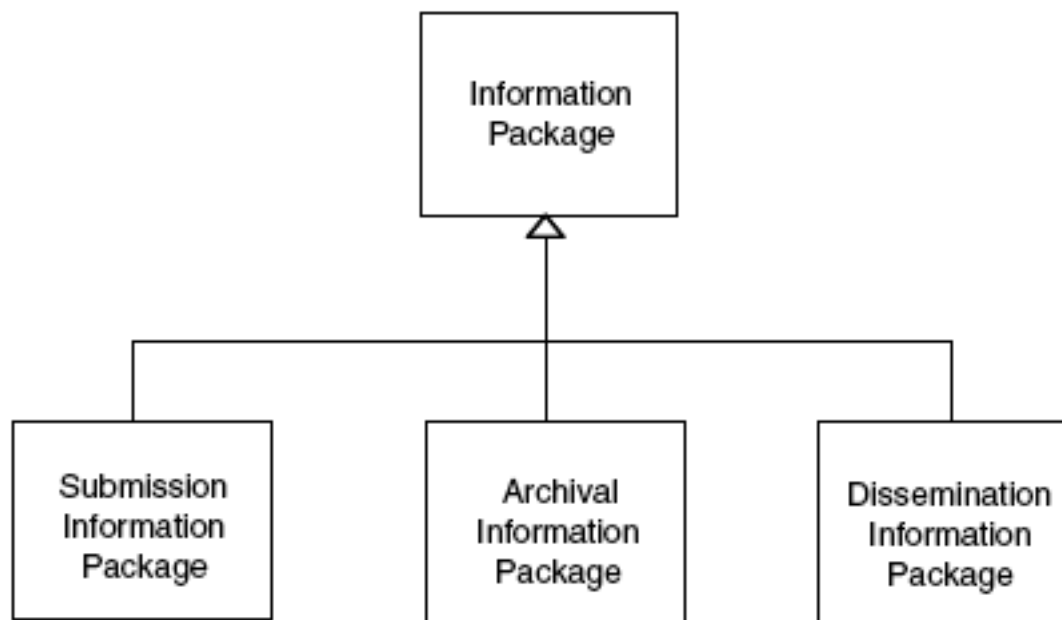
Paquetes de información en OAIS

- 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



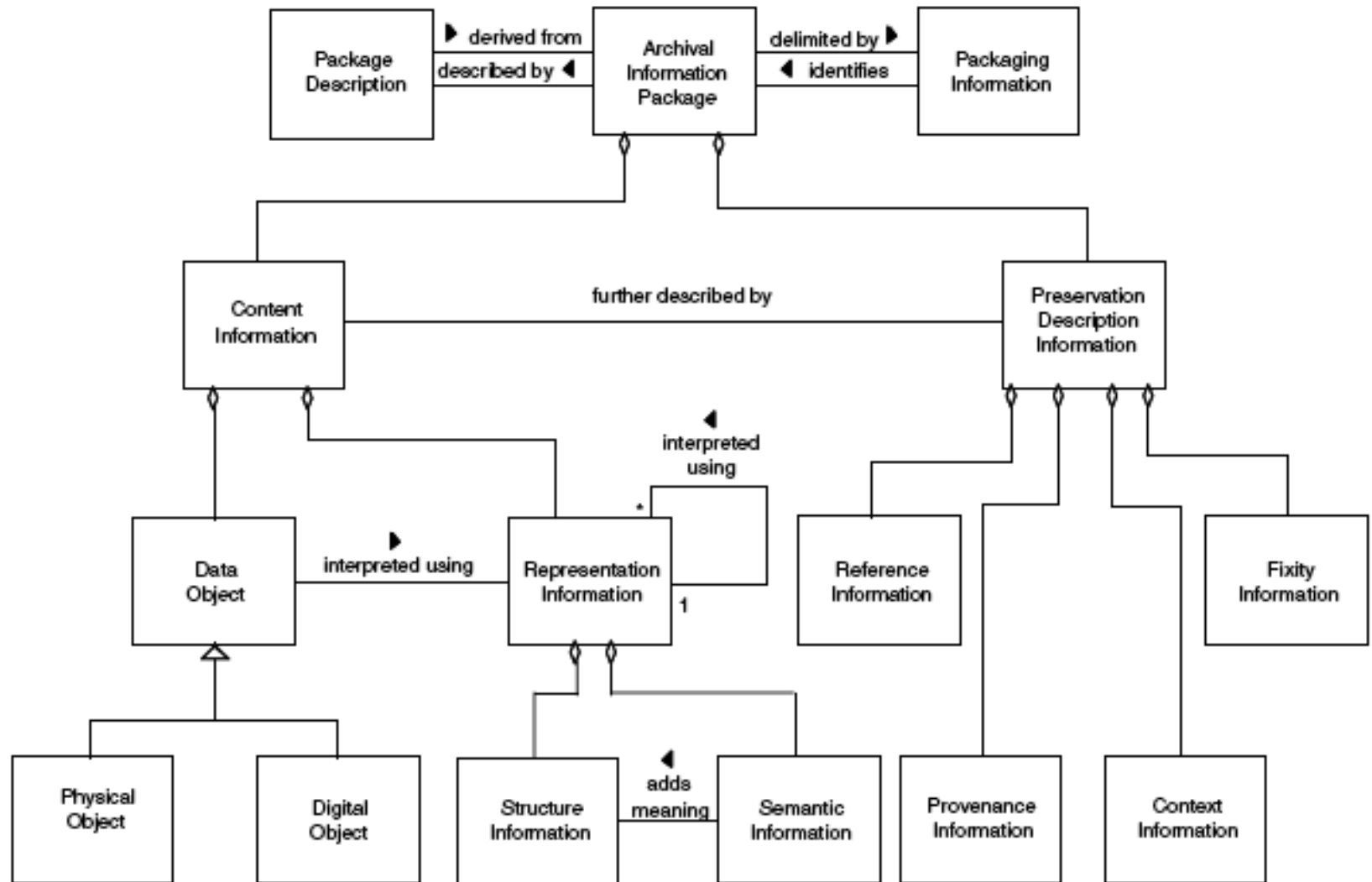
Tipos de paquetes de información



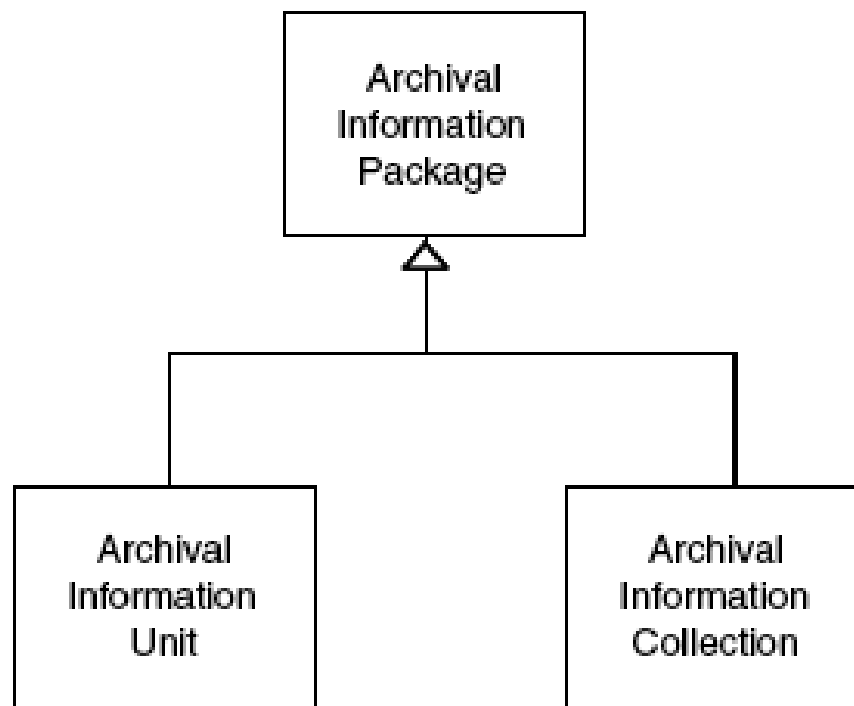
SIP

- 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.
- 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.
- OAIS 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

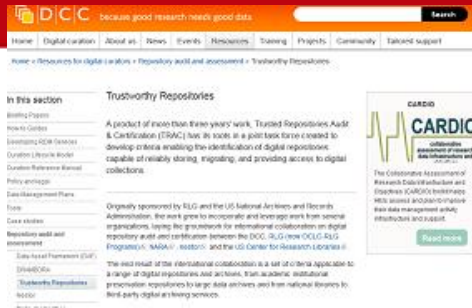
Los resultados de este módulo servirán como pautas para que los administradores diseñen sus políticas, y en última instancia, guiarán las actividades de conservación de los materiales.

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.

El Futuro es y grande, seguimos en la próxima!!!



DICC because good research needs good data

Home | Digital curation | About us | News | Events | Resources | Training | Projects | Community | Contact us

Trustworthy Repositories

A project of more than three years work, Trusted Repositories Audit & Certification (TRAC) has its roots in a joint task force created to develop criteria enabling the identification of digital repositories capable of reliably storing, ingesting, and providing access to digital collections.

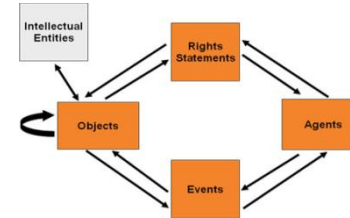
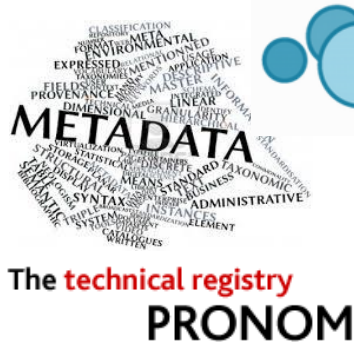
Originally sponsored by RCUK and the US National Archives and Records Administration, the work grew to incorporate and merge work from several organisations, laying the groundwork for international collaboration on digital repository audit and certification between the DCC, SLU, the OCLC-RLG Programme, Noddy, eprints, and the US Center for Research Libraries.

The next result of the international collaboration is a set of criteria applicable to a range of digital repositories and activities, from academic, institutional preservation repositories to large data archives and from national libraries to large early digital archiving services.



The Library of Congress » Standards » **PREMIS Home**

P R E M I S PRESERVATION METADATA MAINTENANCE ACTIVITY

METADATA

Word cloud containing terms related to metadata and digital preservation, such as: CLASSIFICATION, ENVIRONMENTAL, EXPRESSED, PROVENANCE, DIMENSIONAL, STATISTICAL, MEANS, SYNTAX, ADMINISTRATIVE, etc.

The **technical registry**
PRONOM



Open Planets Foundation

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



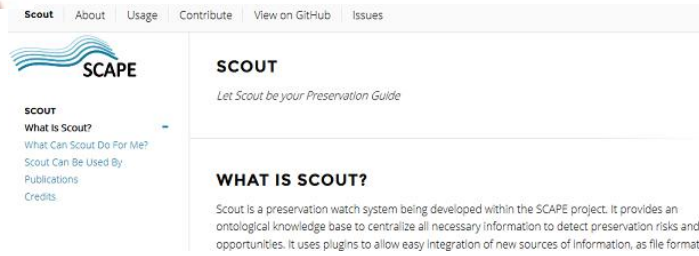
Droid

DROID (Digital Record and Object Identification)



SHAMAN

Sustaining Heritage Access through Multivalent Archiving



SCOUT

Let Scout be your Preservation Guide

WHAT IS SCOUT?

Scout is a preservation watch system being developed within the SCAPE project. It provides an ontological knowledge base to centralize all necessary information to detect preservation risks and opportunities. It uses plugins to allow easy integration of new sources of information, as file format

ISO 14721: 2012
ISO 16363: 2012




APA RSEN

Alliance Permanent Access to the Records of Science in Europe Network

DRAMBORA interactive

Digital Repository Audit Method
Based on Risk Assessment

