



SeDiCI SERVICIO DE DIFUSIÓN
DE LA CREACIÓN INTELECTUAL



Bibliotecas y Repositorios Digitales

Tecnología y Aplicaciones

<http://sedici.unlp.edu.ar>



Participantes del dictado

- Marisa De Giusti
- Nestor Oviedo
- Silvia Pelоче
- Matías Cánepa



Objetivo del curso

- Compartir la experiencia del SeDiCi en todas las áreas que hacen al quehacer del repositorio: edición, catalogación, comunicación y difusión, software de soporte e interoperabilidad, servicios asociados y cuestiones legales, entre otras. Cada área temática será dictada por distintos integrantes del SeDiCi, de acuerdo al perfil de sus tareas.
- Crear conciencia sobre el acceso abierto en todas sus formas.



Bibliotecas y repositorios digitales

Capítulo 1: Panorama y definiciones. Movimiento de Acceso Abierto y la crisis del sistema tradicional de comunicación científica. Estrategias: revistas de acceso abierto y repositorios temáticos e institucionales. Bibliotecas digitales y repositorios institucionales: desarrollo y situación actual. Roles institucionales de un repositorio institucional. El rol de la institución. Impacto para los investigadores de una institución: diferentes estudios y perspectivas. Visibilidad e impacto de un repositorio: ROAR, OpenDOAR, Webometrics, estadísticas. Comentarios sobre repositorios disciplinares y de objetos de aprendizaje.

Movimiento de Acceso Abierto

Open access...



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.

Movimiento de Acceso Abierto

Open access...



- Referencia: Wiki - **Timeline of the Open Access movement**

Línea de tiempo de eventos en la historia del OA. Comenzó como una línea de tiempo mantenida por Peter Suber y se trasladó a OAD en 2009.



Peter Suber
Senior research professor of
philosophy at Earlham College

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

- 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*".



1991 - ArXiv

- ArXiv cubre los campos de la física, matemáticas y ciencias de la computación y los artículos eran y son depositados por los autores antes de su remisión a las revistas especializadas donde sufren el proceso de revisión por pares. Originalmente fue alojada en el Laboratorio Nacional de los Álamos y actualmente en la Universidad de Cornell.
- <http://arxiv.org/>
- 784152 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics.



1991 - ArXiv

- Durante todos estos años se ha discutido el modelo de ArXiv y su viabilidad a otros campos del saber.



1992 – ARL - SPARC

- En el año 1992 interesa destacar que 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. Estimula la emergencia de nuevos modelos de comunicación académica.



1993 - WWW

1993 es un año trascendental que desborda el marco de estas iniciativas: 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, uno de los líderes más activos del movimiento lanza la iniciativa por el *autoarchivo* (self-archiving): el propio autor deposita su trabajo en un lugar confiable.



School of Electronics and Computer Science.
University of Southampton.



Stevan Harnad 1994

- Básicamente 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 que ya han tenido revisión.
- Harnad analiza el impacto para la comunidad científica si los autores depositaran sus 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 abierto el acceso a este tipo de trabajos.



1997

- En 1997 se lanza CogPrints el primer depósito de artículos de investigación en las áreas de psicología, neurociencias, lingüística, filosofía y ciencias de la computación.



1997



En 1997 se inicia también el acceso libre al MEDLINE a través del Pubmed, por el lanzamiento de la 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.



1998



A partir de 1998 el curso de los acontecimientos se acelera. Los consejos editoriales de algunas revistas científicas rompen con la casa editora comercial por las serias divergencias en cuanto a la visibilidad de las mismas a través de la red. En torno a estos movimientos, la recién lanzada SPARC lanza su revista Declaración de Independencia que fue una guía para crear publicaciones científicas controladas por los mismos académicos.

1998: la irrupción del mundo hispano en el movimiento



- **La declaración de San José** (Costa Rica) por los delegados del Sistema de Información en Ciencias de la Salud de Latinoamérica y el Caribe (**Bireme**), institución fundadora del Pubmed hispano denominado **SciELO**.
- *“Nos comprometemos a construir en forma cooperativa la Biblioteca Virtual en Salud, que será la respuesta integradora de la Región a este desafío, facilitando el más amplio acceso a la información para el mejoramiento permanente de la salud de nuestros pueblos. Al propio tiempo será una herramienta para el fortalecimiento de los sistemas de salud y para el desarrollo humano sostenible de la Región”.*



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.



Reunión de Santa Fé (Albuquerque)

- Reunión en Santa Fé, Albuquerque, Estados Unidos, establecimiento de la iniciativa Open Archives Initiative (OAI) en 1999, destinada a establecer una serie de principios organizativos y especificaciones técnicas para permitir que los diversos sistemas de archivo y publicación fueran interoperables.



OAI-PMH

- El OAI llevó a la aparición del OAI-PMH (Protocol of Metadata Harvesting) para facilitar el intercambio de los metadatos entre los acervos. Este protocolo se basa en el esquema de metadatos Dublin Core que es muy simple y contiene 15 elementos, la simplicidad estaba orientada a facilitar el autoarchivo por parte de los autores.

Metadatos



Son datos que describen otros datos. Son información estructurada que sirve para describir, explicar, localizar o facilitar la obtención, el uso o la administración de un recurso de información, es decir no son meramente descriptivos.



OAI-PMH

Lo que es importante resaltar es que el protocolo separa los metadatos del objeto digital. La idea es que los acervos funcionan como proveedores de datos que ofrecen sus metadatos para ser cosechados por los proveedores de servicios, los servidores pueden incluso no proveer ambos servicios.

OAI-PMH



El Protocolo OAI-PMH es un mecanismo de baja barrera para la interoperabilidad del repositorio. Los proveedores de datos son repositorios que exponen metadatos estructurados a través del protocolo OAI-PMH; los proveedores de servicios a continuación, realizan vía OAI-PMH solicitudes de servicio para cosechar los metadatos.



OAI-PMH: seis verbos o servicios que se invocan en HTTP

Requiere como mínimo el uso de Dublin Core

1. **Identify** - identifica la colección
2. **ListMetadataFormats** - obtiene formatos de metadatos disponibles en la colección
3. **ListIdentifiers** - obtiene encabezados de objetos (id, fecha, conjunto)
4. **ListSets** - Obtiene la organización de conjuntos de la colección
5. **GetRecord** - obtiene metadatos de un objeto
6. **ListRecords** - obtiene metadatos de objetos filtrando por conjunto o fecha



OAI-PMH

El registro de los metadatos puede o no tener el texto completo asociado, aunque la intención siempre ha sido que todos los registros cuenten también con el artículo, es técnicamente posible implementar OAI-PMH con registros de metadatos y sin objetos digitales como contraparte.



OAI-PMH

Existen dos clases de participantes en el marco del protocolo OAI-PMH:

- 1) **Proveedores de datos:** albergan un repositorio con los recursos que se quieren publicar y exponen los metadatos de dichos recursos para ser recuperados por los proveedores de servicios.
- 2) **Proveedores de servicios:** recuperan metadatos de los proveedores de datos y los utilizan para dar servicios sobre dichos datos (interfaz de búsqueda,...)

Convención de Santa Fe, Albuquerque, USA



En la convención de Santa Fé 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.

Convención de Santa Fe, Albuquerque, USA



La convención consideró que los siguientes eran componentes cruciales:

- 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

En el 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** en Febrero de 2000.



2000



Con menor impacto, pero a destacar por la importancia de la institución que declara es la **Declaration on Science and the Use of Scientific Knowledge** por parte de la UNESCO.



2000



En el 2000 también 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



En el 2000, varios científicos involucrados en el desarrollo de *PubMed* Central, fundaron un grupo llamado *Public Library of Science* (PloS) que circuló una carta abierta en la que se exigía un vuelco al sistema de comunicación científica.

Este es un hito que muchos autores consideran inicia el OA como lo conocemos hoy en día.

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”.





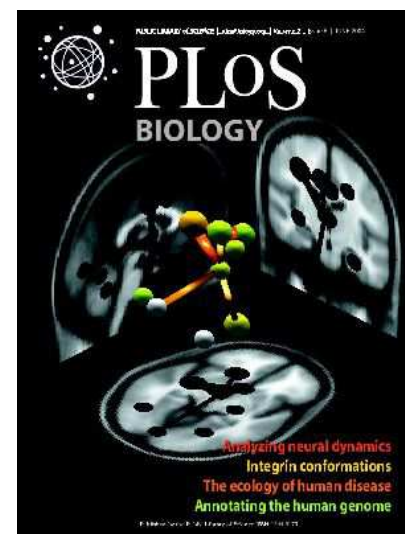
2000 - PLoS

- 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.



2000 - PLoS

PloS que se convirtió en una editorial de acceso abierto y lanzó sus dos primeras revistas de acceso abierto: *PLos Medicine* y *Plos Biology*.





Para reflexionar...

Hoy día Plos mantiene siete publicaciones periódicas en las áreas de medicina, biología y temáticas relacionadas pero, y esto también da para pensar, las tasas de la publicación son pagadas por los autores o por las instituciones que los albergan, o los sponsors con los que cuentan. Los números de los que se habla resultan muy altos.



Cargos de publicación

Publication Fees for PLoS Journals

- *PLoS Biology* US\$2900
- *PLoS Medicine* US\$2900
- *PLoS Computational Biology* US\$2250
- *PLoS Genetics* US\$2250
- *PLoS Pathogens* US\$2250
- *PLoS ONE* US\$1350
- *PLoS Neglected Tropical Diseases* US\$2250

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.



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 en todos los casos dando disponibilidad gratuita en Internet, para que cualquier usuario la pueda leer, descargar, copiar, distribuir o imprimir, así como bucear dentro del artículo sin otras barreras financieras, legales o técnicas que las de acceso a la red, con *“la única función del copyright en este dominio, no puede ser otra que dar a los autores control sobre la integridad de su trabajo y el derecho a ser apropiadamente acreditados y citados”*.



Bethesda

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



Las contribuciones de acceso abierto deben satisfacer dos condiciones:

1. El(los) autor(es) y depositario(s) de la propiedad intelectual de tales contribuciones deben garantizar a todos los usuarios por igual, el derecho gratuito, irrevocable y mundial de acceder a un trabajo erudito, lo mismo que licencia para copiarlo, usarlo, distribuirlo, transmitirlo y exhibirlo públicamente, y para hacer y distribuir trabajos derivados, en cualquier medio digital para cualquier propósito responsable, todo sujeto al reconocimiento apropiado de autoría (los estándares de la comunidad continuarán proveyendo los mecanismos para hacer cumplir el reconocimiento apropiado y uso responsable de las obras publicadas, como ahora se hace), lo mismo que el derecho de efectuar copias impresas en pequeño número para su uso personal.

Berlín



2. Una versión completa del trabajo y todos sus materiales complementarios, que incluya una copia del permiso del que se habla arriba, en un conveniente formato electrónico estándar, se deposita (y así es publicado) en por lo menos un repositorio online, que utilice estándares técnicos aceptables (tales como las definiciones del acceso abierto), que sea apoyado y mantenido por una institución académica, sociedad erudita, agencia gubernamental, o una bien establecida organización que busque la implementación del acceso abierto, distribución irrestricta, interoperabilidad y capacidad archivística a largo plazo.



Estrategias en AA-OA

- El *movimiento de acceso abierto* a la información se basa en dos estrategias fundamentales para garantizar el acceso y diseminación sin restricciones económicas y legales de la información científico-técnica: las revistas de acceso abierto y los repositorios temáticos e institucionales.



Repositorios

- Los propios autores depositan sus artículos en repositorios *temáticos o institucionales*. Estos artículos pueden estar en fase de publicación en una revista tradicional (*preprints*) o haberse publicado (*postprint*). Autoarchivo (*self-archiving*).
- Normalmente a este camino de publicación se le llama “**ruta verde**”. Básicamente es el nombre que se da a la publicación secundaria de las publicaciones principales accesibles por pago.



Revistas de Acceso abierto

- Revistas de acceso abierto (*Open-access Journals*): Ruta dorada. Son revistas cuyos contenidos están disponibles libre y gratuitamente en Internet, pueden ser genuinas o híbridas, según todos o algunos de sus artículos estén gratuitos.

Rutas de publicación



Fuente: Max Planck Society

¡Conocer los derechos!



Resulta importante que los autores revisen una y otra vez sus nociones respecto de derechos para resguardar sus trabajos, de hecho este resulta también un fundamento para la creación y mantenimiento de los repositorios institucionales.



El sistema de comunicación científica y la crisis de su entorno

Desde hace unas décadas consultores, científicos, bibliotecarios y editores han señalado que el sistema tradicional de comunicación científica se encuentra en crisis, porque no se cumplen con los objetivos primarios de este: favorecer la diseminación y el intercambio de los resultados científicos para lograr avanzar en ciencia y obtener mayor progreso científico, técnico y social para la sociedad.



Factores de la crisis

- El incremento sostenido de los precios de las revistas científicas, sobre todo en las áreas de ciencia, tecnología y medicina que se ha denominado en la literatura especializada “serial crisis”.
- Un segundo problema que es de destacar e impacta grandemente en áreas científico técnicas de gran movilidad, como es el caso de la informática, es la extensión del período entre envío de un artículo y su publicación definitiva, necesidad de un cambio de reglas.



Factores de la crisis

- Fusiones y adquisiciones de empresas editoriales –las más pequeñas desaparecen en manos de las más grandes- por lo que se establece un mercado sin competencia.
- Restricciones a los derechos de autor para dar acceso y diseminación de la información científica, que han desvirtuado los objetivos primarios de la comunicación científica y del propio derecho de autor, y aquellos relativos al sistema de recompensa científica, enfocado más a la publicación en revistas “de impacto” que a la amplia diseminación de los resultados científicos.

Sinergia



- Cada vez mayor reconocimiento de que casi toda la investigación se financia con fondos públicos y que para maximizar la disseminación de sus resultados, deben estar disponibles con acceso abierto.
- Las potencialidades de las tecnologías de la información y la comunicación (TICs), han facilitado la creación de revistas electrónicas y otras plataformas que tienen el potencial de permitir un acceso más amplio a la información.
- La sinergia entre los diversos aspectos antes señalados ha contribuido a fortalecer toda una corriente de pensamiento y acción transdisciplinaria e internacional a favor de la ampliación del acceso a la información científica sin barreras económicas ni legales.



Sobre los derechos de autor y el sistema científico

Las legislaciones de derecho de autor se crearon con la intención de proteger al titular de los derechos patrimoniales de una obra (copyright) contra los usos indebidos que terceros podían hacer de estas. Sin embargo, en el caso de las revistas científicas se da la paradoja de que, en un gran número de casos y ya “tradicionalmente” todos los derechos patrimoniales pasan a manos de los editores



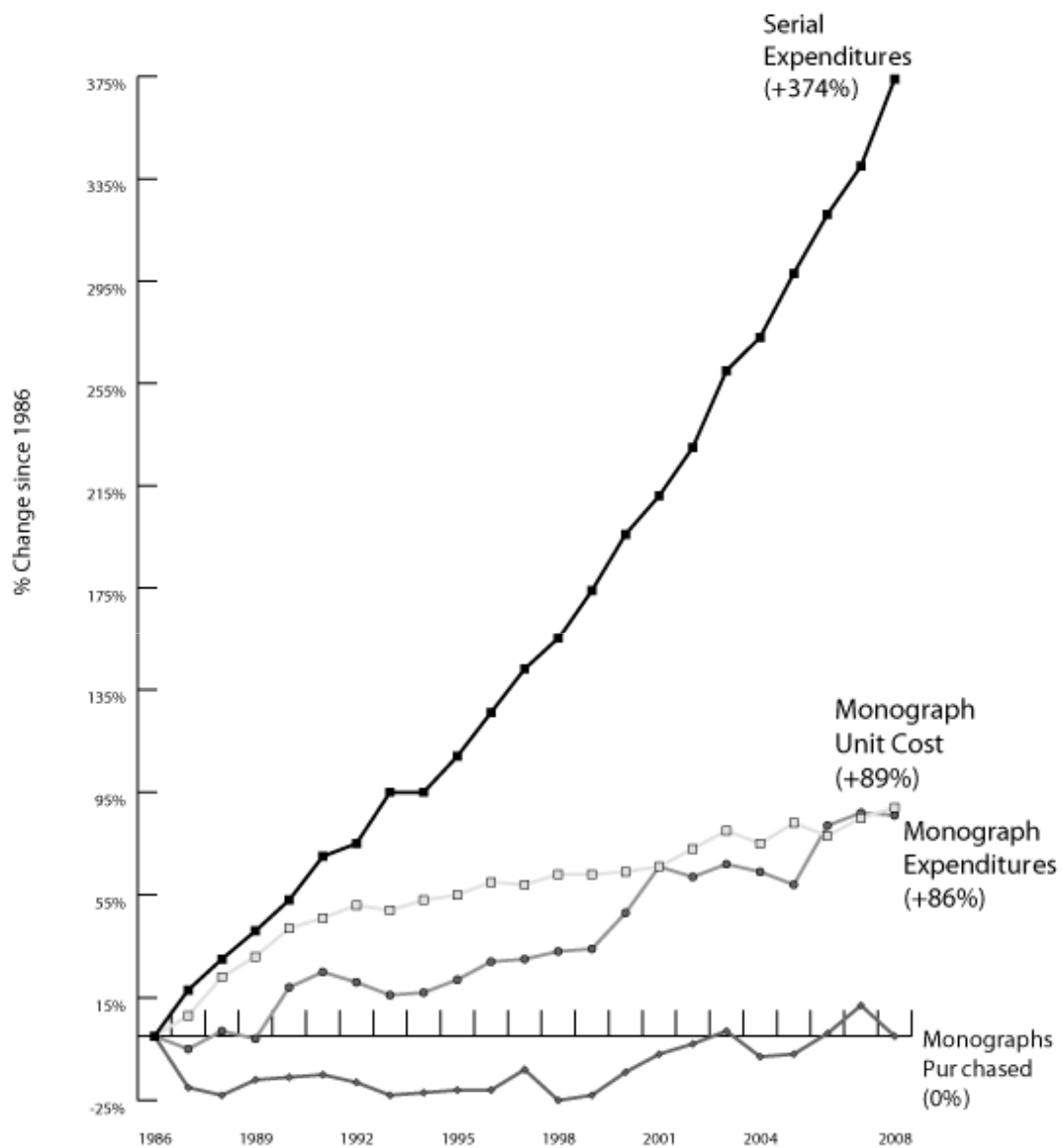
Conocer los derechos

Vale preguntarse qué porcentaje de los editores solicitan hoy día la cesión exclusiva del derecho de autor, qué derechos retiene el autor a usar su propio trabajo, o, en el caso que nos ocupa en nuestro ámbito a hacer depósito en un repositorio institucional de la propia institución que ha apoyado económicamente su trabajo?.

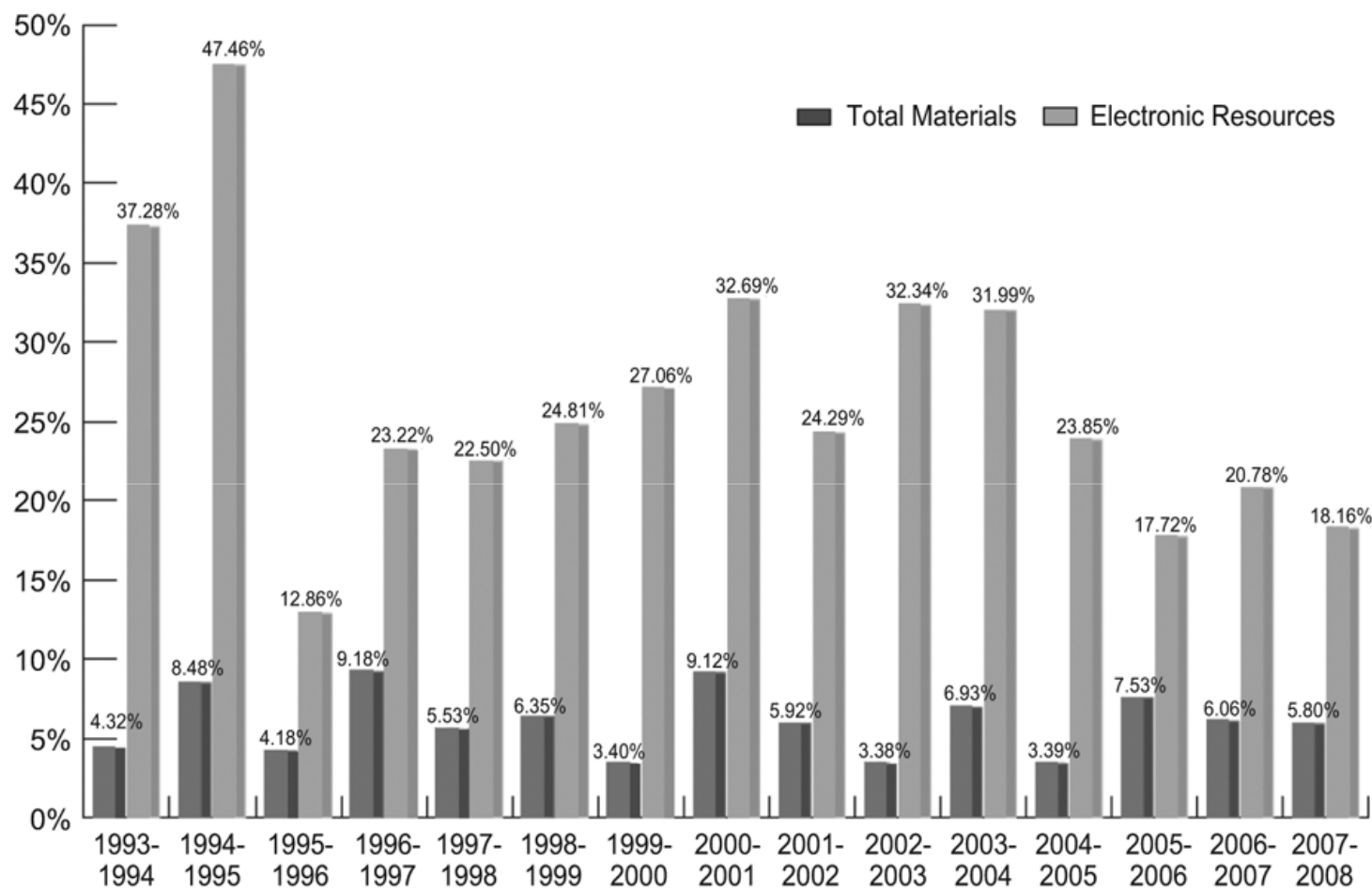


Comentarios en off

Hasta hace un tiempo, esto no constituía una preocupación para los investigadores, acostumbrados a ceder su trabajo y no obtener ningún beneficio a cambio, hoy día el advenimiento de las TICs y especialmente Internet, que habilitan una mayor difusión, tienen como contracara, restricciones cada vez mayores impuestas por las legislaciones de derecho de autor. Un ejemplo más que curioso que cita Sánchez Tarragó es que en Estados Unidos, estas leyes limitan el “uso justo” institucional a sólo cinco artículos publicados en los últimos cinco años de cualquier revista. Una vez que ese límite es alcanzado, cualquier artículo adicional debe pagarse al editor, sea por concepto de préstamo interbibliotecario o por distribución de documentos. Otro detalle que menciona la autora cubana es que existe una práctica común por parte de los editores de prohibir el uso de suscripciones electrónicas para préstamo interbibliotecario, dado que las bibliotecas van aumentando las suscripciones publicaciones en desmedro de las en papel, la disponibilidad va descendiendo.



*Includes electronic resources from 1999-2000 onward.
Source: ARL Statistics 2007-2008, Association of Research Libraries, Washington, DC



Source: Kyrillidou and Bland (2009, p.19)

online

digital.
CSIC



DIGITAL.CSIC

ACCESO ABIERTO
A DOCUMENTOS DIGITALES

DIGITAL.CSIC:

“Catalogación con Dublin Core, gestión de derechos de autor y permisos de editores. Medida, visibilidad e impacto de la producción científica”

25 de mayo 2011

<http://digital.csic.es/>

Isabel Bernal
Unidad de Coordinación de Bibliotecas del CSIC
Oficina Técnica de Digital.CSIC

El impacto de la investigación científica y el acceso abierto



El acceso abierto como modelo alternativo (1/2)

- Entre los objetivos principales del acceso abierto se encuentran:
 - maximizar la visibilidad y la accesibilidad de los resultados de la investigación
 - acortar los tiempos de la comunicación científica
 - analizar y evaluar el *verdadero* impacto y la excelencia de la investigación financiada con fondos públicos

El acceso abierto como modelo alternativo (2/2)

- 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



Factor de impacto

El factor de impacto de una publicación se usa habitualmente para medir el impacto de un trabajo científico y premiar (o castigar) a los investigadores involucrados. El uso del factor de impacto de la revista para fines de evaluación científica genera polémicas, porque muchas veces se asume como representativo de todos los artículos que en la revista se publican y por tanto, como una medida cuantitativa—y objetiva—de la calidad del resultado científico publicado, incluso de esta suposición se desprenden algunas prácticas perversas como las autocitas o las estrategias editoriales para subir el factor de impacto.



Factor de impacto

Los comportamientos de unos y otros resultan tergiversados y se pierden los objetivos prioritarios, es decir, a pesar del interés de la comunidad científica por hacer públicos los resultados de sus trabajos, hay un empuje decidido de uno y otro lado para difundirlos en determinadas publicaciones que garanticen unos resultados a expensas del acceso y con el agregado de unos costos altísimos.



BDs en el mundo digital

El concepto de biblioteca digital ha tenido un crecimiento sostenido desde los albores del año 2000, empujado por el incremento de los recursos de cómputo, las redes y el decremento paralelo de los costos para acceder a servicios de este tipo.



BDs en el mundo digital

Algunos piensan que la biblioteca digital mundial es la propia web, el gran sueño borgiano de la “biblioteca infinita”; sin embargo esta apreciación deja de lado un punto fundamental que comparten todas las bibliotecas, más allá de la naturaleza de los objetos que contienen, esto es las **bibliotecas intrinsecamente conllevan un concepto de selección.**



BDs en el mundo digital

Está claro que desde las apreciaciones vertidas en la creación de la Biblioteca del Congreso, depositaria de los fondos “necesarios” para la actividad del congreso hasta las colecciones gigantescas de hoy en día, el concepto de necesidad o bien de interés parece esfumarse, sin embargo, esto responde al crecimiento exponencial y a la diversidad de sus usuarios, lo que pareciera alejarla de la constrictión de la que partieron, de que todo su material pudiera ser de interés para académicos e investigadores.



BDs primeras definiciones

La primera reflexión que parece importante hacer notar es que hay un plus en las bibliotecas, sean estas de colecciones digitales o en papel, esto es que hay un subconjunto de objetos de información seleccionados (en oposición a otros contenidos excluidos), segregados, disponibles, preservados, cuyo acceso está favorecido por servicios añadidos, entre otros las posibilidades de búsqueda de información.



DLIs

En la década de los 90's, organismos estatales de los Estados Unidos de América como NSF DARPA y NASA llevaron a cabo dos iniciativas: DLI-1 y DLI-2 (Digital Library Initiative) cuyo objetivo fue desarrollar e implementar modelos de bibliotecas digitales.



DLI-1 –Programa –Definición-1994

"El enfoque de la iniciativa es avanzar considerablemente en todos los medios para recopilar, almacenar y hacer disponible para la búsqueda, recuperación y procesamiento a través de las redes de comunicación - todos de una manera fácil de usar. Las BDs, básicamente, almacenan materiales en formato electrónico y manipulan grandes colecciones de dichos materiales con eficacia. La investigación sobre BDs es la investigación en los sistemas de información de la red, concentrándose en cómo desarrollar la infraestructura necesaria para efectivamente manipular la gran masa de información de la red. Las cuestiones técnicas clave son cómo buscar y mostrar las selecciones deseadas a través de grandes colecciones".



Cambio en las definiciones de BDs

Desde 1994 hasta el presente esto ha ido cambiando gradualmente y ha devenido en la cada vez mayor participación de los bibliotecólogos en el quehacer de las BDs. El portal de la JCDL (Joint Conference on Digital Libraries, un importante foro internacional centrado en las bibliotecas digitales y las cuestiones técnicas, prácticas y sociales que aúna la excelencia de conferencias ya establecidos por la ACM y la IEEE-CS muestra el devenir de esta práctica a lo largo del tiempo. Es visible al examinar las conferencias realizadas desde 1994, los temas, autores y la filiación de los mismos.



Cambio en las definiciones de BDs

El portal de JCDL es un buen lugar para comenzar a bucear en definiciones reconocidas y encontrar la convergencia de un término esquivo, refiriéndose a la Conferencia, los responsables se extienden en los muchos significados que abarcan las palabras “bibliotecas digitales”.



Definiciones de un concepto esquivo

Lesk (1997): *“Las Bibliotecas digitales son colecciones organizadas de información digital. Combinan la estructura y concurrencia de la información, que siempre han tenido las bibliotecas y los archivos, con la representación digital que han hecho posible las computadoras”.*



Definiciones de un concepto esquivo

- Borgman (1999), distingue dos sentidos distintos, una definición tecnológica que establece que: *“Las bibliotecas digitales son un conjunto de recursos electrónicos y capacidades técnicas asociadas para crear, buscar y utilizar la información ... son una extensión y mejora de sistemas de almacenamiento y recuperación que manipulan los datos digitales en cualquier medio. El contenido de las bibliotecas digitales incluye los datos y metadatos “*
- *“ las bibliotecas digitales son construidas, recopiladas y organizadas, por (y para) una comunidad de usuarios, y sus capacidades funcionales de apoyo a las necesidades de información y usos de la comunidad” .*



Definiciones de un concepto esquivo

Arms (2000) propuso una definición informal: *“una biblioteca digital es una colección gestionada de información, con servicios asociados, donde la información es almacenada en formato digital y es accesible en toda la red”*. Esta definición enfatiza los aspectos de la gestión de los contenidos.



Definiciones de un concepto esquivo

- La Digital Library Federation (DLF) es una organización establecida en los Estados Unidos en 1995, dedicada a la creación, el mantenimiento, la expansión y distribución de colecciones distribuidas de materiales digitales accesibles para escolares y un sector de público amplio.
- “Digital libraries are organizations that provide resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities”. (DLF, 21 de Abril de 1999).



Definiciones de un concepto esquivo

El concepto de Biblioteca digital fue, tal cual afirma Jesus Tramullas Saz un concepto que se definió cuando las BDs llegaron a un nivel de madurez tal que el objeto de estudio estuvo lo suficientemente definido por sus propias funciones y herramientas así como por los componentes tecnológicos.



Definiciones de un concepto esquivo

Candy Shwartz, recorre un conjunto de definiciones formales e informales para mostrar los recursos de una BD, a quienes sirven los mismos y qué materiales y funcionalidades brindan. La “punta de iceberg” desde la cual parte el trabajo de la Profesora de Boston es la necesidad de proveer un contexto a través de la perspectiva de los componentes del trabajo de una biblioteca digital y también apuntar a los recursos para una exploración más profunda. Un hecho interesante que plantea la autora, es que más allá de que existan las bibliotecas “híbridas”, la noción de una entidad separada denominada biblioteca digital, emerge independientemente, en parte por el impulso enorme de un cuerpo de investigadores de la mitad a finales de los 1990s y fueron solventados por agencias gubernamentales o colaboradores a gran escala.



De esquivo a complejo

DELOS: “Digital Library Manifiesto”. Marco conceptual con tres niveles:

- Biblioteca Digital (DL) es la organización que recopila, gestiona, preserva y ofrece contenidos digitales;
- Sistema de Biblioteca Digital (DLS) que se refiere al sistema de software que proporciona la funcionalidad requerida por una Biblioteca Digital particular y
- Sistema de Gestión de Biblioteca Digital (DLMS) que se refiere a la plataforma: sistema operativo, bases de datos, interfaz de usuario.



Noción compleja

Todas las nociones vinculadas a las BDs necesitan complejizarse para superar la dicotomía de definiciones dadas desde “distintos lugares”: el ámbito de investigación y el ámbito bibliotecario. El área de trabajo amerita definiciones y prácticas más elaboradas porque obliga a la participación de distintos campos del saber.



Repositorios Institucionales

La Universidad Autónoma de Madrid define un repositorio institucional como un conjunto de servicios Web centralizados, creados para organizar, gestionar, preservar y ofrecer acceso libre a la producción científica, académica o de cualquier otra naturaleza cultural, en soporte digital, generada por los miembros de una institución. Las principales características de un repositorio institucional son:

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



Característica de los RIs

A lo dicho precedentemente, resulta importante agregar una característica de todos ellos en conjunto: **la diversidad**. Si existe un elemento que los nuclea es, precisamente que ninguno se parece a otro.

Deslindes terminológicos y aclaraciones



Se confunden y aúnan las definiciones de las bibliotecas digitales con las de los repositorios institucionales, de hecho los Repositorios institucionales comparten muchas características con las bibliotecas digitales, pero se pueden distinguir algunas características importantes. Los repositorios institucionales están diseñados principalmente para recoger, preservar y poner a disposición la producción académica de una institución, alternativamente, las bibliotecas digitales pueden estar organizadas en torno a otros principios: temas, disciplinas, o incluso tipos de documentos en particular.

Deslindes terminológicos y aclaraciones



Los repositorios institucionales y las bibliotecas digitales también difieren en cuanto a cómo adquirir contenidos. Mientras que las colecciones que figuran en las bibliotecas digitales son generalmente el resultado de esfuerzos deliberados de desarrollo de la colección por parte de los profesionales de la biblioteca, los repositorios institucionales son típicamente dependientes de las contribuciones voluntarias de los investigadores. Mientras que algunas instituciones han tratado de exigir el depósito de determinados tipos de materiales, la mayoría dependen de la participación voluntaria. Este hecho está detrás de uno de los mayores desafíos que enfrentan hoy los repositorios institucionales, es decir, la tasa relativamente baja de la contribución de los investigadores.

Deslindes terminológicos y aclaraciones



Otra diferencia entre los repositorios institucionales y bibliotecas digitales es que los repositorios institucionales son fundamentalmente un lugar para almacenar los materiales. En consecuencia, puede haber servicios mínimos ofrecidos a los usuarios. Por el contrario, las bibliotecas digitales a menudo ofrecen servicios a los usuarios, que incluyen el apoyo del personal en la búsqueda de información adicional e incluso en la interpretación.

Deslindes terminológicos y aclaraciones



Los repositorios institucionales pertenecen a una institución académica o de investigación, y se pretende que los materiales de la casa que representan la producción intelectual de esa organización estén presentes; debido a esto es, necesariamente una colección de documentos y objetos, por lo general de varios tipos y formatos. Investigadores afiliados a la organización patrocinadora puede depositar los textos, los conjuntos de datos, archivos de sonido, imágenes o cualquier número de otros artículos.

Significativamente, estos documentos pueden estar en cualquier etapa del proceso de la investigación académica: pre prints, postprints, material que no ha pasado procesos de referato, etc., todo lo cual también depende de la política de la institución en cuanto a qué tipo de documentos pretende alojar.

Finalmente, los repositorios institucionales están estrechamente ligados a los ideales y objetivos del movimiento de acceso abierto y la creencia de que la comunicación científica debe ser lo más abierta y libre posible. La siguiente tabla recoge las diferencias más importantes#



Ris versus BDs

Repositorios Institucionales

- Están organizados en relación a una comunidad institucional en particular.
- A menudo dependen de la contribución voluntaria.
- Son principalmente repositorios y pueden ofrecer servicios limitados a los usuarios.

Bibliotecas Digitales

- Construídas en torno a un número diferente de posibles principios organizativos: tópico, sujeto, disciplina...
- Son producto de una estrategia deliberada para elección de la colección.
- Incluyen aspectos importantes de servicio: referencia, asistencia, interpretación de contenidos...



Ejemplos de BDs y Repositorios

Biblioteca Digital	Repositorios
Disciplinar	arXiv
Alexandria DL	PubMed Central
Perseus Project	EconPapers
Digital Scriptorium	Cogprints
Center for Electronic Texts in the Humanities	
Intitucional	Edinburgh Research Archive
California Digital Library	Dspace at MIT
Illinois Digital Academic Library	Nottingham Eprints
	LSE Research Articles Online

ROAR



All Repositories - Registry of Open Access Repositories - Mozilla Firefox

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

Gmail - Recibidos (474) - mdegiusti@... Google Docs - Home tesis.docx - Google Docs All Repositories - Registry of Open Ac...

http://roar.eprints.org/cgi/roar_search/advanced?location_country=&software=&type=&order=-recordcount%2F-date

Más visitados WR Celsius PrEBi SeDiCI WebUNLP BBVA BN BP CSIC OpenAIRE ACM OR Stats grefine JI IS CRM SS Freebase Cogprints

Registry of Open Access Repositories

Home About Search Search Content Browse

Login New Entry Create Account

Search

All Repositories


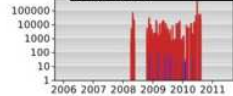
Any Country Any Software Any Repository Type Sort by number of reco Search


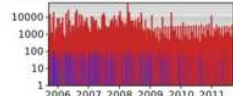
Any Repository Type

- Research Institutional or Departmental (1540)
- Research Multi-institution Repository (10)
- Research Cross-Institutional (194)
- e-Journal/Publication (120)
- e-Theses (189)
- Database/A&I Index (47)
- Research Data (2)
- Open and Linked Data (11)
- Learning and Teaching Objects (36)
- Demonstration (19)
- Other (273)

Export 2441 results as Activity table

Search Content Graphical analysis

1.   [Hispana \(2600798 records\) Ministry of Culture - 8 April 2008 \[Record Details\]](#)

2.   [PubMed Central \(2129250 records\) - 27 February 2001 \[Record Details\]](#)

ES 9:19



Realidad de los RIs

La realidad del repositorio y especialmente la necesidad de recabar los materiales que lo nutren, llevan a introducir un servicio de depósito “mediado” en que los profesionales brinden una gama de servicios incluyendo la digitalización de artículos de papel y orientación sobre derechos de autor. En un sistema así, los bibliotecarios pueden ser los contribuyentes al incorporar los trabajos de los docentes e investigadores.



Realidad de los RIs

Es de observar que la actividad real del repositorio, para que el mismo cumpla el fin de exponer la producción intelectual de una casa de estudios, hace, en definitiva, que sus servicios se aproximen a los de las BDs con lo cual la diferencia establecida previamente cae, como también lo hace la prevalencia del autoarchivo en los RIs.



Repositorio Institucional: definición

En este curso se considerará la definición más amplia de RI que contempla todos los precedentes, manteniendo como elemento diferencial exclusivamente lo mencionado en relación a alojar la producción de una institución en todas sus formas y alcances.

Repositorios de objetos de aprendizaje



Considerando que el desarrollo de repositorios con contenidos de investigación representa una migración relativamente intuitiva de prácticas de publicaciones de pre-print a post-print, con muchas de las mismas marcas todavía visibles (pre-print, revisión por pares, títulos de revistas y factor de impacto) cuando se cambia al ámbito de la enseñanza es posible observar una transición mucho menos coherente.

Repositorios de objetos de aprendizaje



La elaboración de material didáctico en forma digital abarca tanto material institucional y de autor de muchas especies distintas: material de lectura, las notas, las colecciones de imágenes, animaciones, hasta materiales publicados externamente (por lo general los libros de texto). Así que, si bien es posible que una institución podría llegar a captar la totalidad de sus resultados de la investigación en un repositorio institucional, es más complicado de ver cómo su material de aprendizaje puede ser capturado tan extensamente.

Repositorios de objetos de aprendizaje



Los derechos de autor presentan un gran obstáculo para esto en sí mismo, y las instituciones no son tan fuertes en condiciones de buscar exenciones, incluso los derechos de autor de libros de texto a los que puede haber contribuido - como lo están con los trabajos de investigación escritos por sus propios académicos.

Repositorios de objetos de aprendizaje



Sin embargo, hay un argumento obvio que se hizo para el almacenamiento de material didáctico: que pueda ser reutilizado por los colegas en contextos de enseñanza, incluso modificados.

Repositorios de objetos de aprendizaje



Los objetos de aprendizaje, sin embargo, son un grupo heterogéneo de los materiales que varían enormemente en su formato, en los requisitos de los metadatos, y en tamaño. Agrupar a todos juntos en un solo repositorio presenta unos retos considerables. Las ventajas de hacerlo, sin embargo, son las mismas que las que se aplican a productos de la investigación. Es hacer un uso más eficiente de los recursos de la institución, permitir que los contenidos digitales sean conservados en el tiempo, ofrecer una visión completa de los productos institucionales, brindar herramientas de apoyo para búsquedas pertinentes, y permitir la interoperabilidad con establecimientos similares en toda la web, contribuyendo así a un servicio mundial.

Repositorios de objetos de aprendizaje



De la misma manera como se aplica con los resultados de la investigación, los repositorios de objetos de aprendizaje, se pueden clasificar en disciplinares e institucionales. En contraste con la situación con los materiales de investigación, sin embargo, nos encontramos con que los repositorios institucionales de objetos de aprendizaje son relativamente poco frecuentes. Es raro que las instituciones vean que el agregado y agrupamiento de una colección objetos de aprendizaje agregue valor y visibilidad.

Repositorios de objetos de aprendizaje



Mientras que los materiales de investigación tienden a ser muy leído por los demás miembros de una comunidad disciplinaria en todo el mundo, el valor de los objetos de aprendizaje radica en su capacidad para ser reutilizados. Por esa razón nos encontramos con que los repositorios de objetos de aprendizaje disciplinario predominan, con unos pocos servicios interdisciplinarios emergentes, a veces de alcance nacional, tales como UK's Jorum financiado por el JISC.

Repositorios de objetos de aprendizaje



Las plataformas de software que se utilizan para estos depósitos no son tan estandarizados como lo son aquellos para los depósitos de materiales de investigación, y no son capaces de ser cosechados a través de OAI_PMH. Esto le pone límites en su valor como parte de la base de datos .



Registros de los repositorios

***OpenDOAR* - Directory of Open Access Repositories**

es un directorio de calidad controlada sobre los repositorios de acceso abierto mantenido por SHERPA en la Universidad de Nottingham. Tiene criterios estrictos para la inclusión, entre ellos que los repositorios contengan items a texto completo y sean accesibles sin requerir usuario y clave. Cada repositorio registrado es visitado por el staff de *OpenDOAR* con el propósito de su indexación. Además de proveer una lista de repositorios, permite buscar repositorios así como también contenidos dentro de los repositorios. Los datos que provee también son útiles en las operaciones de harvesting.



Registros de los repositorios

ROAR - Registry of Open Access Repositories

Es un directorio de calidad controlada sobre los repositorios de acceso abierto mantenido por la Universidad de Southampton. Tiene como punto fuerte la calidad de sus estadísticas entre las cuales es posible ver gráficos referidos al crecimiento de cada repositorio. Además de proveer una lista de repositorios, permite buscar repositorios así como también contenidos dentro de los repositorios. Los datos que provee también son útiles en las operaciones de harvesting.

ROAR



All Repositories - Registry of Open Access Repositories - Mozilla Firefox

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

Gmail - Recibidos (474) - ... x Noticia sobre el curso Se... x =?iso-8859-1?Q?1_Ense=... x Google Docs - Home x tesis.docx - Google Docs x All Repositories - Registry... x Traductor de Google x + -

http://roar.eprints.org/cgi/roar_search/advanced?location_country=&software=&type=&order=-recordcount%2F-date

Más visitados WR WR Celsius PrEBi SeDiCI WebUNLP BBVA BN BP CSIC OpenAIRE ACM OR Stats grefine JI IS CRM 5S Freebase Cogprints >>

Registry of Open Access Repositories

Home About Search Search Content Browse

Login New Entry Create Account

Search


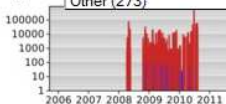
All Repositories


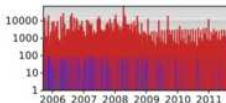
Any Country Any Software Any Repository Type Sort by number of reco Search

- Any Repository Type
- Research Institutional or Departmental (1540)
- Research Multi-institution Repository (10)
- Research Cross-Institutional (194)
- e-Journal/Publication (120)
- e-Theses (189)
- Database/A&I Index (47)
- Research Data (2)
- Open and Linked Data (11)
- Learning and Teaching Objects (36)
- Demonstration (19)
- Other (273)

Export 2441 results as Activity table

Search Content Graphical analysis

1.   [Hispana \(2600798 records\) Ministry of Culture - 8 April 2008 \[Record Details\]](#)

2.   [PubMed Central \(2129250 records\) - 27 February 2001 \[Record Details\]](#)

ES 11:08



Tipos de repositorios

Research Institutional or Departmental: a repository that accepts contents from users based on their affiliation, typically a Higher Education institution.

Research Multi-institution Repository: a single repository that acts as a central resource for many institutions. Defined by a closed set of institutions, not a specific, single discipline.

Research Cross-Institutional: a collection of research material that is defined by its research field, rather than by the bounds of a single institution. Deposits are typically made by researchers working in that field.

e-Journal/Publication: an Open Access electronic journal (should at least provide an OAI-PMH interface).

e-Theses: most content is electronic theses and dissertations.



OpenDOAR - Directory of Open Access Repositories

Es un directorio de calidad controlada sobre los repositorios de acceso abierto mantenido por SHERPA en la Universidad de Nottingham. Tiene criterios estrictos para la inclusión, entre ellos que los repositorios contengan items a texto completo y sean accesibles sin requerir usuario y clave. Cada repositorio registrado es visitado por el staff de *OpenDOAR* con el propósito de su indexación. Además de proveer una lista de repositorios, *OpenDOAR* permite buscar repositorios así como también contenidos dentro de los repositorios. Los datos que provee también son útiles en las operaciones de harvesting.

ROAR - Registry of Open Access Repositories

Es un directorio de calidad controlada sobre los repositorios de acceso abierto mantenido por la Universidad de Southampton. Tiene como punto fuerte la calidad de sus estadísticas entre las cuales es posible ver gráficos referidos al crecimiento de cada repositorio.



¿Por qué crear un Repositorio institucional?

- Los repositorios institucionales están incursionando en la conciencia de sus usuarios, pero, no es menos cierto que la marcha de los contenidos digitales a través de la Web hace que muchos de sus servicios parezcan, a priori, menos importantes de lo que son, e incluso redundantes.
- El concepto de la institucionalidad es aún frágil, la pregunta es si debemos y por qué desarrollar repositorios institucionales.



¿Por qué crear un Repositorio institucional?

- Los repositorios institucionales no aparecen como un desarrollo intuitivo para la mayoría de los académicos.
- Pocas personas todavía piensan que no necesitan una biblioteca física en su puerta del campus.
- Muchos - particularmente aquellos con experiencia en el uso repositorios creados por los propios interesados, como el caso de arXiv - se sorprende al escuchar que se preconiza la creación de repositorios institucionales como nuevos servicios.



Rol del Repositorio institucional

- Un nuevo punto de vista: son una nueva idea fuerte e importante para muchas organizaciones que se ocupan de la difusión de los resultados de la investigación.
- Su atractivo radica en la idea de "arraigo". Las propias instituciones son la base de la que emergen los resultados de la investigación - las ideas, propuestas, hipótesis, experimentos, datos y resultados presentados.
- Los resultados comparten ahora un ADN común en la representación digital. Esta base común permite a las instituciones mirar más de cerca sus mecanismos tradicionales de gestión de resultados de la investigación y descubrir si hay modos nuevos y más eficientes de operación.



Rol del Repositorio institucional

- Los resultados de investigación, tradicionalmente se volcaban en publicaciones (externas a la institución) entregadas a terceros para su posterior procesamiento.
- En este proceso hay una pérdida de control, por parte de la institución y con ello la pérdida de control de los resultados de investigaciones de los cuales la institución es la fuente de financiación, es más una pérdida de difusión a nivel del mundo porque los editores exigen pagos por la difusión.
- Los resultados, en manos de los editores, tienen que "volverse a comprar". Inevitablemente, esto significa que sólo algunos investigadores se beneficiarán. Pero si los resultados, manteniendo su calidad, se mantienen al mismo tiempo en la institución, aparecen nuevas oportunidades.



Rol del Repositorio institucional

- En tiempos pre-digitales, cuando los investigadores escribían sus resultados el editor era el único agente con la tecnología para presentar el trabajo terminado en una forma agradable, y reproducirlo para satisfacer la demanda probable de todo el mundo, en sus diarios. Los editores también gestionaban un proceso muy importante: la verificación de que la investigación era de una calidad acorde, a través del sistema de revisión por pares, y es fundamental para el avance del conocimiento, y, por tanto, a las carreras de los investigadores



Rol del Repositorio institucional

- En la era digital, las funciones de presentación y la reproducción no requieren la intermediación de un editor.
- Esto es lo que un repositorio institucional puede hacer.
- La institución guarda registro de las producciones.
- Si el trabajo total no resulta más difícil conviene mantener los registros y hacer valer los derechos de propiedad - ya sea para la institución o para los propios autores.



Rol del Repositorio institucional

- *“Los académicos se merecen un sistema de comunicación académica innata digital capaz de capturar el registro académico digital, hacerlo accesible, y preservarlo en el tiempo”*. (Van de Sompel et al., 2004).
- El análisis de Van de Sompel se funda en la preocupación por la pérdida de datos, en la necesidad de preservación y en el énfasis en el papel de la institución en el ciclo de vida de las obras.



Rol del Repositorio institucional

- *“Creemos que esta pérdida debe ser subsanada con un sistema de comunicación académica futura que incorpore de forma nativa la capacidad de grabar y exponer esa dinámica, las relaciones y las interacciones en la propia infraestructura de comunicación académica. El registro de este cuerpo de información es sinónimo de registrar la evolución de la academia en una granularidad fina. Esto permitirá rastrear el origen de las ideas específicas hasta sus raíces, analizando las tendencias en un momento específico en el tiempo y proyectando las futuras líneas de investigación”.* (Van de Sompel et al., 2004).

Rol del Repositorio institucional en OFF



- Esta nueva funcionalidad es obviamente deseable, pero se le dio poca atención en el pasado, ya que era prácticamente imposible de administrar, y no había ningún beneficio obvio en cualquier caso. Los repositorios institucionales hoy día hacen la administración relativamente simple, y los beneficios futuros se han puesto de relieve en los últimos tiempos. Estos beneficios se derivan principalmente del extraordinario potencial de las redes de repositorios que han sido posibles gracias al desarrollo del protocolo OAI-PMH.
- Herbert Van de Sompel, desarrollador del protocolo OAI-PMH, ha descrito regularmente cómo la invención tiene la capacidad de servir a los propósitos de la academia - y el público interesado - sin sacrificar ninguno de los elementos de confianza de la difusión de la investigación y el proceso de publicación.



El impacto de la investigación

- Los investigadores son recompensados por su trabajo, no económicamente, sino a través de su impacto.
- Que una investigación sea leída, entendida y consumida.
- Recibir comentarios, crédito y añadir/ampliar el conocimiento.
- Naturalmente, quieren recibir el crédito por añadir al conocimiento y buscan mejorar su entorno.



Impacto de las investigaciones

El método convencional de difusión de la investigación a través de su publicación en revistas es mucho más limitado en su posible impacto (a través de las fuerzas del mercado) que el nuevo método de publicación de la misma investigación en repositorios de acceso abierto. Numerosos estudios han demostrado que los trabajos “en abierto” se leen más en general, y, por tanto, citan con más frecuencia, que los papeles que no se encuentran en los repositorios. La consecuencia de esto es que tienen un mayor impacto.



Factor de Impacto - ISI

- El Instituto de Información Científica (ISI) hace el famoso ranking de impacto de revistas científicas desde hace muchos años, en base a la serie de índices de citas, ahora basado en la web y se conoce como *Web of Knowledge* (wok.mimas.ac.uk).
- Los factores de impacto se basan en el número medio de veces que los documentos en un título de revista dado son citados por otros papeles - una medida justa de su impacto en la investigación, aunque no sin algunas distorsiones, como ISI mismo señala en su publicación periódica que presenta la clasificación de impacto, la revista *Journal of Citation Reports*.



Recomendaciones ISI en OFF

- No debe depender únicamente de los datos de citación en la evaluación de la labor.
- Los datos de citación no están destinados a reemplazar la revisión por pares.
- Debe prestarse cuidadosa atención a las muchas condiciones que pueden influir en las tasas de citación, como historial de la revista, lenguaje y formato, fechas de publicación, y la especialidad en la materia.

online

digital.
CSIC



DIGITAL.CSIC

ACCESO ABIERTO
A DOCUMENTOS DIGITALES

DIGITAL.CSIC:

“Catalogación con Dublin Core, gestión de derechos de autor y permisos de editores. Medida, visibilidad e impacto de la producción científica”

25 de mayo 2011

<http://digital.csic.es/>

Isabel Bernal
Unidad de Coordinación de Bibliotecas del CSIC
Oficina Técnica de Digital.CSIC

El impacto de la investigación

- Medición de la actuación científica: becas y proyectos de investigación concedidos, premios, reconocimientos científicos, patentes, actividades docentes, colaboraciones nacionales e internacionales, indicadores de herramientas 2.0...
- Medición bibliométrica: diversos criterios y modelos que reflejan el impacto de las publicaciones de los investigadores

Modelos dominantes en bibliometría(1/9)

- **Journal ImpactFactor (JIF)**
 - desarrollado por el Institute of Scientific Information (ISI), ahora parte de Thomson Reuters (Web of Knowledge)
 - ediciones anuales: Science Edition, Social Sciences Edition
 - el factor de impacto de una revista es el número de veces que se cita por término medio un artículo publicado en la misma
 - **factor de impacto de revista x en 2008**: veces que las revistas ISI han citado en 2008 artículos publicados por revista x durante 2006-2007 / número de artículos publicados en revista x en 2006-2007

Modelos dominantes en bibliometría(2/9)

- **Journal ImpactFactor**
 - cálculo aplicado a unas 9.000 revistas científicas ...seleccionadas por ThomsonReuterssegún varios criterios (p.e.,la periodicidad de las revistas, contenidos, estándares de calidad etc)
 - se trata principalmente de revistas en inglés procedentes de EE.UUy Europa
 - **modelo de evaluación de revistas PERO se usa incorrectamente para evaluar la producción científica de los investigadores**

Modelos dominantes en bibliometría(3/9)

- **Críticas al Journal ImpactFactor:**

- 1) modelo viable cuando las revistas se difundían solo en formato impreso e ISI (1961) tenía la única gran base de datos referencial de información científica
- 2) deja fuera del análisis un número muy alto de revistas científicas. Según UlrichsDirectory, hay más de 300.000 revistas (+magazines, periódicos, newslettersetc..)
- 3) el 80% de las citas corresponden al 20% de los artículos...p.e: 89% del Factor de impacto de Sciencelo genera el 25% de sus artículos
- 4) no puede predecir el potencial de citas que pueden obtener los artículos a nivel individual..
- 5) ..ni siquiera predice los investigadores potenciales para ganar el premio Nobel: de los 28 físicos que han ganado el Nobelentre 2000-2009, sólo 5 están en la lista de ISI de los más citados en su campo

Modelos dominantes en bibliometría(4/9)

- **Más críticas al Journal ImpactFactor:**

- 6) El periodo de cálculo base para citas es muy corto. Los artículos clásicos son citados aún después de décadas.
- 7) La naturaleza de los resultados en distintas áreas de investigación produce distinta cantidad de publicaciones y a diferente ritmo, lo que tiene un efecto en el factor de impacto. Generalmente, por ejemplo, las publicaciones médicas tienen un factor de impacto más alto que las publicaciones matemáticas.
- 8) A pesar de las críticas, JIF es el modelo dominante, pero mal utilizado: conclusión: ¿sólo cuentan las publicaciones en revistas ISI?



Comparación de Impacto

- Esta metodología no es una forma adecuada de medir el impacto de los trabajos de investigación de acceso abierto, ya que se basa en los títulos de revistas en lugar de en documentos individuales.
- Mientras que crece el número de títulos de revista de acceso abierto (ruta dorada), la comparación directa del factor de impacto contra los títulos existentes, establecidos, casi inevitablemente, producen un menor impacto.
- Harnad y Brody al tanto de esta limitación, y sugieren una manera en la que se puede medir el impacto de los documentos depositados en repositorios de acceso abierto.



Factor de Impacto de unas y otras

- Para obtener una estimación realista de los efectos de la OA en el impacto, no es suficiente para comparar sólo el 2% de las revistas ISI que son revistas de acceso abierto, con el 98% que no lo son, para encontrar que son iguales en el impacto (de esto puede así que comparar manzanas con naranjas, incluso si se equiparan a la materia).
(Harnad y Brody, 2004)



Factor de impacto de unas y otras

¿Qué más conviene comparar?

- (1) el índice de impacto de un porcentaje mucho más alto, tal vez tan alto como el 20-40% de acuerdo a Swan y Brown (2004) de los artículos del 98% de la muestra de revistas no OA revistas, que se han hecho OA por sus autores (por autoarchivo)
- (2) frente al índice de impacto de los artículos (de las mismas revistas) que no se han hecho OA por sus autores.

Relación de impacto abierto/restringido



Sobre la base del trabajo "*En línea o Invisible*" (Lawrence, 2001), el análisis de Harnad y Brody de la literatura física para el año 2001 revelaba que la proporción de citas entre artículos de acceso abierto a los de acceso restringido variaba de 2.5:1 a 5.8 : 1.

Estudios vinculados a abierto/restringido



- (Antelman, 2004) estudió la proporción de documentos de acceso abierto y restringido en cuatro disciplinas - filosofía, ciencia política, ingeniería electrónica y eléctrica, y matemáticas.
- Matemáticas mayor proporción global de documentos disponibles en abierto (69%).
- Ciencias políticas, la menor (29%).
- Antelman: el comportamiento de los investigadores en esos campos cambiará una vez que una masa crítica de los documentos está disponible en los repositorios de acceso abierto.



ISI - 2004

Open Access Journals in the ISI Citation Databases:

Analysis of Impact Factors and Citation Patterns

A citation study from Thomson Scientific

Marie E. McVeigh

October 2004

Estudios vinculados a abierto/restringido



Jean-Claude Guédon (2002) *“En La larga sombra de Oldenburg: bibliotecarios, investigadores, editores, y el control de las publicaciones científicas”*.



Un comentario fuerte...

“If, through the manipulation of the number of articles in a given database, a publisher manages to affect the rate of use of its own articles, it also stands to reason that this publisher is able to affect the citation rate of its articles. If this situation leads just one Ohio scientist to cite one more Elsevier article in one of his/her articles, this affects the impact factor of the journal where the article appears. Of course, with one citation, the effect is too small to be detected, but imagine now that event repeated an untold number of times in Ohio and across other similarly structured consortia. It will lead to increasing the number of citations to Elsevier articles. As a result, the impact factor of Elsevier journals should begin to go up. As a consequence, these journals begin to attract more authors; but then, with a greater choice of authors, the quality should go up. In effect, a kind of quality pump has been successfully primed and it begins to propel the journal up the pecking order ladder among the core journals”. (Guédon, 2002).

Estudios vinculados a abierto/restringido



Uno puede especular que cuando los artículos están a un solo clic, los "malos" comportamientos de los autores que se han descrito en la literatura de análisis de citas serán menos comunes. Un ejemplo es el sesgo de citas, donde los autores referencian revistas que sólo ellos pueden acceder. (Antelman, 2004).

Acceso abierto “real” y “potencial”



Según los patrones de publicación de la comunidad científica argentina en el campo de la Medicina.

- Sandra Miguel, Nancy Gómez y Paola Bongiovani
- BIREDIAL. Bibliotecas y Repositorios Digitales: Gestión del conocimiento, Acceso Abierto y Visibilidad Latinoamericana. (2011)



Acceso abierto “real” y “potencial”

- Una de las primeras conclusiones de este estudio es que los resultados son reveladores de un **panorama muy alentador**, ya que el **68%** de la **producción científica argentina del campo de Medicina publicada entre 2008 y 2010** y visible internacionalmente en SCOPUS se publica en revistas que adhieren a alguna forma de acceso abierto, en una relación del **20,7%** para la vía dorada y del **47,3%** para la vía verde.

SHERPA-ROMEIO



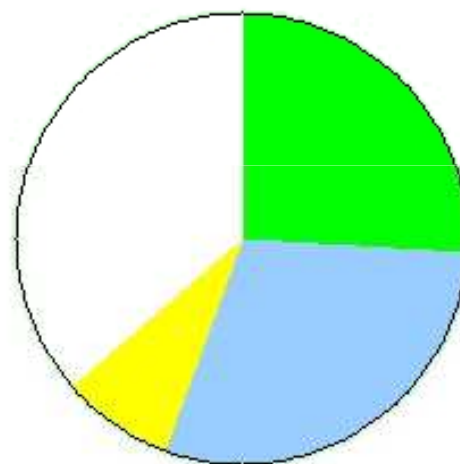
El grupo de trabajo SHERPA (UK) mantiene el **proyecto ROMEIO** (Rights Metadata for Open Archiving) que analiza los términos de copyright de las editoriales respecto al autoarchivo de los trabajos que publican. Según la política de la editorial, **ROMEIO** hace una clasificación por colores: verde (se autoriza al depósito de la versión pre- o post-print), azul (auto-archivo de la post-print, amarillo (auto-archivo de la pre-print) y blanco (la editorial no permite el depósito posterior del artículo de ninguna forma). Actualmente de la lista de 1007 editoriales registradas en su base de datos, el 64 % permiten algún tipo de auto-archivo.

<http://bit.ly/oZzauo>

SHERPA-ROMEO



SHERPA/RoMEO Colours

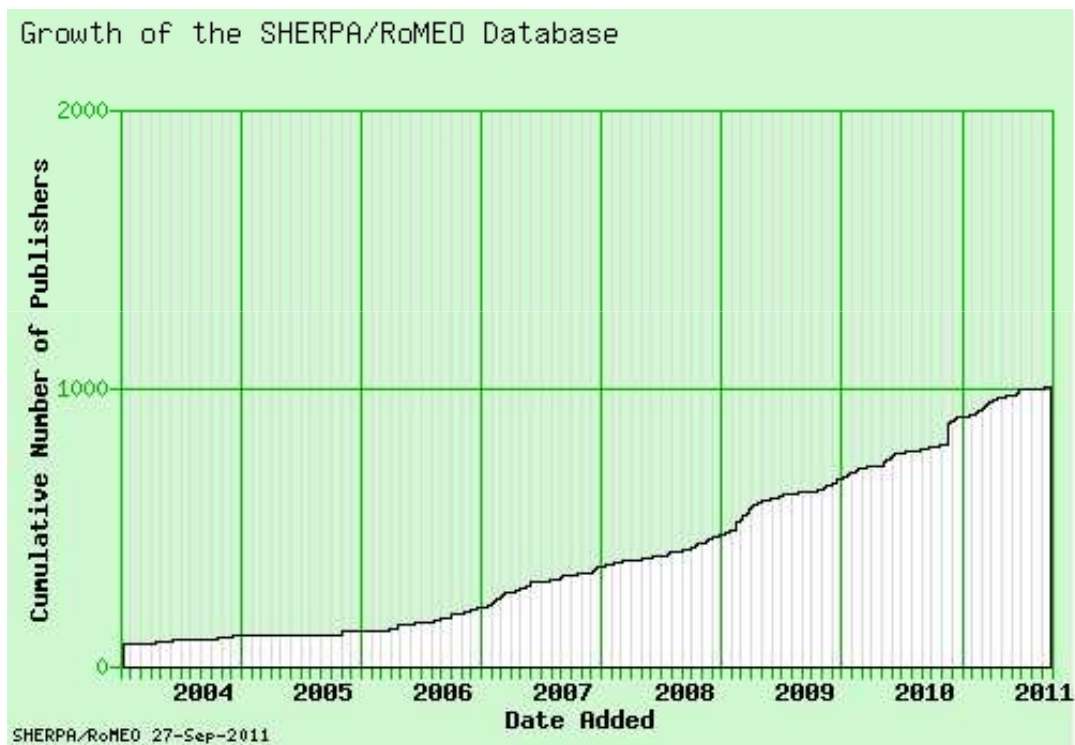


- Green (263 = 26%)
- Blue (297 = 29%)
- Yellow (82 = 8%)
- White (365 = 36%)

SHERPA/RoMEO 27-Sep-2011

Total = 1007 publishers

SHERPA-ROMEO





Webometrics

Es un proyecto que pertenece al Laboratorio Cybermetrics de España, gestiona un ranking de repositorios que pone a disposición dos ediciones anuales a partir del 2008. Según Aguillo, el ranking inicialmente realizaba sus cálculos clasificados en un 50% de peso para el número de páginas, archivos PDF e ítems que se encuentran en Google Scholar y el otro 50% que toma en cuenta la visibilidad a partir de los enlaces externos recibidos por el repositorio

Indicadores Webometrics hasta 1/2012



1. **Size** con un peso de 20%. Representa los enlaces web (links) del repositorio en los buscadores Google, Yahoo SE (Site Explorer) y Bing.
2. **Visibility** con un peso de 50%. Son los diferentes links externos que hacen referencia al repositorio, son conocidos como inlinks, en Yahoo SE.
3. **Rich Files** con un peso de 15%. Son los archivos pdf, doc, docx, ppt, pptx, ps y eps según en el buscador de Google.
4. **Google Scholar** con un peso de 15% representando el trabajo científico presente en el repositorio.

Ref.: Los Indicadores de Webometrics en SeDiCI según las Normas ISO 9126 y 25010

Autor: Jose Daniel Texier R.



Indicadores Webometrics desde 7/2012

- 1. Size (S)** - 10%. Número de páginas web extraídas de Google.
- 2. Visibility (V)** - 50%. Son los diferentes links externos recibidos por el número de dominios referidos que se obtienen de MajesticSEO y ahrefs.
- 3. Rich Files (R)** - 10% Son los archivos pdf, doc, docx, ppt, pptx, ps y eps según en el buscador de Google.
- 4. Scholar I (Sc-1)** - 15%. Google Scholar: normalización de todos los paperes.
Scholar II (Sc-2) - 15%. Google Scholar: normalización de todos los paperes entre 2007 y 2011.



Indicador Scholar SeDiCI 7/2012



site:sedici.unlp.edu.ar



Académico

Aproximadamente 6.280 resultados (0,06 s)

Cualquier momento

Desde 2012

Desde 2011

Desde 2008

Intervalo
específico...

[PDF] [CUADERNOS DE HERFETOLOGIA VOLÚMEN DICIEMBRE 198-5 ASOCIACION
HERPETOLOGICA ARGENTINA](#)

RF Laurent - 1985 - [sedici.unlp.edu.ar](#)

The typical subgenus *Liolaemus sensu skicto* includes most chilean species a few argentinian and some bolivian and peruvian forms. An argentinian group which includes most argentinian, bolivian and peruvian species (several undescribed), can be named ...

Citado por 44 [Versión en HTML](#)

Ordenar por
relevancia

Ordenar por fecha

[CITAS] [Arqueología del área interserrana bonaerense](#)

GG Politis - 1984 - [sedici.unlp.edu.ar](#)

Citado por 44 [En caché](#)

Buscar en la Web

Buscar sólo
páginas en
español

[CITAS] [Los roedores muroideos del Pleistoceno tardío-Holoceno en la región pampeana \(sector e Patagonia \(República Argentina\)\)](#)

UFJ Pardiñas - 1999 - [sedici.unlp.edu.ar](#)

SeDiCI - Repositorio de la Universidad Nacional de La Plata Los roedores muroideos del Pleistoceno tardío-Holoceno en la región pampeana (sector este)



Indicador Scholar SeDiCI 7/2012

Google

Académico

Cualquier momento
Desde 2012
Desde 2011
Desde 2008
Intervalo específico...

Ordenar por relevancia
Ordenar por fecha

Buscar en la Web
Buscar sólo páginas en español

Buscar artículos
con todas las palabras

con la frase exacta

con al menos una de las palabras

sin las palabras

donde las palabras aparezcan **en todo el artículo** ▾

Mostrar artículos escritos por
p. ej., "García Márquez" o Cela

Mostrar artículos publicados en
p. ej., JAMA o Gaceta Sanitaria

Mostrar artículos fechados entre —
p. ej., 1996



Indicadores Webometrics

La clasificación o ranking Webometrics se calcula a partir del máximo valor para el size, visibility, rich files y Google Scholar de todos los repositorios. El trabajo medirá la calidad de SeDiCI a partir de esos 4 valores establecidos por Webometrics. Para observar la mejora se debe establecer un punto de comparación que permita medir la calidad.

Ref.: Los Indicadores de Webometrics en SeDiCI según las Normas ISO 9126 y 25010

Autor: Jose Daniel Texier R.



SeDiCI SERVICIO DE DIFUSIÓN
DE LA CREACIÓN INTELECTUAL



Bibliotecas y Repositorios Digitales

Tecnología y Aplicaciones

<http://sedici.unlp.edu.ar>



Participantes del dictado

- Marisa De Giusti
- Nestor Oviedo
- Silvia Pelоче
- Matías Cánepa



Objetivo del curso

- Compartir la experiencia del SeDiCi en todas las áreas que hacen al quehacer del repositorio: edición, catalogación, comunicación y difusión, software de soporte e interoperabilidad, servicios asociados y cuestiones legales, entre otras. Cada área temática será dictada por distintos integrantes del SeDiCi, de acuerdo al perfil de sus tareas.
- Crear conciencia sobre el acceso abierto en todas sus formas.

Bibliotecas y repositorios digitales



Capítulos 2 y 3: ¿Cómo se establece un repositorio institucional?. Marcando los aspectos importantes de su definición. Análisis previo. Reflexiones institucionales importantes. Determinación de políticas. El ámbito de la UNLP. Costos del establecimiento. Interdisciplinaridad. Continuando con los conceptos básicos. Tipos de objetos. Metadatos. Catalogación. Preservación. Interoperabilidad. Servicios de una Biblioteca Digital.

¿Qué es un repositorio institucional?



Un **Repositorio Institucional** es una base de **datos compuesta de** un grupo de servicios destinados a capturar, almacenar, ordenar, preservar y redistribuir la documentación académica de la Universidad en formato digital.



¿Qué es un repositorio institucional?

La organización SPARC define los Repositorios Institucionales como sigue:

- Pertenecen a una institución.
- Son de ámbito académico.
- Son acumulativos y perpetuos.
- Son abiertos e interactivos \rightleftarrows que cumplen con el OAI y permiten acceso abierto a la documentación académica.



¿Qué es un repositorio institucional?

- Clifford Lynch define los RIs de la siguiente forma: *“Un Repositorio Institucional universitario es un conjunto de servicios que ofrece la Universidad a los miembros de su comunidad para la dirección y distribución de materiales digitales creados por la institución y los miembros de esa comunidad. Es esencial un compromiso organizativo para la administración de estos materiales digitales, incluyendo la preservación a largo plazo cuando sea necesario, así como la organización y acceso o su distribución”*.

Clifford A. Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age" ARL, no. 226 (February 2003): 1-7



Crecimiento y difusión de los RIs

Incremento actual

- En los últimos años ha habido un movimiento importante que ha llevado a las instituciones a considerar la posibilidad de crear repositorios.
- Informe del Comité de Tecnología y Ciencia del Congreso, publicado a principio de año, el cual recomienda que “todas las instituciones de educación superior usen Repositorios Institucionales donde se pueda almacenar lo que se publique y donde se pueda leer de forma gratuita en línea”.



Crecimiento y difusión de los RIs

Boletín oficial del estado español BOE número 35
Sección 1 página 13919. 10 de febrero de 2011:

Artículo 14: Evaluación de la tesis doctoral, punto 5:

“una vez aprobada la tesis doctoral, la universidad se ocupará de su archivo en formato electrónico abierto en un repositorio institucional y remitirá, en formato electrónico, un ejemplar de la misma así como toda la información complementaria que fuera necesaria al Ministerio de Educación a los efectos oportunos”.



Crecimiento y difusión de los RIs

- El MINCyT ha propiciado la creación de un Sistema Nacional de Repositorios Digitales en CyT.
- Convocó a representantes de los organismos e instituciones que componen el Sistema
- Nacional de Ciencia, Tecnología e Innovación (SNCTI) que cuentan con repositorios para el establecimiento de políticas conjuntas en aspectos clave.
- Este Sistema Nacional tiene por objeto impulsar, gestionar y coordinar una red interoperable de repositorios distribuidos físicamente, creados y gestionados por instituciones o grupos de instituciones a nivel nacional para aumentar la visibilidad e impacto de la producción científica y tecnológica de Argentina.



Crecimiento y difusión de los RIs

Como cada vez hay más cantidad de investigaciones y material académico en forma digital, recuperar y conservar este material tiene múltiples finalidades:

- permitir a los autores distribuir sus artículos de investigación gratis vía Internet, y
- asegurar la conservación de los mismos en un entorno electrónico y dinámico.

¿Cómo usa la gente los Repositorios Institucionales?



Las Universidades utilizan los Repositorios Institucionales para:

- Comunicación académica.
- Conservación de materiales de aprendizaje y de cursos.
- Publicaciones electrónicas.
- Organización de las colecciones de documentos de investigación.
- Conservación de materiales digitales a largo plazo.

¿Cómo usa la gente los Repositorios Institucionales?



- Aumento del prestigio de la Universidad exponiendo sus investigaciones académicas.
- Relevancia institucional del papel de la biblioteca.
- Conocimiento sobre la dirección.
- Evaluación sobre la investigación.
- Animación a la creación de un acceso abierto a la investigación académica.
- Conservación de colecciones digitalizadas.



Modelo de RI

- Cada Universidad tiene su propia idiosincrasia, así como valores únicos que requieren un método personalizado. El modelo de información que mejor se ajuste a una Universidad no será el adecuado para todas. Hay que pensar el modelo en función de las necesidades propias. De esto trata planear el repositorio.

Cómo proyectar un repositorio institucional



Cuestiones generales

- Aprendizaje sobre el proceso en base a otros Ris
- Definición y Plan de Servicio
 - Evaluación de las necesidades de la universidad
 - Modelo de costo
 - Planificación y horario
 - Políticas para recopilar contenidos...
- Formación de equipo
- Tecnología
- Marketing
- Difusión del servicio
- Puesta en funcionamiento

Esfera institucional

- Marco legal
- Costos

Principales actuaciones para crear un Repositorio Institucional



- Aprender examinando otros RIs.
- Desarrollar una definición y un plan de servicio:
 - Realizar una evaluación de las necesidades de la Universidad.
 - Desarrollar un modelo de coste basado en este plan.
 - Crear una planificación y un horario.
 - Desarrollar políticas de actuación que gestionen la recopilación de contenidos, su distribución y mantenimiento.



Principales actuaciones para crear un Repositorio Institucional

- Formar el equipo.
- Realizar la planificación
- Tecnología: elegir e instalar el software.
- Marketing.
- Difundir el servicio.
- Puesta en funcionamiento del mismo.



Compromiso

Observaciones: la elección de la tecnología debe reflejar los requisitos especificados en el capítulo de la planificación de servicio.



Retos más comunes

- Grado de acogida por parte del personal académico.
- Proporcionar sustancialidad.
- Desarrollar un plan de actuación.
- Gestionar los derechos de propiedad intelectual.
- Apoyo universitario.
- Control de costes.
- Conservación digital.
- Identificar a las principales partes interesadas.



Cómo empezar

El diseño del servicio es básico para posteriores elecciones sobre presupuesto y tecnología.

Si ya se ha empezado a crear un repositorio...

Se puede comparar la planificación actual con el material del capítulo para comprobar si ya hay pasos o decisiones adicionales a considerar antes de proceder.

Referencias clave: nada es nuevo bajo el sol...



- Durante los dos últimos años ha aumentado la información publicada sobre Ris y está toda en abierto...
- ***Es importante consultar Repositorios Institucionales en funcionamiento:*** a veces es útil ver cómo otras Universidades e instituciones culturales han organizado o presentan sus repositorios.
- Una práctica muy buena es mirar los repositorios mejor rankeados en ROAR, atendiendo la clase de repositorio que se desea armar e inspeccionando para ver la tipología de los materiales que ofrecen.
- Expertos: <http://www.rsp.ac.uk/>



Primera fase: planificación del servicio

- El material de planificación del servicio está dirigido a personal bibliotecario universitario y personal directivo que está creando el RI.
- También serán interesantes, para personal directivo financiero y de personal, las secciones sobre la evaluación de los recursos existentes y las habilidades requeridas para el personal que debe mantener el Repositorio Institucional.



Desarrollando su modelo de servicio

- Es importante definir de forma precisa cómo se usará el sistema y qué tipo de servicios ofrecerá. Por ejemplo, algunas Universidades crean su Repositorio Institucional para guardar solamente investigaciones académicas. Otras amplían la definición de servicio para incluir tesis de alumnos, materiales docentes o informes universitarios.
- Es preferible decidir esto antes de crear la infraestructura técnica de un Repositorio Institucional.



¿Cómo definir el servicio?

- Un Repositorio Institucional no se define únicamente por el software y la base de datos que contiene sus colecciones digitales. Es un conjunto de servicios para aquellos que almacenan contenidos, tanto las comunidades académicas y de investigación que abarca como los usuarios (de distintos tipos).
- Desarrollar una **definición de servicio para el RI** significa especificar lo que ofrecerá a sus usuarios. Los formatos, los servicios, el papel de los administradores con las comunidades que depositarán sus contenidos y la planificación del desarrollo del servicio.

Preguntas a tener en cuenta para crear una definición de servicio



- ¿Qué tipo de contenidos aceptarán?,
- ¿Quién puede depositar contenidos en el repositorio?,
- ¿Quién proporcionará los metadatos?.

Por ejemplo, algunos Repositorios Institucionales aceptan solamente el material revisado por pares, otros aceptan materiales pre-publicados. Otras incluyen objetos de aprendizaje, materiales educativos y otros artículos que no se encuentran en texto impreso.

Constantemente ver en ROAR o en OpenDOAR.



Cómo crear una definición de servicio

- ¿Cuál es la misión del servicio?
- ¿Qué tipo de contenidos aceptará?
- ¿Quiénes son los usuarios principales?
- ¿Quiénes son las principales partes interesadas?
- ¿Qué servicios ofrecería si tuviera recursos ilimitados?
- ¿Qué puede permitirse ofrecer?
- ¿Cobrará por los servicios?



Cómo crear una definición de servicio

- ¿Qué responsabilidades tendrá la administración del RI vs. la comunidad de contenidos?
- ¿Cuáles son sus principales prioridades de servicio?
- ¿Cuáles son sus prioridades a corto y a largo plazo?

Servicios gratuitos vs. Servicios de cobro



La siguiente tabla muestra la variedad de servicios que un equipo de biblioteca puede ofrecer en un Repositorio Institucional.



Servicio de Repositorios Institucionales

Servicios generales (gratuitos)	Servicios premium (prepago)
Establecer los departamentos académicos y las comunidades de contenidos en el Repositorio Institucional	
Servicio de metadatos Consulta de servicios de Metadatos.	Servicio de metadatos Creación de Metadatos de Clientes
Apoyo a los usuarios suministradores de contenidos	Servicios documentales - Escaneado - OCR - Formato de archivos
Asignación de espacio de archivo: básico	Asignación de espacio de archivo: extra
Importación de datos por lotes (batch) - Colecciones históricas - Colecciones digitalizadas recientemente	
Gestión de Sistemas Informáticos	
Otros	

Fuente University of Rochester

Servicios gratuitos vs. Servicios de cobro



- En el ejemplo precedente extraído de la Universidad de Rochester, algunos servicios están disponibles de forma gratuita y otros previo pago. Cada servicio puede ofrecer algunos o la mayoría de éstos dependiendo de cómo se estructure el Repositorio Institucional y las necesidades de la comunidad.
- El cobro por los servicios de un Repositorio Institucional específico depende de su previsión de coste. En el ejemplo que se muestra, se cobra a las comunidades de contenido por contratar servicios extra con la biblioteca. La Universidad de Rochester ofrece una variedad de servicios centrales o gratuitos, junto con servicios “premium” para recuperar costos.



Política de servicios de Digital CSIC

Digital.CSIC pone a disposición de su comunidad de usuarios perteneciente al CSIC (investigadores, autores y bibliotecarios) una serie de servicios para promover un uso fácil y eficiente del repositorio.

La Oficina Técnica de **Digital.CSIC**:

- Define la misión, objetivos, servicios y desarrollo del repositorio.
- Coordina las iniciativas del repositorio.
- Ofrece servicios de consultas y apoyo técnico sobre el funcionamiento cotidiano del repositorio.



Política de servicios de Digital CSIC

- Organiza talleres de divulgación, formación y promoción sobre el funcionamiento del repositorio del CSIC y variadas temáticas relativas al acceso abierto. La Oficina Técnica también tendrá en cuenta las peticiones procedentes de los centros y bibliotecas del CSIC para la organización de sesiones sobre temáticas específicas
- Crea recursos educativos, estudios, informes, manuales, material divulgativo, estrategias de comunicación y promoción
- Fomenta el intercambio de conocimientos e información mediante nuevos canales de comunicación
- Asesora en cuestiones de derechos de autor
- Diseña e implementa la política de preservación digital
- Incorpora nuevas funcionalidades en el software del repositorio, ofreciendo así más y mejores servicios
- Esta política de servicios es revisada por la Oficina Técnica de **Digital.CSIC**.



Formas de organizar el contenido

- Cada servicio de Repositorio Institucional organiza su contenido de la manera que mejor se ajuste a la particularidad de la Universidad y de su estructura académica. Muchas Universidades se organizan según los centros o departamentos de investigación.
- Éste no es el único principio organizativo.

Formas de Organizar el Contenido: Kansas: “comunidades híbridas”



1. **Comunidades Formales:** constan de departamentos, centros de investigación y grupos ya existentes, con directrices de aportación de contenido establecidas y volumen de trabajo. Ejemplo: Departamento de Neurociencia.
2. **Comunidad por Áreas Temáticas:** acceso abierto, todo el personal académico puede aportar contenidos, o a través de terceros. El personal de la biblioteca revisa el contenido antes de colgarlo en la red. Ejemplo: Policy Research Institute.
3. **Comunidades de Interés:** Un grupo ad hoc, de departamentos inter- facultades. La organización de las comunidades de contenido depende de los intereses y acuerdos de sus integrantes.

<http://kuscholarworks.ku.edu/dspace/community-list>

[\[Buscar 🔍 \]](#)[\[Búsqueda_avanzada 🔍 \]](#)

SERVICIOS

[Login](#)[Mis suscripciones](#)

LISTAR

[Comunidades](#)[Títulos](#)[Autores](#)[Materias](#)[Fecha Publicación](#)[ESTADÍSTICAS](#)[Políticas Digital.CSIC](#)[Revista CSIC
ABIERTO](#)[Memoria Digital.CSIC](#)[Envía tus trabajos](#)

RECURSOS

🏠 Bienvenidos al Repositorio Institucional del CSIC

[VER PRESENTACIÓN](#)

Bienvenidos a **Digital.CSIC**, el repositorio institucional del **Consejo Superior de Investigaciones Científicas**.

Digital.CSIC es un depósito de documentos digitales, cuyo objetivo es organizar, archivar, preservar y difundir en modo de acceso abierto la producción intelectual resultante de la actividad investigadora del CSIC.

I Comunidades en Digital.CSIC

- [+] Biología y Biomedicina [3740]
- [+] Ciencia y Tecnología de Alimentos [1105]
- [+] Ciencia y Tecnología de Materiales [2803]
- [+] Ciencia y Tecnologías Físicas [5729]
- [+] Ciencia y Tecnologías Químicas [3919]
- [+] Ciencias Agrarias [5672]
- [+] Humanidades y Ciencias Sociales [9150]
- [+] Recursos Naturales [7395]
- [+] Servicios Centrales CSIC [244]

I Digital.CSIC en cifras

- ★ Número de registros depositados: 37.309
- ★ Visualizaciones de registros: 5.278.656
- ★ Descargas de registros: 6.287.449

[CRECIMIENTO LINEAL](#)[EVOLUCIÓN MENSUAL](#)



Open access to 705,282 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics

Subject search and browse:

29 Jul 2011: [arXiv Sustainability Initiative Update](#): An update on the effort to design a new collaborative support model for arXiv, including a short FAQ, is provided in the [July 2011 update](#).

25 Apr 2011: [Data sets accepted as part of Data Conservancy pilot](#); [ScienceWISE provides article annotation and bookmarking](#)

See cumulative "What's New" pages. Read [robots beware](#) before attempting any automated download

Physics

- **Astrophysics** ([astro-ph new](#), [recent](#), [find](#))
includes: [Cosmology and Extragalactic Astrophysics](#); [Earth and Planetary Astrophysics](#); [Galaxy Astrophysics](#); [High Energy Astrophysical Phenomena](#); [Instrumentation and Methods for Astrophysics](#); [Solar and Stellar Astrophysics](#)
- **Condensed Matter** ([cond-mat new](#), [recent](#), [find](#))
includes: [Disordered Systems and Neural Networks](#); [Materials Science](#); [Mesoscale and Nanoscale Physics](#); [Other Condensed Matter](#); [Quantum Gases](#); [Soft Condensed Matter](#); [Statistical Mechanics](#); [Strongly Correlated Electrons](#); [Superconductivity](#)
- **General Relativity and Quantum Cosmology** ([gr-qc new](#), [recent](#), [find](#))
- **High Energy Physics - Experiment** ([hep-ex new](#), [recent](#), [find](#))
- **High Energy Physics - Lattice** ([hep-lat new](#), [recent](#), [find](#))
- **High Energy Physics - Phenomenology** ([hep-ph new](#), [recent](#), [find](#))
- **High Energy Physics - Theory** ([hep-th new](#), [recent](#), [find](#))
- **Mathematical Physics** ([math-ph new](#), [recent](#), [find](#))
- **Nuclear Experiment** ([nucl-ex new](#), [recent](#), [find](#))
- **Nuclear Theory** ([nucl-th new](#), [recent](#), [find](#))
- **Physics** ([physics new](#), [recent](#), [find](#))
includes: [Accelerator Physics](#); [Atmospheric and Oceanic Physics](#); [Atomic Physics](#); [Atomic and Molecular Clusters](#); [Biological Physics](#); [Chemical Physics](#); [Classical Physics](#); [Computational Physics](#); [Data Analysis, Statistics and Probability](#); [Fluid Dynamics](#); [General Physics](#); [Geophysics](#); [History and Philosophy of Physics](#); [Instrumentation and Detectors](#); [Medical Physics](#); [Optics](#); [Physics Education](#); [Physics and Society](#); [Plasma Physics](#); [Popular Physics](#); [Space Physics](#)
- **Quantum Physics** ([quant-ph new](#), [recent](#), [find](#))

Mathematics

- **Mathematics** ([math new](#), [recent](#), [find](#))



Communities

Welcome to Europeana's community space.

From these pages, we link you to useful online resources that can help you explore, learn and share more about Europe's cultural and scientific heritage. They include photo sharing sites, forums and innovative digital projects.

The websites and social media groups featured here are organised around themes and categories that relate to the content on Europeana. The pictures on each page highlight some of the items in Europeana that are associated with each topic. Descriptions of the sites we link to are in English.

View Communities

Art Nouveau - [visit](#)

Biodiversity - [visit](#)

Maps and Cartography - [visit](#)

Medieval/Renaissance - [visit](#)

Do you have any suggestions for great communities online, or themes to explore? We'd love to hear your ideas, and may feature them on these pages. Please email us here: info@europeana.eu

Europeana is not responsible for the content of external websites. Please read our [External link disclaimer](#).



The Southampton Quadrilles - Provided by The British Library



DIGITAL COLLECTIONS



- [Digital Collections Home](#)
- [About Digital Collections](#)
- [Digital Reference](#)
- [Digital Library Standards](#)
- [Digital Preservation](#)

More Resources

- [Bibliographies and Guides](#)
- [Finding Aids](#)
- [Virtual Reference Shelf](#)
- [Ask a Librarian](#)
- [Library Catalogs](#)

Digital Collections & Services

Access to print, pictorial and audio-visual collections and other digital services

|| Featured Digital Collections & Services ||



American History & Culture

A digital library of historic maps, photos, documents, audio and video from **American Memory**.



Historic Newspapers

Enhanced access to America's historic newspapers through the **Chronicling America** project.



International Collections

Materials and bilingual presentations from overseas libraries from **Global Gateway**.



Legislative Information

THOMAS provides searchable full text of bills, Congressional Record and more.



Browse By Topic

Browse selected collection content available online by **Topic**.



Performing Arts

Collections, articles and special presentations on music, theater and dance materials.



Prints and Photographs

Catalog of about half of the Library's pictorial holdings with over 1 million digital images.



Veterans History

Experience first-person stories of wartime service through personal artifacts.



Early Sound Recordings

The National Jukebox features over 10,000 78rpm disc sides issued by the Victor, 1900-25.



Web Site Archiving

Searchable archives of historic Web pages including national elections, 9/11 and more.

|| Additional Digital Collections & Services ||

- [Children and Families: America's Library](#)
- [Country Studies/Area Handbooks](#)
- [Digital Interlibrary Loan](#)
- [Exhibitions from the Library](#)
- [Local Legacies: Community Roots](#)
- [Manuscript Collections](#)
- [Map Collections](#)
- [Poetry 180: For the School Year](#)
- [Rare Books and Special Collections](#)





Home



ePrints Soton

Policies

Latest Additions

Browse by Year

Browse by Subject

Browse by School

EPrints Search



Login

[RSS 1.0](#)

[RSS 2.0](#)

[Atom](#)

[Home](#) > [Research](#) > [EPrints](#)

Welcome to ePrints Soton

Welcome to the University of Southampton Institutional Research Repository ePrints Soton.

[Atom](#) [RSS 1.0](#) [RSS 2.0](#)

[Repository Policies](#)

University of Southampton policies regarding the eprints soton research repository.

[Latest Additions](#)

View items added to the repository in the past week.

[Search Repository](#)

Search the repository using a full range of fields. Use the search field at the top of the page for a quick search.

[Browse Repository](#)

Browse the items in the repository by subject.



Buscar

Participantes

Acceso a 3,111,451 objetos digitales de 153 repositorios.

Proyectos europeos

- Europeana
- Europeana local
- Consejo de proveedores de contenido y agregadores

Últimas contribuciones

- CEU Repositorio Institucional. Conservación - Fundación Universitaria San Pablo-CEU
- Prisma social: revista de ciencias sociales - Fundación IS+D para la Investigación Social Avanzada (IS+D)
- PoQ - CSIC

Directorio de colecciones digitales

[Relación de los proyectos que se están realizando en España](#)

Vídeo de presentación

[Canal del Ministerio de Cultura](#)

© Ministerio de Cultura
Aviso Legal || Accesibilidad



SeDiCI SERVICIO DE DIFUSIÓN DE LA CREACIÓN INTELECTUAL



<http://sedici.unlp.edu.ar>



US National Library of Medicine
National Institutes of Health

PMC

Limits Advanced Journal List

Search

Help



PMC

PMC is a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM).

Get Started

[PMC Overview](#)

[Users' Guide](#)

[Journal List](#)

[PMC FAQs](#)

[PMC Copyright Notice](#)

Participate

[Add a Journal to PMC](#)

[Participation Agreements](#)

[File Submission Specifications](#)

[File Validation Tools](#)

Keep Up to Date

[New in PMC](#)

[PMC News Mail List](#)

[PMC News RSS](#)

Other Resources

[PMC International](#)

[Open Access Subset](#)

[E-utilities](#)

[NLM LitArch](#)

[PMC Citation Search](#)

Number of journals that deposit

ALL articles	NIH-funded articles	Selected articles
896	296	1534

Number of articles archived in PMC

2.2 MILLION

NIH Public Access

[NIH Public Access and PMC](#)

[NIH Manuscript Submission System](#)

[My Bibliography](#)

[PMCID/PMID/NIHMSID Converter](#)



Humanities Text Initiative

The Humanities Text Initiative, a unit of the University of Michigan's [Digital Library Production Service](#), has provided online access to full text resources since 1994. The Humanities Text Initiative (HTI) is an umbrella organization for the creation, delivery, and maintenance of electronic texts, as well as a mechanism for furthering the library community's capabilities in the area of online text.

- ♦ *text collections*
- ♦ *Making of America*
- ♦ *sgml resources*
- ♦ *about HTI*

The collections on this site are freely available to the Internet community. Resources which are restricted to use by University faculty, staff, and students only can be found at the [DLPS List of all Collections](#) page.

This site is made possible in part by a generous equipment grant from Sun Microsystems Inc.

The Humanities Text Initiative site is maintained at <http://www.hti.umich.edu/>.
For further information or to give feedback please contact hti-info@umich.edu.



[INICIO](#)

[ABOUT NDLTD](#)

[NEWS](#)

[EVENTS AND AWARDS](#)

[RESOURCES](#)

[JOIN NDLTD](#)

[JOIN THE NDLTD EMAIL LIST](#)

[CONTACT NDLTD](#)

[accesibilidad](#)
[Entrar](#)



Welcome to the Networked Digital Library of Theses and Dissertations (**NDLTD**), an international organization dedicated to promoting the adoption, creation, use, dissemination, and preservation of electronic theses and dissertations (ETDs). We support electronic publishing and open access to scholarship in order to enhance the sharing of knowledge worldwide. Our website includes resources for university administrators, librarians, faculty, students, and the general public. Topics include how to find, create, and preserve ETDs; how to set up an ETD program; legal and technical questions; and the latest news and research in the ETD community.

[for Researchers:
Find ETDs](#)

[for Authors:
Submit ETDs](#)

[for Institutions:
Manage ETDs](#)

subnavigation

Últimas noticias

- ▶ [Blog Pays Tribute to ETD 2011 Conference](#) 13/09/2011
 - ▶ [Report on the 2011 Inaugural United States Electronic Theses and Dissertations Association \(USETDA\) Conference](#) 18/07/2011
 - ▶ [NDLTD ETD Awards 2011 Winners Announced](#) 08/06/2011
 - ▶ [USETDA 2011 Conference a Great Success](#) 26/05/2011
 - ▶ [Brazilian Professor Visits WVU to Gain ETD Expertise](#) 12/05/2011
- [Más noticias...](#)

Próximos eventos

- ▶ [2011 Seminario Internacional de Repositorios y Tesis Electrónicas / 2011 International Seminar on Electronic Theses and Repositories](#) Lima, Peru, 20/10/2011
 - ▶ [ETD 2012 - 15th International Symposium on Electronic Theses and Dissertations](#) Lima, Peru, 12/09/2012
- [Eventos anteriores...](#)

Website development funded by UNESCO, Adobe Systems Inc. and NDLTD members.

RePEc

[How you can use RePEc](#) | [RePEc information for participants](#) | [Major participants and activity](#)

RePEc (**R**esearch **P**apers in **E**conomics) is a collaborative effort of hundreds of volunteers in [75 countries](#) to enhance the dissemination of research in economics. The heart of the project is a decentralized database of working papers, journal articles and software components. All RePEc material is freely available. Participation in RePEc as a provider only involves the cost of your time in preparing and maintaining metadata describing your publications.

You may add your own materials to RePEc through a department or institutional archive. All institutions, nonprofit and commercial, are welcome to join and [contribute their materials](#) by establishing and maintaining their own RePEc archive. If your institution does not yet participate in RePEc, you may submit your own papers to [MPRA](#) (the Munich Personal RePEc Archive), and they will automatically be included in RePEc. RePEc does not support personal archives: only institutional archives.

RePEc collaborates with the American Economic Association's [EconLit](#) database to provide content from leading universities' working paper series to [EconLit](#). If your university does not contribute its working paper series to RePEc, please contact us for assistance, or view the ["step by step"](#) instructions at IDEAS.

Please note that RePEc does **not** contain full-text journal articles. RePEc services provide links to many full-text articles, but you may need a personal or institutional subscription to follow those links. **If a working paper or journal article is not indicated as "downloadable", please contact the author, publisher or your local library for assistance.** We cannot provide full texts on request.

The RePEc database holds over **1,085,000** items of interest, over **955,000** of which are available online:

410,000 [working papers](#)
645,000 [journal articles](#)
2,200 [software components](#)
27,000 [book and chapter listings](#)
29,000 [author contact and publication listings](#)
12,000 [institutional contact listings](#)

Bookmark this page to easily locate our services to the economics profession.

How you can use RePEc:

The following web sites offer all or part of the RePEc database for you to browse or search:



[RePEc Author Service](#): Author registration and "RePEc CV" maintenance.



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



Alma Mater Digital Library
A cura dell'Area Sistemi Dipartimentali e Documentali



[AlmaDL Home](#)

[AlmaDL Search](#)

Collezioni

[Servizi](#)

Collezione AMS Campus - AlmaDL - Università di Bologna

[Home](#) | [Cerca](#)

[Login](#)

Benvenuto su AMS Campus - AlmaDL - Università di Bologna

AMS Campus è il deposito istituzionale dei materiali didattici dell'Università di Bologna. Consente di accedere e consultare in modo agevole ed efficace i materiali didattici preparati dai docenti dell'Ateneo nell'ambito dei propri corsi.

Per utilizzare il servizio e consultare i materiali didattici nel rispetto delle disposizioni vigenti in materia di diritto d'autore si leggano le [Condizioni del servizio](#).

Attualmente sono disponibili **36478** materiali didattici depositati da **1758** Docenti dell'Ateneo. Il numero totale di Docenti registrati è **3043**. Sono stati effettuati **5771143** download di materiali didattici da questo servizio.

[Atom](#) [RSS 1.0](#) [RSS 2.0](#)

[Cerca tra i materiali depositati](#)

Usa questa interfaccia per effettuare una ricerca in AMS Campus.

[Ultimi materiali didattici depositati](#)

Visualizza l'elenco dei materiali che sono stati depositati di recente.

[Aiuto & Documentazione](#) | [Help desk](#)



AMS Campus 1.1.1 © ALMA MATER STUDIORUM - Università di Bologna, 2007-2011. Tutti i diritti riservati.
Servizio implementato e gestito da [AlmaDL - Area Sistemi Dipartimentali e Documentali](#)
[Informativa sulla privacy](#) - [Condizioni del servizio](#)

[inizio pagina](#)

SeDiCI SERVICIO DE DIFUSIÓN
DE LA CREACIÓN INTELECTUAL



<http://sedici.unlp.edu.ar>

Organización anterior de SeDiCI (hasta enero de 2012)



- [Tipos de Documentos](#)
- [Temario](#)
- [Tesis](#)
- [Repositorios](#)
- [Unidad](#)



Organización actual de SeDiCI

Tesis



Tesis de grado, post-
grado y otros
documentos

Revistas



Publicaciones en
revistas científicas

Eventos



Ponencias realizadas
en congresos y
conferencias

Libros



Libros digitalizados y e-
books

Red UNCI



Artículos y ponencias de
la Red UNCI



Cómo formar un equipo

Equipo de planificación: en primer lugar habrá que elegir al personal responsable de cada una de las siguientes tareas:

- Encuestas o entrevistas académicas y de personal.
- Evaluación de necesidades.
- Modelo de servicio.
- Modelo de costos.
- Evaluación de recursos.
- Llevar a cabo un análisis gap.
- Documentos de requisitos (soft y hard).
- Presentaciones para personal académico y usuarios potenciales.

Cómo realizar una encuesta de evaluación de necesidades



- Una pieza clave para el plan de servicio es conocer las necesidades específicas de la comunidad académica.
¿Cómo se publican y almacenan en la actualidad los trabajos académicos del campus?,
- ¿quiénes son las principales partes interesadas que se deben incluir en la planificación?,
- ¿qué recursos informáticos están disponibles en el campus?,
- ¿cuál es la percepción académica de los problemas a la hora de dirigir materiales digitales?...



Una evaluación de necesidades incluye encuestas a la comunidad.

Encuestas informales: se incluyen encuentros presenciales con administradores y profesores de forma individual, contactos por email y supervisión de los servicios de publicación en línea existentes en el campus.

Encuestas formales: pueden incluir encuestas al personal académico bien en papel o en línea, así como presentaciones y sesiones de Q&A con departamentos y grupos académicos.

➤ Lo importante: saber cómo publican los docentes e investigadores en la actualidad sus investigaciones y como podrían usar un Repositorio Institucional universitario.



Cómo crear un plan de servicio

- Un típico plan de servicio de Repositorio Institucional se divide en tres períodos: comienzo, desarrollo y madurez. Estos períodos son piedras angulares en el desarrollo por fases y permiten diferenciar niveles de compromiso de recursos para afrontar estos objetivos.
- Muchas instituciones prefieren comenzar con un servicio pequeño y manejable, acompañado de un programa de marketing cuyo objetivo inicial sea hacerse visible en la comunidad. Conforme crece el interés, se añaden comunidades de contenido y departamentos académicos adicionales.



Actividades y servicios posibles

Planificación Temporal, por ejemplo:

- ✓ Identificar el equipo del servicio y perfilar las responsabilidades.
 - ✓ Identificar a los primeros usuarios para empezar a añadir contenidos en una fase piloto o beta.
 - ✓ Identificar colecciones históricas para adquirir contenidos (digitalización??).
 - ✓ Identificar nuevos contenidos de investigación.
 - ✓ Identificar líderes para pasar la idea en el campus.
- ✓ Desarrollar planes de actuación.
- ✓ Elegir e identificar tecnologías.
- ✓ Desarrollar un equipo asesor.



Personal: Servicio y apoyo

Dependiendo del tamaño y alcance del servicio del Repositorio Institucional pueden necesitarse recursos para realizar las siguientes tareas:

- Ayudar al establecimiento de la comunidad.
 - Diseño de la página web.
 - Definición de la colección.
 - Definición del volumen de trabajo.
 - Cargar lotes de colecciones históricas.



Personal: Servicio y apoyo

- Ayudar a los usuarios.
 - Ayuda telefónica.
 - Ayuda en línea.
 - Documentación en línea.
 - Páginas FAQ.
- Revisar metadatos.
- Hacer que el personal de la biblioteca cree metadatos.
- Desarrollar esquemas de metadatos personalizados.
- Dirigir colecciones.
- Consultar con comunidades y con autores sobre la conservación.



Personal: Servicio y apoyo

- ***Funciones de ayuda al usuario y roles tecnológicos***
- Para un servicio grande o ya establecido, hay dos funciones primordiales: una eminentemente **técnica (sistemas informáticos / tecnología)** y otra de **ayuda a los usuarios, que gestione el Repositorio Institucional** con los profesores e investigadores.
- También puede pensarse una lista con las habilidades necesarias para cada uno de los papeles más destacados en el equipo.



Un programa piloto

- Cada servicio de Repositorio Institucional se compone de colecciones de contenido, a veces llamadas **comunidades**. Estos son grupos que añaden contenidos a un Repositorio Institucional, departamentos académicos o administrativos, Universidades, centros de investigación, unidades de laboratorios etc.
- Muchas Universidades han encontrado útil poner en marcha un **programa piloto** para su repositorio común, eligiendo en un primer momento a unos pocos usuarios iniciales que prueban el software y han aceptado unirse al programa previamente. Esto facilita la atención a un grupo de contenido y de usuarios asequibles al sistema, probar el software, suprimir procedimientos, y comprobar la política de actuación y las expectativas antes de poner el servicio a disposición de toda la Universidad.
- Una vez que se ha elegido una colección para añadir se pasa al programa piloto, y se elige un enlace elija a alguien como enlace en ese departamento para trabajar con el personal del repositorio y el equipo de implementación.



La administración...

- DSpace es de administración distribuida en todo sentido.
- Primero, porque es un sistema completamente web, permitiendo que cualquier usuario con acceso a internet acceda y dependiendo de los permisos en el sistema se le permite hacer determinadas cosas.
- Segundo, porque está pensado para que los usuarios "normales" hagan autoarchivo de los trabajos, permitiendo definir restricciones de subida por comunidad o colección a determinados usuarios, y luego de ese autoarchivo, los documentos entran en el circuito de revisión, en el que pueden ser rechazados, modificados, aceptados, etc
- Todo eso sería la administración distribuida cabe destacar que el circuito de revisión es manejado por usuarios con privilegios de acceso a esa sección del software siempre a través de la web.



Un programa piloto

- Enviar información por e-mail sobre el servicio de Repositorio Institucional a los foros de discusión de la institución.
- Sondear/entrevistar/encuestar al profesorado para averiguar lo que necesita de un Repositorio Institucional.
- ***Criterios de selección*** para los primeros usuarios:
 - Un grupo o departamento que esté de acuerdo con la misión del RI.
 - Diversidad entre las disciplinas.
 - Diversidad de tipos de contenido o formatos.
 - Incluir ejemplos diferentes sobre temas de gestión de los derechos de propiedad intelectual.



Un programa piloto

- Colecciones de tamaño manejable en lugar de grandísimas colecciones.
- Opinión de los principales responsables o cargos del campus.
- Proporcionar una red de recuperación de contenido, identificar el contenido que está a punto de perderse.
- Una comunidad predispuesta a la conservación digital y experiencia en la interpretación de datos.
- Lo más importante de todo es un enlace fuerte y de confianza dentro de la comunidad para trabajar con el equipo de servicio del Repositorio Institucional.

Los criterios de selección para los primeros usuarios:



Los criterios de selección varían dependiendo de la fase de desarrollo que se esté implementando.



Cómo hacer el marketing del RI

- Cuando se crea un servicio, es imprescindible difundir cómo va a beneficiar dicho servicio a la comunidad universitaria y anunciar el nuevo servicio al campus.
- Algunos equipos han tenido éxito al centrarse en un grupo de “líderes” del campus, haciendo que se comprometieran para fomentar el interés por el servicio, otros centrándose en grupos: investigadores, informáticos para difundir la información entre los que enviaban contenidos académicos a las páginas web de la Universidad.



Cómo usar un método descendente

- Un método descendente se centra en los Vicerrectores, Decanos y Administradores. Se usa el boca a boca y la influencia directa.
- Convencer a los cargos directivos de la institución para tener un repositorio digital a través de los profesores y demás personal.
- Desarrollar algún ejemplo que demuestre el valor del repositorio para la institución como conjunto, mostrando investigaciones universitarias y el beneficio para la investigación académica.



Cómo usar un método ascendente

Ofrece servicio al personal, a las comunidades que publican en la página web de la Universidad.

- Se crea la necesidad de un Repositorio Institucional antes de pedir apoyo a un nivel superior.
- Comité docente asesor que describa a sus colegas los beneficios que comporta usar un RI.
- Valorar las particularidades en cuanto a producción académica, necesidades digitales, etc.
- Usar la amplia gama de disciplinas y sus necesidades
- de publicación digitales.
- Reunirse con editores y webmasters y ofrecer el servicio. Ellos conocen los retos del uso y conservación del contenido en línea.
- Colaborar con otras iniciativas del campus interesadas en
- tener contenido en línea, cursos, etc.

Cómo fomentar el interés en la Universidad



- Presentar el servicio en reuniones del campus, a grupos de investigación, departamentos, personal, al servicio de informática...
- Escribir notas (prensa) anunciando el RI y distribuir folletos.
- Coordinar la publicidad con todo el ámbito universitario.
- Compartir pósters, folletos con las oficinas, páginas web, etc.
- Planificar actividades en el campus y en el entorno de las comunidades de contenido para publicitar el lanzamiento.
- Reunir al personal para dar a conocer el RI y plantee dudas.
- Justificar la existencia de un programa de RI antes de lanzarlo.
- Hacer publicidad tanto dentro como fuera de la Universidad.
- Escuchar al profesorado y los usuarios finales del campus, y ser flexible respecto a sus expectativas.
- Promover el interés en la conservación a largo plazo en el campus.
- Ofrecer presentaciones sobre temas de interés relacionados con los Ris: copyright, etc.

Cómo mantenerse en contacto con comunidades de contenido



- Encuestar a las comunidades de contenido anualmente para conseguir información y reunir nuevos requisitos etc.
- Usar una solicitud anual para verificar las decisiones sobre la política de actuación.
- Poner en funcionamiento una línea de ayuda de forma que los que envían contenido y el personal directivo se dirija al directorio de ayuda al usuario .
- Hacer un seguimiento de resolución de problemas y sugerencias de mejora.
- Compartir FAQs con las comunidades de contenido del campus.
- Ver la experiencia en marketing de otros Ris.

Cómo añadir contenido al servicio



- Entre los equipos de Ris encuestados, la
- prioridad y mayor reto es conseguir el contenido.
- Una vez que conocen el servicio y valoran su importancia, interesa la idea del repositorio.
- Conseguir que envíen contenido es el siguiente reto.

Cómo elegir la plataforma de software del Repositorio Institucional.



- Una vez determinada la necesidad de crear un Repositorio Institucional y empezar a planificar lo que oferta el servicio, toca examinar cuidadosamente los sistemas disponibles para elegir el que se adapte a sus necesidades.
- Tener en cuenta los recursos disponibles.
- Cada plataforma tiene puntos fuertes.
- Hay que mirar lo que usan los otros Ris.



Usos

- Archivos de pre-print y e-print
- Tesis en línea.
- Materiales educativos...



Componentes de un sistema de repositorio

- Los componentes esenciales de un RI son:
- Interfaz para añadir contenido al sistema.
- Interfaz para buscar/ comprobar/ recuperar contenido.
- Base de datos para almacenar contenido
- Interfaz administrativa para apoyar la gestión de las colecciones y las actuaciones de conservación.
- Una característica adicional puede ser la integración con otros sistemas universitarios incluidos cursos en línea, *etc.*

Cómo elegir la plataforma de software del Repositorio Institucional.



- Este documento presenta los temas a considerar, los requisitos a tener en cuenta y
- los puntos fuertes de las plataformas de los principales RI disponibles en la actualidad.

Cómo elegir la plataforma de software del Repositorio Institucional.



- Tecnología básica de un Repositorio Institucional.
- Las características de los productos.
- Modelos de producto.
- Otros aspectos técnicos sobre el funcionamiento de un servicio.
- Pasos para la implementación.
- Consideración de los costos.
- Principales proveedores de software de RI.
- Listas con las características principales.



Cómo elegir una plataforma de Software

- Equipo con distintos perfiles.
- Cada miembro aporta experiencia sobre cómo debería funcionar el sistema y las características necesarias, tanto las de servicio (metadatos, aportación de datos, tipos de contenido, etc.) como las relacionadas con los servidores subyacentes (sistemas operativos, bases de datos, mecanismos de búsqueda, etc).

Requisitos a tener en cuenta y puntos fuertes de las plataformas de RIs



- Tecnología básica de un RI.
- Características del producto.
- Modelos de producto.
- Otros aspectos técnicos sobre el funcionamiento de un servicio.
- Pasos para la implementación.
- Consideración de costos.
- Principales proveedores de software de RI.
- Listas con las características principales.



Tecnología básica de un Repositorio Institucional

- Servidores Windows o Unix/Linux Servers.
- Servidor Web, como Apache y herramientas de aplicación web relacionadas.
- Bases de datos, como MySQL, DB2, Oracle, Postgres, servidor SQL.
- Software de un Repositorio Institucional.
- **Observaciones:** puede que necesite varios servidores para este servicio, para cada una de estas fases de desarrollo, prueba y producción.



Características deseables del producto

- Apoyo a formatos de archivo: textos, imágenes, conjuntos de datos, vídeo, audio, etc.
- Estándares de metadatos (descriptivo, técnico, de conservación, derechos).
- Interoperatividad: OAI compliance, Z39.50, SRW, etc.
- Dirección o localizador permanente del artículo.
- Búsqueda/vista de metadatos.
- Búsqueda de texto completo.



Características deseables del producto

- Volumen de trabajo, aportación para la aprobación del contenido??
- Autenticación y autorización del usuario:
 - Usuario administrador: proveedor de contenido, editor, administrador, editor de metadatos.
 - Usuario final: acceso al contenido.
- Personalización: API (interfaz de aplicación de la programación) para personalizar el software, aumentar las características según sea necesario.



Productor/distribuidor del software

- Gratuito vs. software comercial (licencia, suscripciones).
- Código abierto vs. propietario.
- Apoyos técnicos disponibles.



Modelos de la tecnología del producto

- Las plataformas de software están disponibles en diferentes modelos de licencia y distribución.



Modelos de la tecnología del producto

Las plataformas de software de un RI están disponibles en diferentes modelos de licencia y distribución.

- **Software de propiedad.**
- Se paga por el software y de manera opcional consigue cualquier suscripción adicional o de consulta, o bien posee el uso del software y con una suscripción consigue actualizaciones de software.
- Con una interfaz de programación, o API, se puede personalizar el software, pero el vendedor del software posee, crea y
- mantiene el código fuente.



Modelos de la tecnología del producto

Software de código abierto.

- Se descarga la plataforma de software, en la mayoría de los casos gratuito, y el personal de informática puede examinar, personalizar y mejorar el código fuente. Un organismo central dirige el código fuente, pero está abierto a cambios y mejoras por parte de la comunidad de desarrollo (por ejemplo, CDSware, DSpace, EPrints, Fedora, Greenstone).



Modelos de la tecnología del producto

Modelo de Servicio de Software.

Un vendedor de software posee y distribuye una plataforma de software, o también guarda y gestiona los datos del cliente. En este modelo, el vendedor de software proporciona servicios adicionales prepago y también controla y pone al día el código fuente del software (por ejemplo, Open Repository o Bepress). Hay tres opciones:

- Hacerlo funcionar y gestionarlo de forma local.
- Hacerlo funcionar de forma local y ser gestionado de forma remota por el vendedor.
- Hacerlo funcionar y ser gestionado de forma remota por el vendedor.



Modelos de la tecnología del producto

- El modelo de software que se ajuste mejor depende de las necesidades y recursos, de si se está usando una plataforma de software de código abierto que desarrolle la comunidad o se prefiere una relación convencional de software vendedor-cliente, donde se puede pagar por apoyo técnico, actualizaciones y consultas según se necesite.
- Hay que atender a los costos y ver otras experiencias.

Cuestiones técnicas una vez que el servicio está funcionando



- Dependiendo de la plataforma de software que elija, el personal técnico puede manejar los siguientes aspectos de entrega de servicios:
- Escalabilidad (crecimiento).
- Carga y recuperación.
- Mantenimiento del sistema.
- Extensibilidad: acceso a recursos sistemas de otras Universidades.
- Personalización.
- Internacionalización/ apoyo multilingüe.
- Carga de datos.



Pasos de implementación

- Examen de las necesidades y requisitos del servicio.
- Elección de la plataforma de software.
- Selección e instalación del hardware necesario, servidores.
- Instalación y configuración del software.
- Creación de una versión test / demo de su sistema.
- Personalización de la interfaz según lo necesite.
- Prácticas para el personal.
- Creación de volúmenes de trabajo para aprobación del contenido: aceptar, editar, rechazar, etc.
- Carga de contenido.
- Sistema de pruebas.



Consideración de costos

- Los costes de las plataformas del software y del hardware son, en general, predecibles.
- Los costes de la planificación total, implementación, prácticas de personal y funcionamiento del servicio, dependen del plan.
- Varias plataformas de Ris son de fuente gratuita y abierta. Esto significa que el software está libre de costos, y su personal de desarrollo técnico puede modificar el código fuente y mejorar sus características.

Consideraciones sobre el costo de la tecnología



- Costos del software (actuales y futuros).
- Hardware, servidores, etc.
- Personal de operaciones.
- Personal de programación.
- Carga y recuperación.
- Conservación.



Planificar a largo plazo

- Cuando se elige una plataforma de software para RI se consideran no sólo sus necesidades actuales sino pensando en el servicio a futuro.
- ¿Qué tipos de contenidos podría necesitar albergar?.
- ¿Qué importancia tiene para la institución la preservación de contenidos?
- ¿Cuántos miembros de la Universidad adoptarán el servicio?.
- ¿Cómo podría aumentar la cantidad de envíos al tiempo que crece la curva de aceptación?.
- ¿Qué iniciativas de conservación aplicar para almacenar archivos y datos?.
- La estrategia de migración o exportación de datos si se cambia el sistema.



Conservación digital

Además del acceso abierto al material de investigación la conservación 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. Aunque el objetivo de los RIs no es la conservación digital se trata de una herramienta importante para alcanzarla.



En conservación: SHERPA

- La información digital se pierde cuando no se actualiza mientras que el hardware, el software y los medios de comunicación siguen desarrollándose.
- Si no se transforma, la edición electrónica podría perder relevancia con el tiempo, aunque esté actualmente en alza, ya que sería inaccesible su contenido si el software y el hardware cambiasen.



Estrategias de conservación

- **Copias Bitstream.** Archivo de datos(backup), donde se puede hacer un duplicado del objeto digital.
- **Medios de comunicación duraderos y permanentes.** Donde se conservan los medios físicos en los cuales el objeto se almacena.
- **Migración.** Pasaje de datos de una tecnología a otra para evitar que se vuelvan obsoletos por el hardware o el formato.
- **Estándares.** Confianza en estándares reconocidos, a largo plazo, sobre formatos propietarios.
- **Emulación.** Proceso de reproducción de entornos de software y hardware para traducir códigos de un programa informático para que funcione en otro.
- **Encapsulación.** Como parte de una estrategia de emulación, donde los objetos y metadatos se agrupan juntos para ayudar a decodificar y ejecutar el objeto más tarde.
- **Conservación de Metadatos.** Describe el software, hardware y los requisitos de un objeto digital que se deben usar para conservar el objeto.

Fuente <http://www.nla.gov.au/padi/topics/18.html>



Proveedores de Software de Repositorios Institucionales

Productos:

- Archimede.
- Bepress.
- CDSware.
- CDS-Invenio
- CONTENTdm.
- DSpace.
- EPrints.
- Fedora.
- Greenstone.
- Open Repository.

Mostrar panorama en ROAR



Software de Repositorios Institucionales comparativa

- Se podría utilizar la tabla comparativa de:
- <http://www.rsp.ac.uk/start/software-survey/results-2010/>
- Aunque no es exacta la misma lista de productos...
- Otra:
http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_Table_v3.pdf



Marco legal regulador y política de desarrollo.

- Entender los derechos de la propiedad intelectual para RI.
- Copyright.
- Licencia.
- Gestión de derechos.
- Legislación relevante.
- Cambios en el copyright.
- Depósito Legal.
- Decreto de Libertad de Información.
- Directrices de política de actuación para Repositorios Institucionales.

Conocer las normas sobre propiedad intelectual para los RIs



- Esta sección describe las principales cuestiones sobre **derechos de**
- **propiedad intelectual a tener en cuenta al crear un RI** incluyendo el copyright, las licencias de contenido y los derechos de gestión.
- **La normativa sobre propiedad intelectual se refiere en general** a la propiedad y derechos sobre el trabajo producido y distribuido tanto en línea como impreso.



Copyright y licencias de contenido

El copyright permite a los creadores de contenido controlar el uso y la distribución de su material. Para más detalles sobre cómo se define y se obtiene el copyright, ver la excelente página web de la oficina de patentes de UK que describe todos los términos y condiciones del copyright en el Reino Unido:

<http://www.patent.gov.uk/coy/indetail/basicfacts.htm>



Copyright

Conocer a fondo los temas sobre copyright es vital para el éxito de un proyecto de Repositorio Institucional. La persona entendida en copyright de su Universidad es la más adecuada para interpretar cómo estas leyes afectan a la Universidad y a un Repositorio Institucional. Los Repositorios Institucionales tienen que ver con el **copyright** en dos frentes: a la hora de recoger contenido del profesorado mediante el cual se deben asegurar los derechos para distribuir y conservar dicho contenido, y en la distribución del mismo a usuarios finales, donde hay que compatibilizar el acceso abierto con la protección de copyright.



Copyright

Cuando se trabaja con centros académicos que aportan contenidos a su Repositorio Institucional, puede que sea mejor que conserven el copyright de su producción o al menos los derechos para publicar su trabajo de forma electrónica. Ofrecemos detalles y directrices a continuación.



Licencias de contenido

Son los acuerdos legales mediante los cuales se puede distribuir dicho contenido. Normalmente un RI tiene estas dos licencias:

- **Licencia de depósito:** un acuerdo entre el creador (o poseedor de copyright) y la institución que le da al repositorio derecho para distribuir y conservar el trabajo.
- **Licencia de distribución:** un acuerdo entre el creador (o poseedor de copyright) y el usuario final que trata sobre el uso que puede hacerse del trabajo.



Licencias de contenido

Ofrece a creadores y distribuidores de contenido una gran variedad de licencias, permitiéndole al creador de contenido estipular las condiciones para usar dicha licencia de contenido.



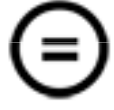

<http://creativecommons.org/>, para obtener información sobre las licencias ofertadas y herramientas para los creadores y distribuidores de contenido.

<http://creativecommons.org/learn/legal/>

Ofrece excelente información sobre los conceptos legales fundamentales sobre propiedad intelectual.



Licencias CC

-  **Reconocimiento (Attribution):** En cualquier explotación de la obra autorizada por la licencia hará falta reconocer la autoría.
-  **No Comercial (Non commercial):** La explotación de la obra queda limitada a usos no comerciales.
-  **Sin obras derivadas (No Derivate Works):** La autorización para explotar la obra no incluye la transformación para crear una obra derivada.
-  **Compartir Igual (Share alike):** La explotación autorizada incluye la creación de obras derivadas siempre que mantengan la misma licencia al ser divulgadas.

Directrices de copyright para investigadores



- Los docentes que depositan su investigación en Ris puede que necesiten información adicional sobre el copyright.
- Algunas organizaciones proporcionan información y guías sobre el copyright enfocado a la investigación académica.
- El grupo Creative Commons:

<http://creativecommons.org/learn/licenses/>



- Las editoriales suelen acceder a la petición del autor de mantener los derechos cuando envía contenido a una página web o a un RI.
- Hay que potenciar que los centros académicos conserven estos derechos antes y después de publicar para que puedan contribuir con su contenido a repositorios en línea.



- El proyecto RoMEO es un recurso excelente para el profesorado interesado en archivar su material.
- Su página web proporciona información pertinente sobre acuerdos de negociación de contenido entre los centros y las editoriales.

<http://www.lboro.ac.uk/departments/ls/disresearch/romeo/>



El proyecto EPrints publica una extensa información y guía de auto-archivo y archivos abiertos, así como un glosario de términos en esta área

<http://www.eprints.org/glossary/>

y enlaces con las páginas web más importantes

<http://www.eprints.org/self-faq/>

Gestión de derechos



- La terminología sobre la gestión del copyright en la era digital es confusa.
- **La gestión digital de derechos se refiere en general a la normativa** de copyright aplicable a la edición en formato digital. Hay varias iniciativas de investigación en la actualidad que tratan estos asuntos.



- **Los proyectos de derechos de metadatos tratan sobre los** derechos de la información en metadatos. Han aparecido dos estándares destacados: XRML, una iniciativa privada, y ODRL, un estándar abierto. Muchos proyectos de código abierto usan la especificación abierta de ODRL.

XrML: <http://www.xrml.org/>

ODRL: <http://odrl.net/>

Para los Repositorios Institucionales, la gestión de derechos se refiere



- A cómo se trata el contenido según las normas del copyright y a quién pertenece. Los Repositorios Institucionales suelen ser de acceso abierto. Sin embargo, hay ocasiones en las que el acceso es restringido, como por ejemplo en la información relacionada con materiales de patentes.
- La plataforma de software que usa para crear su servicio puede estar preparada tecnológicamente para gestionar estos derechos de acceso. En MIT, por ejemplo, el sistema de control de acceso DSpace es restringido para materiales relacionados con las patentes tecnológicas.



Información sobre derechos

- Legislación reciente
- Cambios de Copyright
- Recursos de Copyright.
- Depósito Legal.
- Recursos de depósito legal.
- ¿Libertad de información?.



Directrices sobre la política de actuación de un RI

- Cada Universidad que desarrolla un Repositorio Institucional en línea necesita investigar y especificar políticas de actuación y regulaciones para sus colecciones. Ésta sección ofrece directrices para diseñar las políticas de actuación que se adapten a su Repositorio
- Institucional e incluye los siguientes temas:
 - Creación de directrices de políticas de actuación.
 - Formación de un equipo asesor.
 - Asuntos a tener en cuenta.
 - Implicaciones tecnológicas...



Ejemplo: Políticas de Digital.CSIC

- Política de contenidos y colecciones
- Política de servicios
- Política de metadatos
- Política de datos
- Política de depósitos
- Política de edición, retención, sustitución y eliminación de registros
 - Política de estadísticas
 - Política de preservación digital
 - Política de soporte de formatos
 - Política de privacidad



Política del SeDiCI

- [De contenidos](#)
- [De metadatos](#)
- [De datos](#)
- [De diseminación](#)
- [De depósito](#)
- [De preservación digital](#)
- [De estadísticas](#)

<http://sedici.unlp.edu.ar/about/politicas.php>

Política de contenidos del SeDiCI



SeDiCi alberga la producción de los alumnos, docentes e investigadores de la Universidad Nacional de La Plata.

Los tipos de documentos aceptados, en cualquier idioma, son: tesis (grado y posgrado), artículos, libros, presentaciones en congresos, producciones artísticas, publicaciones periódicas, proyectos de investigación, patentes, revisiones, producidas en cualquiera de las unidades académicas de la UNLP y cualquier otra forma de producción de valor académico.

SeDiCi puede trabajar con una gran cantidad de formatos digitales: textos, imágenes, archivos de audio, de video, programas, objetos de aprendizaje, páginas web, etc.

Política de metadatos del SeDiCI



Los metadatos son la información descriptiva que se aplica sobre un recurso para facilitar su organización, recuperación, preservación e interoperabilidad. Básicamente son datos que describen otros datos. Los metadatos se refieren a un recurso y sirven como índices para organizar y localizar contenidos. Esa es su importancia: la inclusión de metadatos correctos asegura una localización del recurso y posibilita una mayor visibilidad de las obras.

SeDiCi utiliza un formato de metadatos propio, los cuales están traducidos al inglés y al portugués, y los transforma al formato [Dublin Core](#) para que sean accedidos vía [OAI](#).

Política de datos del SeDiCI



Los documentos alojados en SeDiCi pueden ser visualizados y descargados con fines educativos, de estudio, y no lucrativos, sin necesidad de pedir permisos explícitos, siempre y cuando, en caso de citas textuales,

- 1) se anexe el nombre de los autores, el título y los detalles bibliográficos completos,
- 2) el contenido no sea alterado, excepto con la expresa autorización del autor.

Política de disseminación del SeDiCI



SeDiCI fomenta la compartición de sus registros en favor de la difusión a través de protocolos estándares de intercambio de metadatos, como por ejemplo [OAI-PMH](#).

Política de depósito del SeDiCI



Cualquier miembro alumno, docente o investigador de la UNLP, o con vínculos probados con ella, puede publicar sus trabajos en forma gratuita en SeDiCi, personalmente o en línea (autoarchivo). Los tipos de materiales aceptados son obras de todo tipo, de naturaleza académica producidas por autores vinculados a la UNLP como se detalla en la política de contenidos.

Para completar la publicación de los documentos en SeDiCi (mediante autoarchivo o personalmente) es necesaria la aceptación de una licencia no exclusiva de depósito, permitiendo a SeDiCi el archivo, preservación y difusión de la obra. La condición “no exclusiva” significa que el creador no pierde los derechos de autor sobre su trabajo, pudiendo publicar el mismo en publicaciones, otros repositorios, etc.

Política de preservación digital del SeDiCI



SeDiCI tiene el propósito de preservar digitalmente y a tiempo indefinido todos los documentos que alberga. Para ello está desarrollando un plan de acción que incluye medidas capaces de asegurar la perdurabilidad del material digital presente en el Repositorio, entre otras:

- Backups
- Conversión de formatos
- Comprobaciones periódicas de la integridad y accesibilidad de los archivos
- Vigilancia tecnológica
- Identificadores persistentes

Parte de esta política es la digitalización de material antiguo (tesis, libros, publicaciones, etc.) alojado en las diferentes bibliotecas pertenecientes a la UNLP, no sólo en función de la preservación del mismo, sino como memoria de la producción académica de la Universidad.

Política de estadísticas del SeDiCI



SeDiCi realiza estadísticas para analizar no sólo la difusión y visibilidad de la producción de la UNLP, sino también el crecimiento de los contenidos disponibles en el repositorio, las visitas, las descargas y las modalidades de sus usuarios, estas estadísticas son de carácter público y pueden ser consultadas en la web de SeDiCi por todos los visitantes.



Dos componentes en la creación de RI: tecnologías y metadatos

- **Metadatos:** Son datos que describen otros datos, información estructurada para describir, explicar, localizar o facilitar la obtención, uso o administración de un recurso de información.
- Abarcan dos aspectos:
 - **Descripción** del documento
 - **Interoperabilidad**

Datos y metadatos



Los datos que describen un documento (autor, fecha, palabras clave) deben ser mapeados (traducidos) a su equivalente en el esquema de metadatos

dc.date.available	2012-12-12T13:39:33Z	
dc.date.issued	2012-12	
dc.identifier.uri	http://hdl.handle.net/10915/25158	
dc.description.abstract	Objetivos: - Que los autores conozcan las posibles vías de publicación de sus trabajos. - Que conozcan sus derechos. - Destacar la importancia de gestionar correctamente los derechos de autor de una obra. - Informar acerca de cómo mantener los derechos. - Recomendar dónde publicar. - Pensar qué deben conocer los gestores de repositorios sobre estos tópicos para la gestión e inclusión de las obras. - Generar preguntas, plantear dudas y debatir con los especialistas y los involucrados.	es
dc.format.extent	87 diap.	es
dc.language	es	es
dc.title	Vías de publicación y derechos de autor en la academia	es
dc.type	Objeto de conferencia	es
sedici.creator.person	De Giusti, Marisa Raquel	es
sedici.subject.materias	Ciencias Informáticas	es
sedici.subject.materias	Bibliotecología	es
sedici.subject.keyword	derechos de autor	es



Tipo de metadatos

- **Metadatos administrativos:** información de “mantenimiento” sobre el propio registro – su creación, modificación, relación con otros registros, etc. (Número de registro; Fecha de creación del registro; Fecha de la última modificación)
- **Metadatos descriptivos:** Descripción e identificación de recursos de información (Título; Fecha; Autor; etc.)
- **Metadatos analíticos o temáticos:** información que analiza y facilita el acceso a los contenido del recurso (Encabezamientos de materia; Tesauros; Palabras clave)



Dublin Core

DC es un modelo de metadatos elaborado por la **DCMI** (Dublin Core Metadata Initiative). Actualmente es el esquema de datos más utilizado.

Su objetivo es elaborar normas interoperables sobre metadatos y desarrollar vocabularios especializados en metadatos para la descripción de recursos que permitan sistemas de recuperación más inteligentes.

<http://www.dublincore.org/documents/2004/09/10/library-application-profile/>



Dublin Core

Básico

- DC trabaja con quince elementos básicos (todos ellos optativos, repetibles y extensibles)

Contenido	Propiedad intelectual	Instanciación
Title	Creator	Date
Subject	Publisher	Type
Description	Contributor	Format
Source	Rights	Identifier
Language		
Relation		
Coverage		

Cualificado

- Propone un perfil de aplicación que amplía los elementos del Dublin Core básico.
- Ejemplos:
 - Date
 - DateCopyrighted
 - DateSubmitted
 - DateAccepted
 - DateCaptured



Dublin Core - Interoperabilidad

La interoperabilidad es la capacidad de un sistema de información para comunicarse y compartir datos, información, documentos y objetos digitales de forma efectiva con uno o varios sistemas de información



Interoperabilidad

La interoperabilidad se manifiesta en:

- La capacidad de los sistemas para trabajar entre sí en tiempo real.
- La capacidad del Software para trabajar en diferentes sistemas.
- La capacidad de los datos para ser intercambiados entre diferentes sistemas (portabilidad)



Interoperabilidad

Para esto se debe potenciar el desarrollo de:

- Formatos estandarizados de documentos
- Formatos estandarizados de Metadatos
- Formatos estandarizados de protocolos de comunicación y recuperación.
- Medios estandarizados de autenticación y seguridad.



Interoperabilidad-directrices

DRIVER

“Digital Repository Infrastructure Vision for European Research” (Visión de infraestructura de repositorios digitales para la investigación europea), es un proyecto realizado por un consorcio financiado por la Unión Europea que está creando un marco de trabajo tecnológico y organizativo para implementar una capa paneuropea de datos, que permita el uso avanzado de los recursos de contenido en el ámbito de la investigación y la educación superior. DRIVER desarrolla una infraestructura de servicios y una infraestructura de datos. Ambas están concebidas para orquestar los recursos y los servicios existentes en la red de repositorios.



Interoperabilidad-directrices

- Las directrices son una herramienta para hacer corresponder (o traducir) los metadatos empleados en el repositorio con los metadatos de Dublin Core tal como los recolecta DRIVER.
- No están pensadas para utilizarse como instrucciones de introducción de datos en la operación de inserción de metadatos en el sistema de repositorios.

Se centran en cinco cuestiones:

- colecciones
- metadatos
- implementación del protocolo OAI-PMH
- prácticas recomendadas
- vocabularios y semántica

<http://www.driver-support.eu/>

http://www.driver-support.eu/documents/DRIVER_2_0_Guidelines_Spanish.pdf



Interoperabilidad-directrices

- A nivel latinoamericano: BDCOL (Biblioteca digital colombiana) creó sus propias directrices tomando como base la propuesta europea de DRIVER.
- A nivel nacional se crea el SNRD (Sistema nacional de repositorios digitales). Dependiente del Consejo Asesor de la Biblioteca Electrónica del Ministerio de Ciencia, Tecnología e Innovación Productiva (MINCYT)

Sistema Nacional de Repositorios Digitales



- Para optimizar los recursos y para asegurar la sostenibilidad de los repositorios argentinos el Consejo Asesor de la Biblioteca Electrónica del Ministerio de Ciencia, Tecnología e Innovación Productiva (MINCYT), impulsa la creación de un Sistema Nacional de Repositorios Digitales (SNRD) el que se crea por la Resolución MINCYT N° 469/11

[Sitio](#)

Sistema Nacional de Repositorios Digitales



Son sus objetivos:

- a) Promover el acceso abierto a la producción científico-tecnológica generada en el país.
- b) Promover el intercambio de la producción científico-tecnológica e incrementar su accesibilidad a través de una red nacional de repositorios interoperables entre sí.
- c) Generar políticas conjuntas que favorezcan la sostenibilidad de los repositorios digitales de ciencia y tecnología.
- d) Delinear estrategias dirigidas a garantizar el respeto por los derechos de los autores de los objetos digitales incluidos en los repositorios.

Sistema Nacional de Repositorios Digitales



Son sus objetivos:

- e) Definir estándares generales para el correcto funcionamiento del Sistema.
- f) Dotar de proyección internacional a la producción científico-tecnológica producida en el país a través de su difusión en redes virtuales y su interoperabilidad con repositorios internacionales.
- g) Contribuir a la formación de recursos humanos capacitados a través de programas comunes de desarrollo tanto a nivel local como regional e internacional.
- h) Contribuir a las condiciones adecuadas para la gestión y preservación de los repositorios digitales.
- i) Generar líneas de acción coordinadas con otros Sistemas Nacionales de Bases de Datos.

Planificación de costos del RI



Equipamiento del sistema:

- Software
- Hardware
- Personal de operaciones
- Personal de programación
- Carga y recuperación
- Conservación
- Servicios
- Marketing
- Prensa y difusión



SeDiCI SERVICIO DE DIFUSIÓN
DE LA CREACIÓN INTELECTUAL



Bibliotecas y Repositorios Digitales

Tecnología y Aplicaciones

<http://sedici.unlp.edu.ar>

Participantes del dictado



- Marisa De Giusti
- Nestor Oviedo
- Silvia Peloche
- Matías Cánepa

Bibliotecas y repositorios digitales



Capítulo 4: Aspectos tecnológicos e informáticos. Software de gestión del repositorio. Requerimientos a nivel local. Preservación digital. Servicios: búsqueda, exploración, autoarchivo, DSI, citas, etc.

Contenido



- Software del repositorio
 - Características deseables
 - Alternativas libres
- Representación de recursos
 - Formatos planos vs. jerárquicos
 - Vocabularios controlados simples
 - Entidades abstractas
 - Representación física de los datos

Contenido



- Identificadores persistentes
 - Importancia
 - Algunas opciones disponibles
- Servicios de un repositorio digital
 - Búsqueda y recuperación
 - Exploración
 - Diseminación selectiva de la información
 - Autoarchivo

Contenido



- Estadísticas del repositorio
 - Objetivos
 - Estadísticas frecuentes
- Preservación de contenido
 - Digital obsolescence
 - Estrategias de solución

Contenido



- Repositorio semántico
 - Introducción
 - Problemas relativos a la representación
 - Recuperación de la información y navegación de las relaciones
 - Posibilidad de nuevas estadísticas



Software del repositorio

Software del repositorio



- Es uno de los pilares en la construcción de un repositorio digital.
- Tiene la capacidad de potenciar o limitar todos los aspectos del repositorio (servicios, tamaño, descripción de los recursos, etc.).
- Debe perdurar en el tiempo.

Software del repositorio



Aspectos a evaluar de un software de repositorio

Licencia: es un contrato entre el propietario de los derechos del software y los usuarios que lo utilizan. Este contrato especifica las condiciones bajo las cuales el primero cede derechos o permite actividades sobre el software a los segundos. Licencias conocidas son GPL, Creative Commons, BSD, LGPL, MIT, Apache, etc.

Nivel de impacto: nivel de uso del software por parte de la comunidad de repositorios digitales. Un nivel elevado proporciona confianza y promueve la constante actualización de la aplicación (reporte de errores y mejoras continuas).

Software del repositorio



Aspectos a evaluar de un software de repositorio

Nivel de personalización: medida de las posibilidades de adaptación, tanto de interfaz de usuario como de funcionalidad, para reflejar la identidad y las necesidades de la institución a la que representa. Esto incluye extensiones del software, logos y colores, estructura y organización de contenidos, etc.

Nivel de documentación: cantidad y calidad de la información de todos los aspectos relacionados al software. Desde la instalación y configuración hasta el uso del sistema por parte de usuarios finales y administradores.

Software del repositorio



Aspectos a evaluar de un software de repositorio

Frecuencia de actualizaciones: corrección de errores (de funcionamiento y seguridad) de forma continua, mejora en las funciones existentes e inclusión de nueva funcionalidad que amplíe las características del sistema.

Centros de soporte: listas de correo, wiki, foros, canal de chat y cualquier otro punto de contacto entre un usuario del sistema y los desarrolladores y/o la comunidad de usuarios del software, desde donde puede obtenerse asistencia ante dudas y problemas concretos.

Software del repositorio



Aspectos a evaluar de un software de repositorio

Facilidad de uso: medida referente a la curva de aprendizaje respecto del uso del sistema y todas sus funciones, tanto por usuarios como por administradores.

Formato de metadatos soportado: conjunto de elementos usado para almacenar los datos de cada recurso. Se destaca como un punto importante porque:

- propicia o limita parte de la funcionalidad
- influye en la precisión y completitud de la información
- es un factor de rechazo

Software del repositorio



Aspectos a evaluar de un software de repositorio

Performance: tiempos de respuesta del sistema ante cada solicitud, recursos físicos consumidos (disco, memoria, procesador, etc). La performance habla del balance entre velocidad de respuesta, consumo de recursos, costos, etc.

Escalabilidad: capacidad del software de mantener sus cualidades (performance, simplicidad, mantenibilidad, etc) en niveles aceptables aún cuando el volúmen de recursos, cantidad de usuarios, etc. aumenten considerablemente con el tiempo.

Software del repositorio



Aspectos a evaluar de un software de repositorio

Interoperabilidad: capacidad del sistema de comunicarse e interactuar con otros sistemas. En general los roles de un repositorio pueden ser:

- recolector de recursos/consumidor de servicios
- expositor de recursos/proveedor de servicios

Administración: sección del software de acceso restringido a usuarios con privilegios. Permite acceder a sectores privados del sistema para realizar principalmente acciones de control y mantenimiento.

Software del repositorio

Aspectos a evaluar de un software de repositorio



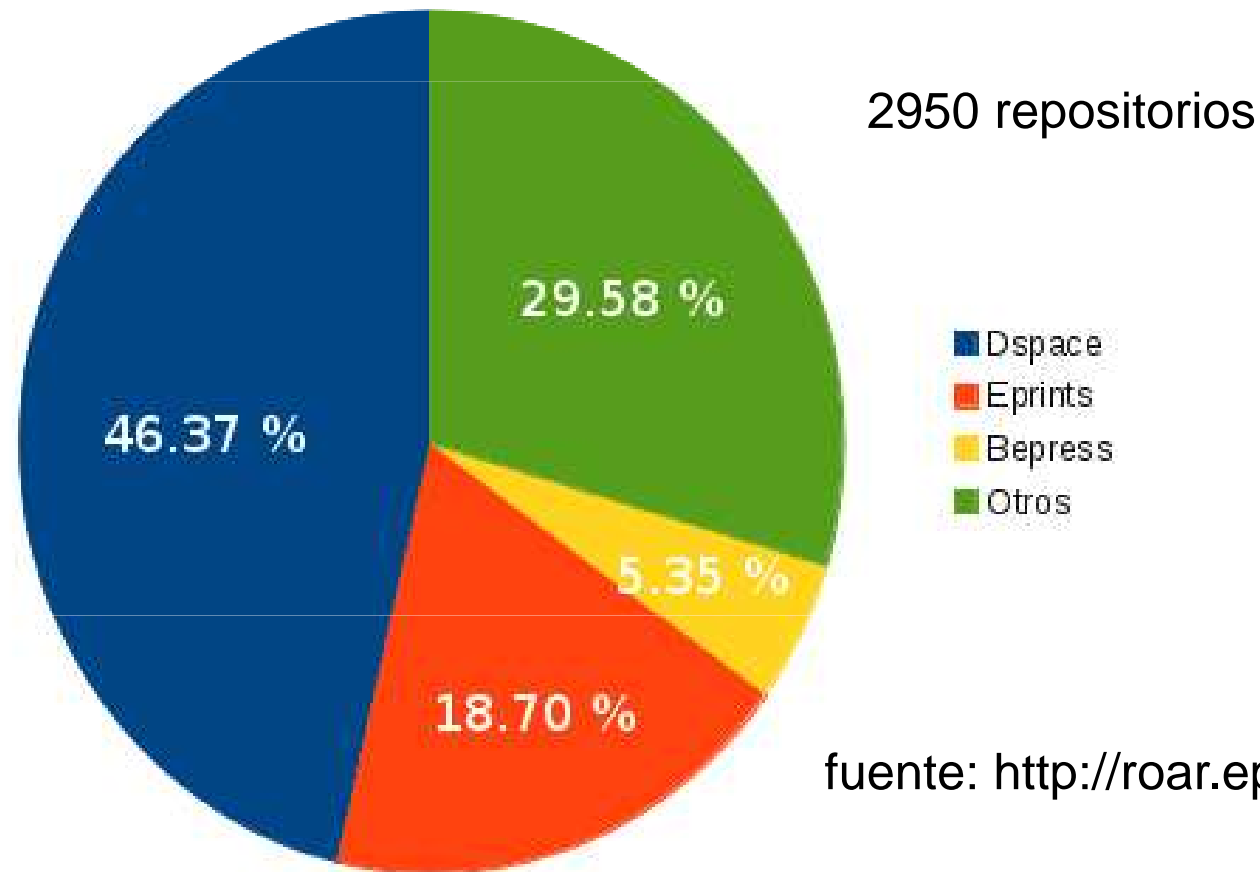
¿Qué buscamos en cada aspecto a analizar?

- Licencia
- Nivel de impacto
- Nivel de personalización
- Nivel de documentación
- Frecuencia de actualizaciones
- Centros de soporte
- Facilidad de uso
- Formato de metadatos
- Performance
- Escalabilidad
- Interoperabilidad
- Administración

Software del repositorio



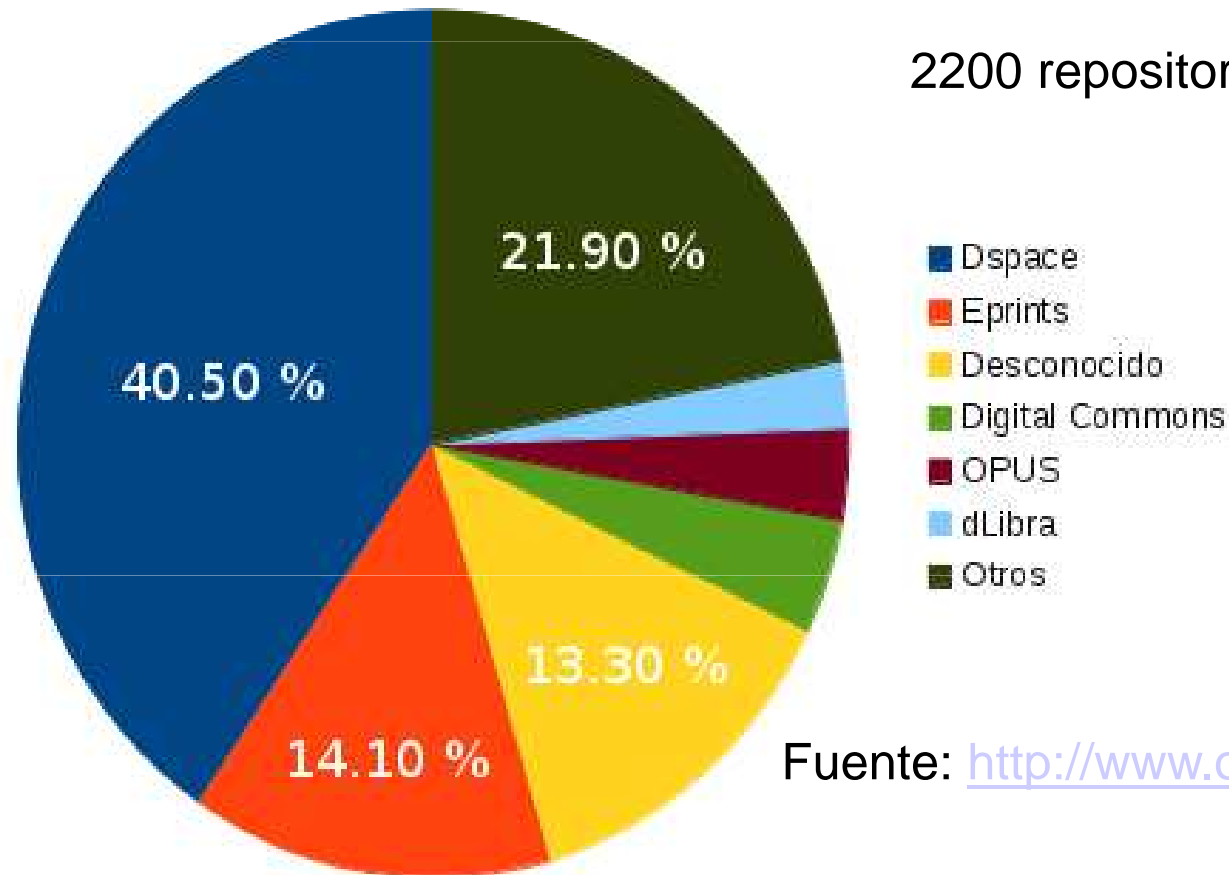
Software de repositorios mas usados



Software del repositorio



Software de repositorios mas usados



Fuente: <http://www.openoar.org>

Software del repositorio



Breve comparativa entre DSpace y EPrints *(más utilizados a nivel mundial)*

	.DSpace	.EPrints
•Sitio web	• http://www.dspace.org	• http://www.eprints.org
•Creadores	•MIT (USA)	•University of Southampton (UK)
•Lenguaje	•Java	•Perl
•Plataforma	•Multiplataforma	•UNIX-like (portado a Windows)
•Base de Datos	•PostgreSQL	•MySQL
•Licencia	•BSD	•GPL v2
•Formato de Metadatos	•Qualified DC, formatos planos	•cualquier formato (incluso jerárquicos)
•Soporte para búsquedas	•Apache Solr (DSpace-Discovery)	•MySQL indexes



Representación de recursos

Representación de recursos



¿Qué se entiende por recurso?

Es todo objeto, físico o digital, que puede ser descrito a partir de la enumeración de un conjunto de datos específicos de dicho elemento, que lo distinguen entre otros objetos.

¿Qué significa representar un recurso?

Habla de registrar de forma persistente el conjunto de datos asociado a un recurso, usando este conjunto de datos como síntesis y reemplazo del objeto "real", permitiendo distribuir el recurso sin necesitar el objeto real (es decir, se usa su representación).

Representación de recursos



La representación que se elija para los recursos del repositorio influye directamente en aspectos como:

- **complejidad del software:** una representación simple implica que los modelos de datos, los procesos de carga e incluso la interfaz de usuario, son más simples.
- **escalabilidad y performance:** cuando el número de recursos aumenta considerablemente, la representación de los recursos comienza a tomar un rol importante. Por ejemplo, en representaciones complejas basadas en bases de datos, la complejidad de las consultas aumenta considerablemente, y por lo tanto también aumentan los tiempos de respuesta.

Representación de recursos



- **Interoperabilidad:** para interoperar es necesario exponer los recursos propios en formatos entendibles por otros sistemas. La elección de la representación influirá en las capacidades del sistema para **derivar** otras representaciones (para su exposición) o bien generar recursos internos a partir de representaciones externas. Esto es, representaciones demasiado simples pueden llevar a transformaciones deficientes, mientras que representaciones muy complejas pueden llevar a procesos de transformación complicados.

Representación de recursos



Formatos de metadatos para la representación de recursos

Según estructura:

- Planos: no existe anidamiento de metadatos
- Jerárquicos: existe anidamiento de metadatos

Según especificidad:

- Simples: pocos elementos, más generales
- Complejos: muchos elementos, más específicos

Representación de recursos



Formatos de metadatos planos

```
<documento>  
  <titulo>...</titulo>  
  <autor>Gomez P.</autor>  
  <filiacion>UNLP</filiacion>  
  ...  
</documento>
```

Parece adecuado, pero ¿qué sucede, por ejemplo, si se tiene más de un autor con disitintas filiaciones?

Representación de recursos



Formatos de metadatos planos

```
<documento>  
  <titulo>...</titulo>  
  <autor>Gomez P.</autor>  
  <filiacion>UNLP</filiacion>  
  <autor>Lopez R.</autor>  
  <filiacion>UTN</filiacion>  
  ...  
</documento>
```

¿Cómo determinar de forma segura qué filiación corresponde a qué autor?

¿Qué pasa si el orden cambia en algún proceso de manipulación de metadatos?

Representación de recursos



Formatos de metadatos jerárquicos

```
<documento>
  <titulo>...</titulo>
  <autor>
    <nombre>Gomez P.</nombre>
    <filiacion>UNLP</filiacion>
  </autor>
  <autor>
    <nombre>Lopez R.</nombre>
    <filiacion>UTN</filiacion>
  </autor>
</documento>
```

Soluciona el problema planteado anteriormente, pero **complejiza el software** del repositorio, ya que la interpretación de estos datos para su validación, procesamiento y presentación ya no son tan simples.

Representación de recursos



La representación de un formato de metadatos plano es relativamente simple. Es decir, básicamente se trata de un listado de elementos con un nombre y un valor (sin considerar por el momento restricciones de tipos de datos, formatos, etc).

Su tratamiento y su representación son relativamente simples

Representación de recursos



Tratar con un formato de metadatos jerárquico dificulta considerablemente su representación. En bases de datos relacionales por ejemplo, debido a la naturaleza anidada de estos formatos, se tiende a crear consultas SQL demasiado complejas, con múltiples JOINS entre las mismas tablas, degradando la performance de forma considerable.

La opción mas viable para este tipo de formatos suele ser alguna forma de representación inherentemente anidada, como ser XML. Esto significaría la necesidad de contar con una Base de Datos XML (posiblemente solo para los documentos).

Representación de recursos



Formatos de metadatos simples frente a complejos

El caso **simple** se destaca por poseer poca cantidad de metadatos, cuya definición es amplia y, en general, poco restrictiva en cuanto a formatos.

En el caso **complejo** existe una mayor cantidad de metadatos, con contenidos mas explícitos y por lo tanto una definición mas restrictiva para cada uno.

Representación de recursos



Ejemplo: al catalogar una tesis con un formato simple como Dublin Core sin calificar, es probable que el director y co-director, junto con la institución de desarrollo, sean catalogados utilizando un mismo elemento: *dc:contributor*, ya que no existe una distinción para estos datos en la definición del formato.

Desde el punto de vista informático esto dificulta:

- presentación: no se puede distinguir de qué dato se trata
- validación: solo puede esperarse texto libre

Representación de recursos

Vocabularios controlados simples



Para determinados metadatos, se indica que su contenido se extrae de un vocabulario controlado, especificando además el vocabulario al que se hará referencia.

- Tesauros
- Sistemas de clasificación
- Idiomas
- Referencias geográficas
- Tipos de recursos
- Materias
- Frecuencias de entrega (mensual, bimestral, trimestral, etc)

Representación de recursos

Vocabularios controlados simples



Se necesita una forma de **Representación**

- Depende del tipo de vocabulario (lista simple de elementos o elementos relacionados).
- Puede ser una tabla en la base de datos, un archivo XML con un *schema* particular, un archivo de texto, etc.
- Debe permitir generar respuestas rápidas.
- Complejidad aportada por las **relaciones** entre elementos.

Representación de recursos

Vocabularios controlados simples



Se necesita **Referenciar** elementos

- Depende de la representación elegida para los recursos (XML, Bases de Datos, etc).
- Debe permitir distinguir de forma unívoca un elemento específico en un vocabulario determinado.
- Decisión entre:
 - Metadato vacío, con un dato adicional para la referencia
 - Metadato con valor del vocabulario replicado y un dato adicional para la referencia
 - Metadato con la referencia como valor

Representación de recursos

Vocabularios controlados simples



Se necesita una forma de **Presentación**

- Debe ser simple e intuitiva (suggest, select, search)
- Debe proporcionar respuestas rápidas
- De ser posible, debe ser **internacionalizable**
- Se debe utilizar desde un formulario de carga, desde una página de presentación de metadatos, desde la exportación de recursos, etc.

Representación de recursos

Entidades abstractas



¿A qué llamamos Entidades Abstractas?

Conjunto de elementos que poseen información descriptiva propia, utilizados en los procesos de catalogación de recursos como elementos de un vocabulario controlado.

Mismas consideraciones que para vocabularios controlados simples, adicionando algunos problemas.

Representación de recursos

Entidades abstractas



Ejemplos:

- Autores: apellido, nombres, email, institución de origen, etc.
- Instituciones: nombre, institución de la que depende, localidad, dirección, mail, responsables, etc.
- Revistas y sus números: nombre, ISSN, director, editor, staff, volúmen, tapa, etc.
- Eventos y sus instancias: nombre, año, ubicación, organizador, etc.

Representación de recursos

Entidades abstractas



Desafíos: **Representación**

- Se debe definir un formato de metadatos (considerar los mismos problemas que para la representación de recursos)
- Opción de usar de WebServices como proveedor de entidades (hay que considerar qué información se incluye en la respuesta del servicio)

Representación de recursos

Entidades abstractas



Desafíos: **Referencia**

Una vez seleccionada una entidad abstracta, es necesario guardar la referencia.

Pueden suceder **problemas de compatibilidad** entre la representación elegida para la entidad abstracta y el o los metadatos del recurso a los cuales esa entidad se asocia.

Representación de recursos

Entidades abstractas



Ejemplo de problemas de compatibilidad

Entidad Autor:

- apellido
- nombre

Metadato autor:

(del formato de catalogación)

```
<author>  
  <lastName/>  
  <firstName/>  
</author>
```

¿Cómo se indica que el campo *apellido* debe ir en el metadato `/author/lastName` y el campo *nombre* en `/author/firstName`?

Representación de recursos

Entidades abstractas



Desafíos: **Presentación**

Además de los elementos a tener en cuenta para los vocabularios simples, es necesario considerar los problemas de compatibilidad entre el formato de la entidad abstracta y el formato de catalogación utilizado.

Representación de recursos

Entidades abstractas



Alternativas de referencia que influyen en la presentación, según en qué momento se realiza la transformación de la entidad abstracta al metadato correspondiente

En ambos casos se asume que la referencia se guarda en un campo independiente

1. en el momento de catalogación
2. en el momento de presentación

Representación de recursos

Entidades abstractas



1. En el momento de la catalogación

- Una única transformación
- Problema de duplicidad de información
- Tiende a generar problemas de consistencia

Representación de recursos

Entidades abstractas



2. En el momento de la presentación

- Se requiere transformación cada vez que se muestra el recurso
- Mayor carga de procesamiento cada vez que se muestra el recurso
- Se evita la duplicidad de la información
- Se asegura la consistencia

Representación de recursos

Representación física de los datos



Es necesario analizar alternativas para el almacenamiento

- Performance
- Flexibilidad
- Escalabilidad

Algunas opciones:

- Base de datos XML (eXist)
- Base de datos relacional
- Base de datos orientada a objetos
- Base de datos RDF

Se pueden adoptar soluciones mixtas



Identificadores persistentes

Identificadores persistentes



¿Qué es un Identificador persistente?

Es un método de resolución de direcciones (URL) que busca garantizar el acceso a los objetos en internet, aún cuando éstos cambien su ubicación (URL de acceso).

Handle: hdl.handle.net/123456789/1234

DOI: dx.doi.org/10.4100/jhse.2010.52.15

PURL: purl.org/net/example/purlName

Identificadores persistentes



Importancia

Las URL cambian con el tiempo

- Dominio: cambios poco frecuente
- Ruta: en general cambios frecuente

El servicio se basa en redireccionar la solicitud de una URL persistente a una URL (no persistente) real, la que efectivamente apunta hacia el recurso.

Cuando la URL real del recurso cambia, se informa de este cambio solo al manejador de identificadores persistentes contratado y este modifica las reglas de redirección.

Identificadores persistentes

Algunas alternativas disponibles



10045/13546



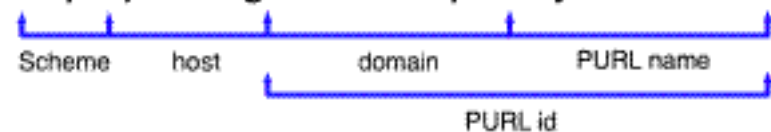
Handle

10.4100/jhse.2010.52.15



DOI

http://purl.org/net/example/myFirstPURL



PURL

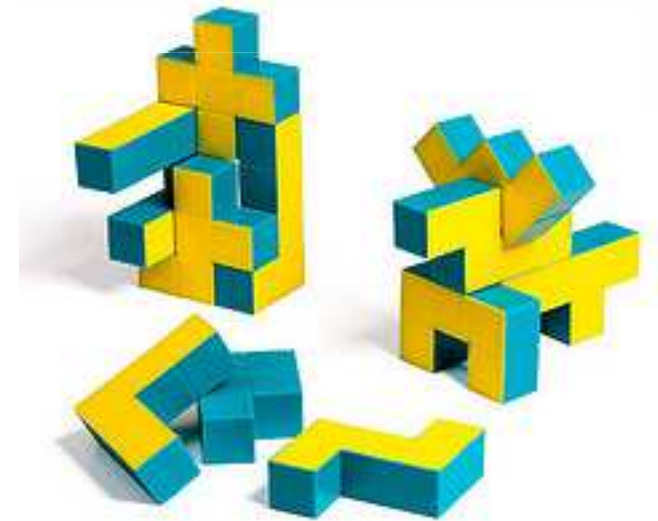


Servicios de un repositorio digital

Servicios de un Repositorio digital



- Búsqueda y Recuperación
- Exploración
- Diseminación Selectiva de Información
- Autoarchivo
- Servicios a otros sistemas



Servicios de un Repositorio digital



Búsqueda y Recuperación

- Un repositorio digital puede alojar cientos, miles o millones de recursos
- Es necesario proveer a los usuarios de mecanismos para buscar y recuperar estos recursos
- Los usuarios pueden estar buscando un recurso específico y único, o pueden requerir recursos con alguna característica en común (por ejemplo, artículos que traten sobre determinada área del conocimiento)
- A veces, los usuarios no saben bien que están buscando; suelen refinar los criterios de búsqueda una y otra vez hasta que localizan los recursos

Servicios de un Repositorio digital



Búsqueda y Recuperación

Un repositorio tiene que proveer un servicio de **búsqueda simple**, que permita ingresar algunos términos de búsqueda y retorne un conjunto de recursos como resultado

También debe proveer una **búsqueda avanzada**, que permita parametrizar los criterios de búsqueda y acotar así el conjunto resultante: por fecha de publicación de los recursos, por tipo de recurso, por idioma, por autor...

En cualquier caso, las búsquedas deben cumplir ciertos criterios mínimos:

Servicios de un Repositorio digital



Búsqueda y Recuperación

- Simpleza: el formulario de búsqueda debe ser simple, y mostrar campos de búsqueda avanzada si el usuario lo requiere. De todos modos, la búsqueda avanzada también debe permanecer simple
- Eficiencia: las búsquedas deben resolverse casi inmediatamente, en cuestión de milisegundos, o muy pocos segundos a lo sumo
- Relevancia: Todos los resultados de una búsqueda tendrán un valor de relevancia. Cuanto más relevante, más arriba deberá mostrarse entre los resultados

Servicios de un Repositorio digital



Búsqueda y Recuperación

- Filtrado: la búsqueda avanzada permite definir ciertos criterios a aplicarse durante la búsqueda
- En ocasiones, es deseable aplicar **filtros** una vez realizada la búsqueda
- Para ello, es necesario definir criterios de agrupamiento de resultados, y permitir al usuario agregar o eliminar criterios
- Una técnica muy utilizada es el *faceting* (*faceted search*, *faceted navigation* o *faceted browsing*), que permite a los usuarios explorar filtrando la información disponible en los resultados de la búsqueda

Servicios de un Repositorio digital

Búsqueda y Recuperación . Faceting



Buscar desarrollo de software 

Mostrando ítems 1-10 de 5181 1 2 3 4 ... 519 [Página siguiente](#) ▶

Desarrollo de software sensible al contexto

Quincoces, Viviana Elizabet; Gálvez, M.; Cáceres, N. R.; Vega, Ariel; Brouchy, C. V.; Velázquez, E. C.; González, O. M.; Guzmán, A. H. (2011)

Los Sistemas Informáticos fueron evolucionando desde aplicaciones científicas, comerciales y de escritorio, hasta el momento actual, en que pueden brindar servicios de acuerdo a la ubicación, tiempo y perfil del usuario. ...

Framework de mejora de procesos de desarrollo de software

Barbieri, Sebastián (2007)

El trabajo desarrolla un framework de mejora de procesos sobre organizaciones que realicen desarrollo o mantenimiento de software independientemente del tamaño de la organización. Este framework no está atado a un modelo ...

Automatización de procesos de desarrollo de software definidos con spem

Zorzán, Fabio; Riesco, Daniel Eduardo (2007)

Esta línea de investigación propone una alternativa para lograr la automatización de la gestión de los procesos de desarrollo de software especificados con el Software Process Engineering Metamodel(SPEM). La idea es utilizar ...

Tipo de documento

- Artículo (1310)
- Capítulo de libro (1)
- Comunicación (50)
- Contribución a revista (83)
- Documento de trabajo (40)
- Informe técnico (25)
- Libro (42)
- Objeto de conferencia (2561)
- Preprint (6)
- Revisión (99)
- Tesis de doctorado (505)
- Tesis de grado (194)
- Tesis de maestría (210)
- Trabajo de especialización (55)

Fecha de publicación

- 2000 - 2012 (4780)
- 1900 - 1999 (401)

Materia

- Ciencias Informáticas (2393)
- Educación (680)
- Humanidades (548)
- Ciencias Naturales (462)
- Ciencias Exactas (292)

Servicios de un Repositorio digital



Exploración

- Mediante la exploración, los usuarios pueden acceder a los recursos a partir de *un orden* preestablecido
- Este *orden* puede variar de repositorio en repositorio: colecciones, temas, fechas, etc.
- La exploración permite obtener un pantallazo general del repositorio

Servicios de un Repositorio digital

Exploración. Ejemplos



Colecciones en SeDiCI

Desde aquí usted puede navegar todas las colecciones de documentos disponibles en el repositorio

Tesis

Revistas

Eventos

Red UNCI

Publicaciones en revistas científicas

Acta Farmacéutica Bonaerense

Alp

Analecta Veterinaria

Anales de la Facultad de Ciencias Jurídicas y Sociales

Anuario del Instituto de Historia Argentina

Aportes para la Integración Latinoamericana

Archivos de Ciencias de la Educación

Archivos de Pedagogía y Ciencias Afines

Arkadin

Arte e Investigación

AUGMDOMUS

Auster



Servicios de un Repositorio digital

Exploración. Ejemplos



Navegar por autor

0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Mostrando ítems 1-60 de 20939

[Página siguiente](#) ►

Nombre de los autores	Nombre de los autores	Nombre de los autores
AA, Jiye	Abatedaga, Nidia	Abdala, Juan
Aamir, Muhammad F.	Abate, Stella Maris	Abdala, Lidia R.
Aamir, Muhammad N.	Abba, Agustín Manuel	Abdala, Virginia
Abad-Grau, María M.	Abba, Martín Carlos	Abdalla, Dulcinéia S. P.
Abadía, Anselmo	Abbas, Alaa M.	Abdelahad, Corina
Abadi, Florencia	Abbas, Ash Mohammad	Abdel-Hamid, Magdi
Abajo, Rosaura	Abbas, Ghulam	Abdel Masih, S.
Aballay, Alicia	Abbas, Khizar	Abdel Masih, Samira
Aballay, Laura	Abbas, Mateen	Abdo, Diego
Aballay, Laura N.	Abbas, Tanveer	Abdo Ferez, María Cecilia
Aballay, P.	Abbate, Florencia	Abdulameer, Shaymaa A.
Aballay, P.	Abbate, Horacio Antonio	Abdul Rasool, Bazigha K.
Abalo, Facundo	Abbate, Sandro Giuseppe	Abedini, Walter

Servicios de un Repositorio digital



Diseminación Selectiva de Información

- DSI es una técnica de envío de información de interés a los usuarios
- En un servicio DSI, los usuarios **solicitan** que se les envíe información
- Esta solicitud debe estar acompañada de algunos criterios de selección de información: temas, idiomas, tipos de recursos, períodos...
- En algunos casos, los usuarios pueden *suscribirse a búsquedas*; el software del repositorio ejecutará la misma búsqueda periódicamente, y enviará al usuario aquellos recursos que aparecen como nuevos entre los resultados

Servicios de un Repositorio digital

Diseminación Selectiva de Información



Google Scholar: Alertas por correo

Google scholar Search [Advanced Scholar Search](#)

Scholar [Create email alert](#)

[New York Times Co. v. Sullivan](#)

376 US 254, 84 S. Ct. 710, 11 L. Ed. 2d 686 - Supreme Court, 1964 - Google Scholar

... Thus we consider this case against the background of a profound national commitment to the principle that debate on public issues should be uninhibited, robust, and wide-open, and that it may well include vehement, caustic, and sometimes unpleasantly sharp attacks on ...

[Cited by 21634](#) - [How cited](#) - [Related articles](#) - [All 4 versions](#)

[Brown v. Board of Education](#)

349 US 294, 75 S. Ct. 753, 99 L. Ed. 1083 - Supreme Court, 1955 - Google Scholar

... The defendants in the cases coming to us from South Carolina and Virginia are awaiting the decision of this Court concerning relief. Full **implementation** of these constitutional principles may require solution of varied local school problems. ...

[Cited by 7052](#) - [How cited](#) - [Related articles](#) - [All 3 versions](#)

Criterios de búsqueda avanzada

Servicios de un Repositorio digital



Autoarchivo

- Es importante que todos los miembros de la organización se involucren con el repositorio. Una forma de hacerlo es que ellos mismos aporten su propia producción
- El servicio de *autoarchivo* permite a los miembros de la organización cargar sus propios recursos al repositorio
- De este modo, los autores se aseguran la publicación y difusión de sus trabajos en forma rápida y sencilla
- Este servicio implica la carga de un archivo, y una pre-catalogación del recurso por parte de quién realiza el autoarchivo
- La interfaz de catalogación debe ser muy simple, y se presenta un subconjunto de metadatos al usuario

Servicios de un Repositorio digital



Autoarchivo

- Existen restricciones en cuanto al tipo de archivo a enviar, y también en cuanto al tamaño de los mismos
- Los recursos enviados mediante autoarchivo quedan en un estado *pendiente de revisión*: debe hacerse un control de calidad sobre los recursos subidos, especialmente sobre aquellos subidos por personas no especializadas en catalogación
- Los autores deben seleccionar una licencia CC para su obra
- Los autores deben aceptar una licencia de difusión para SeDiCI

Servicios de un Repositorio digital

Autoarchivo



Envío de ítems

Describir → Describir → Adjuntar → Revisar → Licencia CC → Licencia → Finalizar

Tipo de documento:

Seleccione el Tipo de Documento que desea cargar

Artículo ▼

Autor (*):

Autores de la obra

Oviedo, Néstor



+ Agregar Otro

Título (*):

El título principal de la obra

Extract, Transform and Load architecture for metadata collection

Fecha de Publicación:

Fecha en la que la obra fue publicada en una revista, libro, etc. No debe confundirse con la fecha de entrega o defensa de una tesis, que debe cargarse en el campo Fecha de Presentación. Los valores posibles para este campo son día/mes, mes/año o día/mes/año.

15

mayo ▼

2011

Día

Mes

Año

Resumen:

Resumen de la obra

|

Servicios de un Repositorio digital

Autoarchivo



<http://e-archivo.uc3m.es/>

> Página Principal

E-Archivo. Guía general



- > Acerca de E-Archivo
- > Organización de Contenidos
- > Depósito de Contenidos
- > Navegación por Índices
- > Cómo Buscar
- > Usuarios
- > Utilidades
- > Comentarios y sugerencias
- > Contacto

Acerca de E-Archivo

¿Qué es E-Archivo?

E-Archivo, el Archivo Abierto Institucional de la Universidad Carlos III de Madrid, se crea con el investigador de la comunidad universitaria, en formato digital. Es un sistema en línea de acces

E-Archivo form incrementand conocimiento.



Biblioteca Universitaria
Universidad de Málaga

Vicerrectorado de Innovación y
Desarrollo Tecnológico

¿Qué se pue

E-Archivo pret artículos de re

¿Cuáles son

- RIUMA
- BÚSQUEDA AVANZADA
- SOBRE RIUMA
- MANUAL DE USO
- DERECHOS DE AUTOR
- CONTACTO
- SUGERENCIAS



Índice

INTRODUCCIÓN

PÁGINA PRINCIPAL

ACCESO A RIUMA

Crear una cuenta de usuario

Acceder a mi cuenta

BÚSQUEDA DE INFORMACIÓN

Depositar Documentos

Para depositar un documento es necesario ser docente o investigador de la Universidad de Málaga, estar registrado como usuario en RIUMA y disponer de la autorización correspondiente para realizar un envío. Por defecto, sólo tendrá permiso para depositar en las colecciones de su departamento, pero si necesita realizar un envío sobre otra colección puede solicitarlo al administrador de RIUMA mediante correo electrónico a riuma@uma.es.

El depósito de un documento supone la aceptación de la licencia de distribución de la Biblioteca de la UMA para permitir a RIUMA reproducir, traducir y distribuir su envío. Asimismo tendrá la posibilidad de asignar una licencia Creative Commons para seleccionar las condiciones de acceso y protección de su obra de usos indebidos (más información en los apartados 4.6 y 4.7 del Manual).

<http://riuma.uma.es/>

SeDiCI SERVICIO DE DIFUSIÓN
DE LA CREACIÓN INTELLECTUAL



<http://sedici.unlp.edu.ar>

Servicios de un Repositorio digital



Servicios a otros sistemas

- Un Repositorio Institucional no está aislado en el mundo: debe ser capaz de interactuar con otros sistemas y otros repositorios, de compartir recursos y de recuperar recursos remotos
- Esto aumentará la visibilidad del repositorio en la web y maximizará la difusión de los recursos
- El repositorio podrá también aumentar la cantidad de recursos disponibles para sus usuarios
- Algunos servicios comunes: OAI PMH, SRU/SRW, RSS

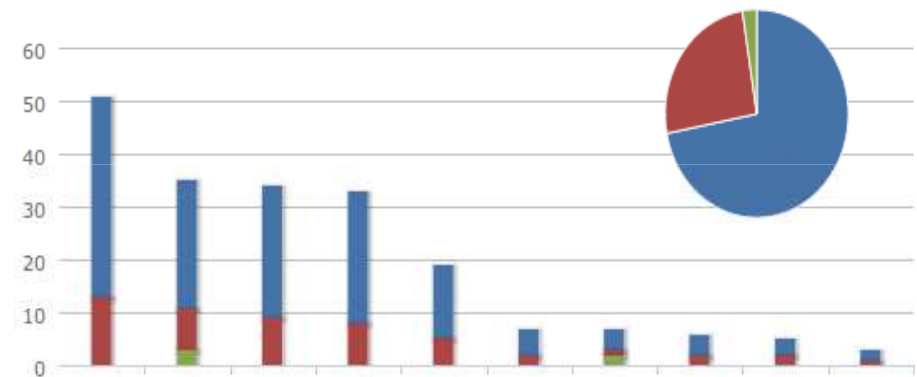
Estadísticas del repositorio



Necesidad e importancia

Clasificación de estadísticas

- a partir de la información que nos brindan
- a partir de quién las genera



Ejemplos

Estadísticas del repositorio



Necesidad de las estadísticas

- Las estadísticas son una herramienta clave a la hora de medir nuestro repositorio
 - Tamaño y Tasa de Crecimiento
 - Nivel de Impacto
- Obtener tablas y gráficos estadísticos avanzados, y no aprovechar esta información es casi lo mismo a no tener estadísticas
- El repositorio debe *retroalimentarse* con estos datos y utilizarlos bajo una política de *expansión y mejora continua*

Estadísticas del repositorio



Necesidad de las estadísticas

- Los datos obtenidos sirven como control de calidad, para saber dónde estamos parados como repositorio
- La interpretación de estos datos permitirá la toma de decisiones en varios niveles:
 - político/estratégico: cooperar con otros grupos, interactuar más con determinados actores...
 - táctico: cambiamos la forma de agrupar cierto tipo de recurso, incorporamos un nuevo tipo de recurso, implementamos una nueva metodología de carga
 - tecnológico: necesitaremos más hardware y mejor conectividad, debemos ampliar nuestro software para integrar cierta tecnología, será mejor revisar los índices de la base de datos...

Estadísticas del repositorio



Necesidad de las estadísticas. Tamaño y Tasa de Crecimiento

- Estadísticas de Tamaño y Tasa de Crecimiento
 - Necesitamos conocer cuántos recursos aloja nuestro repositorio
 - Es importante saber cómo han crecido estos recursos en el tiempo
 - de este modo, podemos detectar mesetas en las curvas de crecimiento y apuntalar donde sea necesario
 - podemos también predecir tendencias, como períodos de mayor o menor actividad, y prepararnos con antelación

Estadísticas del repositorio



Necesidad de las estadísticas. Tamaño y Tasa de Crecimiento

- El concepto de "tamaño" es muy amplio
 - cantidad de recursos locales
 - cantidad de recursos en full-text
 - cantidad de usuarios registrados
- Tasa de crecimiento también puede interpretarse de diferentes maneras
 - recursos incorporados año tras año
 - usuarios registrados cada semana
 - alertas por correo creadas mes a mes

Estadísticas del repositorio



Necesidad de las estadísticas. Tamaño y Tasa de Crecimiento

- Además de las cantidades mencionadas, tenemos otras "cantidades" de interés
 - Cantidad de Recursos locales
 - Tesis de grado, de posgrado
 - Artículos de revista, en congresos
 - Libros, e-books
 - Recursos a partir del origen
 - por dependencia, por departamento, área...
 - Por área temática
 - informática, ingeniería, literatura y letras, ciencias jurídicas...

Estadísticas del repositorio



Necesidad de las estadísticas. Tamaño y Tasa de Crecimiento

- Las clasificaciones nos permiten detectar *desequilibrios*
- Algunos desequilibrios son normales y esperables
 - "en el último año, se sumaron más de 2000 tesis de grado y solamente 50
⇒ libros" natural, considerando la cantidad de alumnos que se recibe anualmente
- Otros desequilibrios puede ser indeseables y podrían corregirse si se detectan a tiempo
 - "El 70% de los recursos proviene del 35% de las dependencias" ⇒ quizás debemos promocionar el uso del repositorio en el 65% restante, o quizás debemos adaptar el repositorio para que les sea de mayor utilidad
- Nuevamente, las estadísticas serán de utilidad si brindan información **precisa**, y si dicha información es **utilizada** apropiadamente

Estadísticas del repositorio



Necesidad de las estadísticas. Nivel de Impacto

- Nivel de Impacto: debemos medir el alcance global y local del repositorio
 - quiénes lo utilizan y para qué
 - desde dónde acceden los usuarios (países, regiones, instituciones)
 - cómo se posiciona en rankings y en buscadores
 - qué se busca y qué no se busca
 - con qué dispositivos y plataformas se accede (computadoras, tablets, sistemas operativos, navegadores)
 - a partir de cuáles servicios llegamos a nuestros usuarios (web, feeds, SRU/SWR, DSI, e-mail...)

Estadísticas del repositorio



Necesidad de las estadísticas. Nivel de Impacto

- Aquí también podremos tomar decisiones en niveles muy diversos:
 - Incorporar nuevos idiomas, a partir del origen de los usuarios
 - Optimizar las páginas web para maximizar su visibilidad en los buscadores
 - Reorganizar los contenidos para darles mayor relevancia a aquellos menos utilizados
 - Promocionar servicios con bajo nivel de uso
 - Desarrollar servicios, herramientas y estrategias para aumentar el acceso desde ciertos dispositivos
 - Mejorar las herramientas de búsqueda

Estadísticas del repositorio

Clasificación de estadísticas



Podemos clasificar las estadísticas a partir de dos grandes criterios:

- a partir del tipo de información que nos brindan
 - información sobre recursos, usuarios, servicios del repositorio, búsquedas realizadas, descargas ...
 - información del entorno o contexto: visitas, visibilidad en la web, navegadores utilizados, hardware de acceso...
- a partir del encargado de recolectarlas y generarlas
 - el software que sustenta al repositorio
 - otras herramientas integradas al repositorio
 - servicios de terceros

Estadísticas del repositorio

Clasificación de estadísticas



- A partir de Tipo de información
 - Información interna:
 - es específica para el repositorio
 - dependiente del software en uso
 - qué datos se almacenan
 - con cuánta granularidad
 - qué estadísticas se generan a partir de estos datos
 - podemos incorporar nuevas estadísticas y obtener datos mucho más precisos
 - recursos almacenados, usuarios registrados, accesos, servicios del repositorio, búsquedas realizadas, descargas

Estadísticas del repositorio

Clasificación de estadísticas



- Entorno o contexto:
 - obtenemos información acerca del entorno del repositorio
 - está muy relacionado con el nivel de impacto
 - **este entorno no es controlado por nosotros**
 - por lo general, no debemos preocuparnos por registrar estos datos

- Incluye cantidad de visitas al portal, visibilidad del portal en la web, tipos de navegadores utilizados, dispositivos desde los que acceden los usuarios

Estadísticas del repositorio

Clasificación de estadísticas



- Recolectadas y generadas por el mismo software
 - La recolección de datos debe estar en todos los rincones del software
 - Podremos controlar por completo las estadísticas, generar versiones más simples y más avanzadas, derivar nuevas estadísticas, etc...
 - Software más complejo
 - mayor dificultad de desarrollo y mantenimiento
 - importancia del diseño en capas
 - podría degradar la performance
 - diseñar un módulo de generación estadísticas no es una tarea simple

Estadísticas del repositorio

Clasificación de estadísticas



- Recolectadas por herramientas integradas al repositorio
 - El software que sustenta nuestro repositorio requiere otros programas para funcionar. Como mínimo, tendremos:
 - un sistema operativo, ej. Linux, Windows
 - un servidor web, ej. Apache, IIS, Tomcat, Jetty
 - una base de datos, ej. MySQL, Oracle
 - un servidor de correos, ej. Postfix, Exim

Estadísticas del repositorio



Clasificación de estadísticas

- Todos estos programas generan registros de acceso, de errores, de potenciales problemas (slow-log)... No nos preocupamos por guardar la información
- El desafío es cómo explotarla: debemos interpretarla, procesarla y mostrarla de manera útil (análisis de logs, minería de datos...)
- Afortunadamente, hay programas que realizan esto por nosotros
- Desafortunadamente, si bien podemos controlar parcialmente qué datos se registran, no tendremos la misma flexibilidad comparado con las estadísticas recolectadas por el software del repositorio

Estadísticas del repositorio

Clasificación de estadísticas



- Servicios de terceros
 - Como tercer alternativa, podemos tercerizar la recolección de estadísticas
 - Existen varios servicios externos capaces de recolectar y generar estadísticas
 - Puede requerir mínimos cambios en nuestro software, aunque a veces los sistemas están preparados para integrarse con algunos servicios populares
 - Aquí tendremos estadísticas de acceso, visibilidad, crecimiento del repositorio...
 - Algunos servicios son gratuitos, otros poseen una parte gratuita y otra paga, otros son solamente pagos

Estadísticas del repositorio

Ejemplos



Estadísticas de SeDiCI

Aplicaciones instalables

Awstats

Servicios on line

Google Analytics

StatCounter

Yahoo! Site Explorer

Rankings y registros globales

Webometrics

Roar <http://roar.eprints.org/1193/>



Preservación de contenido

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

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://www.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



Repositorio semántico

Repositorio semántico

Web semántica



Si bien en general se afirma que la web es una base de datos gigante, colaborativa, distribuida, en continuo crecimiento, etc, también existe consenso respecto de que esa base de datos tiene algunos problemas:

- Información mayormente semi-estructurada o completamente desestructurada
- Mucha información desactualizada
- Información redundante
- Información íntimamente relacionada, aunque sin ningún vínculo

Repositorio semántico

Web semántica



¿Por qué todos estos problemas?

Carga descentralizada

Flexibilidad (texto, imágenes, videos, etc)

Libertad de expresión



Repositorio semántico

Web semántica



El objetivo de la web semántica es aportar "**significado**" a toda la información disponible, de forma que sea "interpretable" por máquinas a través de agentes inteligentes, para así proveer información coherente, completa, competente, etc., de forma automática o semi-automática.

De esta forma se logra

- Aumentar la interoperabilidad entre sistemas
- Generar nuevos tipos de servicios de búsqueda y recuperación

Repositorio semántico



Los repositorios digitales cuentan con:

- Información estructurada (metadatos)
- Carga controlada (reglas de catalogación)
- Vocabularios controlados (tesauros, entidades abstractas)
- Base de datos centralizada (en general)



Repositorio semántico



Un repositorio semántico se caracteriza por la existencia de relaciones entre sus componentes (documentos, entidades, etc).

Algunas relaciones posibles son:

- Composición
- Traducciones
- Misma temática
- Autores relacionados
- Instituciones relacionadas

Repositorio semántico

Problemas relativos a la representación



Es necesario encontrar una forma flexible y eficiente para representar estas relaciones.

- Por inferencia, a través de relaciones establecidas en el modelo de datos
- De forma explícita, por ejemplo con **Ontologías**

Repositorio semántico

Problemas relativos a la representación



Por inferencia en base al modelo

Título: Función endotelial en el embarazo

Autor: Ros, Natalia

Descriptor: Cardiología; Embarazo

Título: Donantes de tejidos valvulares cardíacos: modelo de selección

Autor: Olano, Ricardo Daniel

Descriptor: Cardiología; Cultivo de tejidos

Título: Diferencias en la forma de presentación y diagnóstico de la enfermedad coronaria en la mujer

Autor: Corneli, Mariana

Descriptor: Cardiología; Enfermedades cardiovasculares

Autores relacionados con *Cardiología*

- *Ros, Natalia*
- *Olano, Ricardo Daniel*
- *Corneli, Mariana*

Repositorio semántico

Problemas relativos a la representación



Ontologías

Representa conceptos/objetos y las relaciones entre ellos

Las componentes mas importantes son:

- **Clases:** tipos de objetos
- **Instancias:** objetos concretos que pertenecen a una clase particular
- **Atributos:** características de una Clase (y por lo tanto de todas las instancias de esa clase)
- **Relaciones:** formas en las que los objetos se conectan entre sí

Repositorio semántico

Problemas relativos a la representación



Ontologías del dominio

Este tipo particular de ontologías se limita a representar elementos de un dominio particular, aportando un *contexto* a los conceptos/objetos que representa.

Ej.: el concepto *Ratón*

- En el dominio *Informática* hace referencia a un periférico de entrada
- En el dominio *Animales* hace referencia a un roedor

Repositorio semántico

Problemas relativos a la representación



Las ontologías se construyen usando un **lenguaje**.

Los más destacados son:

- **RDFS** Schema: Resource Description Framework Schema
Provee un conjunto de clases base, utilizando RDF como lenguaje de base.
- **OWL**: Ontology Web Language
Es una familia de lenguajes (OWL Lite, OWL DL y OWL Full), con varias sintaxis alternativas (una es una extensión al vocabulario de RDFS)

Repositorio semántico

Problemas relativos a la representación



Ejemplo de RDFS

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://www.animals.fake/animals#">

  <rdfs:Class rdf:ID="animal" />

  <rdfs:Class rdf:ID="horse">
    <rdfs:subClassOf rdf:resource="#animal"/>
  </rdfs:Class>
</rdf:RDF>
```

Repositorio semántico

Problemas relativos a la representación



Ejemplo de OWL (usando RDFS como sintaxis)

```
<rdf:RDF xmlns:owl = "http://www.w3.org/2002/07/owl#" ...  
  xmlns:rdfs = "http://www.w3.org/2000/01/rdf-schema#"  
  xmlns:xsd = "http://www.w3.org/2001/XMLSchema#">
```

```
<owl:Class rdf:ID="Animal">  
  <rdfs:label>Animal</rdfs:label>  
  <owl:Class rdf:ID="Male">  
    <rdfs:label>Male</rdfs:label>  
    <rdfs:subClassOf rdf:resource="#Animal"/>  
  </owl:Class>  
  <owl:Class rdf:ID="Female">  
    <rdfs:label>Female</rdfs:label>  
    <rdfs:subClassOf rdf:resource="Animal"/>  
    <owl:disjointWith rdf:resource="Male"/>  
  </owl:Class>  
</owl:Class>  
</rdf:RDF>
```

Repositorio semántico

Recuperación de la información y navegación de relaciones



Se necesita una forma de almacenamiento y recuperación eficiente

Existen:

- Bases de datos RDF (openRDF, Mulgara)
- Lenguajes de consulta RDF (SPARQL)

Además es necesario adaptar la interfaz de usuario para proveer elementos de navegación pertinentes, contextuales, eficientes, simples, etc.

Repositorio semántico

Posibilidad de nuevas estadísticas



El agregado de relaciones al repositorio permite generar un nuevo conjunto de estadísticas mas complejas, pero más interesantes:

- Tendencias en cuanto a temáticas
- Relación entre autores e instituciones
- Instituciones y autores mas productivas



SeDiCI SERVICIO DE DIFUSIÓN
DE LA CREACIÓN INTELECTUAL



Bibliotecas y Repositorios Digitales

Tecnología y Aplicaciones

<http://sedici.unlp.edu.ar>

Capítulo 5



Catalogación

Concepto

- Objetivo de la catalogación
- Reseña histórica
 - Ficha
 - Formatos bibliográficos (Marc)
 - Metadatos

Normalización

- ISBN
- ISSN
- ISBD
- AACR2
- FRBR
- RDA

Catalogación



Indización

- Lenguajes libre y controlados
 - Tesouro
 - Encabezamientos de materia
 - Palabras clave
 - Sistemas de clasificación

- Contol de autoridades
 - Puntos de acceso
 - Catálogo de autoridades

- Manual de procedimientos



Catalogación

La catalogación es un conjunto de operaciones destinadas a representar la forma y el contenido de un documento según normas estandarizadas.

- Análisis de la forma (descripción bibliográfica)
 - Título, autor, idioma
 - Puntos de acceso para su recuperación

- Análisis de contenido (indización)
 - Materias
 - Palabras clave
 - Resumen



Objetivos de la catalogación

- Identificar los documentos de forma inequívoca
- Dar uniformidad a la información
- Organizar y preservar el material alojado en una Institución
- Facilitar la ubicación, el acceso y la recuperación
- Permitir el intercambio con otras instituciones

Reseña histórica



- Ficha bibliográfica:
 - Tarjeta de papel o cartulina de forma rectangular y dimensiones pequeñas en la que se fija de forma sintetizada el contenido y los caracteres externos de un documento o libro
 - Se ordenan alfabéticamente y se reúnen en los catálogos
 - Por cada documento se hacen 3 fichas, cuyo contenido es el mismo pero cambia el encabezamiento: en las fichas de autor, el nombre de este aparece primero; en las fichas de materias, el contenido aparece primero y en las fichas de títulos, el título aparece en primer lugar.



Reseña histórica

- Ficha catalográfica con indicación de encabezamiento principal y secundarios

Santillana, Iñigo López de Mendoza, Marqués de ← Encabezamiento principal

Reflexiones entre amigos: Las cartas desconocidas de Iñigo López de Mendoza (Marqués de Santillana) enviadas a sus amigos/ reunidas por Stefan Johan von Foster; traducidas al castellano por Luisa d'Ors –edición de lujo–, Madrid [etc.] : Retamar, D.L. 2008, 235 pág.: il. en col.; 22 cm. + 1 CD-ROM
(Personajes de la Historia; 105. Epistolarios)
D.L. M 55.218-2008. - ISBN 978-84-28991-55-X

I. Foster, Stefan Johan von, comp. II. Ors, Luisa d', trad. III. Título: Reflexiones entre amigos: Las cartas desconocidas de Iñigo López de Mendoza (Marqués de Santillana) enviadas a sus amigos IV. Serie: (Personajes de la historia; 105. Epistolarios).] Encabezamientos secundarios

Reseña histórica



Registro bibliográfico

- Un registro bibliográfico es una presentación codificada y económica de un documento.
- MARC: el formato MARC (Machine Readable Cataloging) es un registro catalográfico legible por máquina.
- Fue desarrollado por la Biblioteca del Congreso de los Estados Unidos (Library of Congress) en la década de los '60.
- Es un conjunto de normas que permite almacenar información en registros de cualquier tipo, para posteriormente, poder tratarla, localizarla, intercambiarla o ponerla a disposición del usuario.
- Un registro MARC bibliográfico consiste en tres componentes principales:
 - Cabecera, Directorio y los campos variables.



Reseña histórica

- Esquema de metadatos
- **Metadatos:** Son datos que describen otros datos, información estructurada para describir, explicar, localizar o facilitar la obtención, uso o administración de un recurso de información.
- DC es un modelo de metadatos elaborado por la **DCMI** (Dublin Core Metadata Initiative). Actualmente es el esquema de datos más utilizado.

Contenido	Propiedad	Instanciación
	Intelectual	
Title	Creator	Date
Subject	Publisher	Type
Descripción	Contributor	Format
Source	Rights	Identifier
Lenguaje		
Relation		
Coverage		



Normalización

- Para normalizar los procedimientos, las bibliotecas disponen de herramientas que han sido creadas a tal fin, siendo las más importantes y difundidas:
 - Reglas de catalogación
 - Sistemas de clasificación
 - Lenguajes controlados
 - Normas ISO

Normalización



- La finalidad de usar normas técnicas en diferentes Unidades de Información y servicios de información es proveer compatibilidad e interoperabilidad entre equipos, datos, información, políticas y procedimientos.
- La creación de estándares acordes para las tecnologías de información contribuirán a permitir el intercambio de registros bibliográficos, anteriormente, o mejor dicho en sus inicios, intercambios manuales y ahora, actualmente, automatizados; a la vez nos permite optimizar los tiempos y los recursos humanos de las diferentes Unidades de Información y/o bibliotecas.
- La importancia de la aplicación de estándares en las bibliotecas nos permite estar mejor preparados y en condiciones óptimas a la hora del intercambio de información y/o de compartir recursos.



Lista de Normas

- Normas ISO
- ISBD
- AACR2
- FRBR
- RDA
- ISSN-ISBN
- Dewey
- CDU

ISBN



- El **ISBN** -International Standard Book Number- (norma ISO 2108) es un número creado para dotar a cada libro de un código numérico que lo identifique. Este número permite coordinar y normalizar la identificación de cualquier libro, utilizar herramientas informáticas para localizarlo y facilitar su circulación en el mercado editorial.
- Naturalmente, todo libro tiene un autor, un título, un editor, etc. que en sí mismos lo identifican. Sin embargo, la creciente producción de títulos a nivel mundial, obligó a utilizar un sistema automatizado para el control editorial. La norma original ha sido modificada a medida que los libros y otros artículos similares han empezado a publicarse en nuevos soportes, y la estructura básica del ISBN, que se define en la norma ISO 2108, ha sido modificada (2005) con la implantación del ISBN-13. El ISBN se usa actualmente en cerca de 150 países. Con esta numeración normalizada se pretende conseguir que cada obra publicada tenga un número de identificación único y simple: un "DNI" para cada libro.
- En la Argentina lo asigna la [Cámara del Libro](#)

ISSN



- El **ISSN** -International Standard Serial Number- es un código numérico reconocido internacionalmente para la identificación de las publicaciones seriadas, impresas o no. El ISSN puede utilizarse siempre que haya que recoger o comunicar información referente a las publicaciones seriadas, evitando el trabajo y posibles errores a la hora de transcribir el título o la información bibliográfica pertinente. El ISSN identifica sin ambigüedades ni errores la publicación seriada a la que va asociado. Es el equivalente para las publicaciones seriadas de lo que es el ISBN para los libros. Se rige por la norma ISO 3297 de 1975, cuya última actualización es de 2007.
- En la Argentina lo asigna el [CAYCIT \(Centro Argentino de Información Científica y Tecnológica.\)](#)

ISBD



- Descripción Bibliográfica Internacional Normalizada
- En 1969, el Comité de Catalogación de la IFLA (Federación Internacional de Bibliotecas Americanas) promovió la Reunión Internacional de Expertos en Catalogación que tomó una resolución de establecer normas que regularan la forma y el contenido de las descripciones bibliográficas



ISBD

- El resultado de este trabajo fue el nacimiento del concepto Descripción Bibliográfica Internacional Normalizada (ISBD).
- Publicada en 1971, la Descripción Bibliográfica Internacional Normalizada para Publicaciones Monográficas (ISBD(M)) fue la primera de las ISBD. A continuación se desarrollaron las ISBD para Publicaciones Seriadas, Material no Librario, Material Cartográfico, Libros Raros, Música Impresa y, más recientemente, Recursos Electrónicos

Áreas ISBD



- Las reglas de catalogación dividen la información en áreas:
- Área del título y mención de responsabilidad
- Área de la edición
- Área de los detalles específicos del material
- Área de publicación, distribución, etc.
- Área de la descripción física
- Área de la colección o serie
- Área de las notas
- Área del número normalizado y condiciones de disponibilidad



AACR2

- Desde principios del SXX la ALA (American Library Association) y la Library Association del Reino Unido trabajaron juntas para la elaboración de reglas
- La primera edición de las AACR (*Anglo-American Cataloguing Rules*) se publicó en 1967 en dos versiones, una bajo un texto norteamericano y la otra bajo uno británico, pero hubo desacuerdos entre ambos países.
- Recién en 1978 siguiendo los acuerdos posteriores de 1969 sobre Descripción Bibliográfica Internacional Normalizada (ISBDs) dentro de IFLA y junto al deseo de los países de habla inglesa de llegar a un acuerdo respecto a las reglas, se publicaron las AACR2



AACR2

- Hubo revisiones de las AACR2 en 1988, 1998 y 2002 - todas seguían básicamente la misma estructura de las AACR2 pero con reglas actualizadas que reflejaban los cambios experimentados a través del tiempo, tales como una nueva perspectiva sobre los recursos electrónicos, las publicaciones seriadas y los recursos integrados



AACR2

- Dividida en dos partes:
 - **Parte I:**
 - contiene las 8 áreas de ISBD
 - un capítulo para cada tipo de documentos
 - **Parte II:**
 - encabezamientos
 - títulos uniformes
 - referencias

FRBR



- La idea principal del modelo FRBR (Requerimientos Funcionales para los Registros Bibliográficos) es la siguiente: los documentos no son entidades independientes sino que se relacionan con otros documentos: versiones, ediciones, traducciones de la misma obra, etc.
- El modelo nos permite ver las relaciones de todas las variaciones dentro de una misma obra: todas las obras asociadas con una persona, todas las expresiones de la misma obra, todas las manifestaciones de la misma expresión y obras y expresiones relacionadas
- Publicadas en el año 1997 en la **Sección de Catalogación del Comité Permanente de la IFLA**, en la **63ª Conferencia General** de la misma institución



FRBR

- El modelo FRBR ayuda a relacionar los distintos tipos de entidades bibliográficas y permite de esta forma, un modelo más adecuado a las bases de datos en redes y le permite al usuario, navegar y relacionar la información de diferentes tipos de entidades, de forma más eficiente y eficaz para la satisfacción de sus necesidades de información. Es decir, el modelo FRBR permite agrupar de manera más eficiente los objetos que comparten el mismo contenido intelectual y artístico

FRBR



Las entidades representan los objetos claves de interés de los usuarios en los registros bibliográficos. Dichas entidades se dividen en tres grandes grupos:

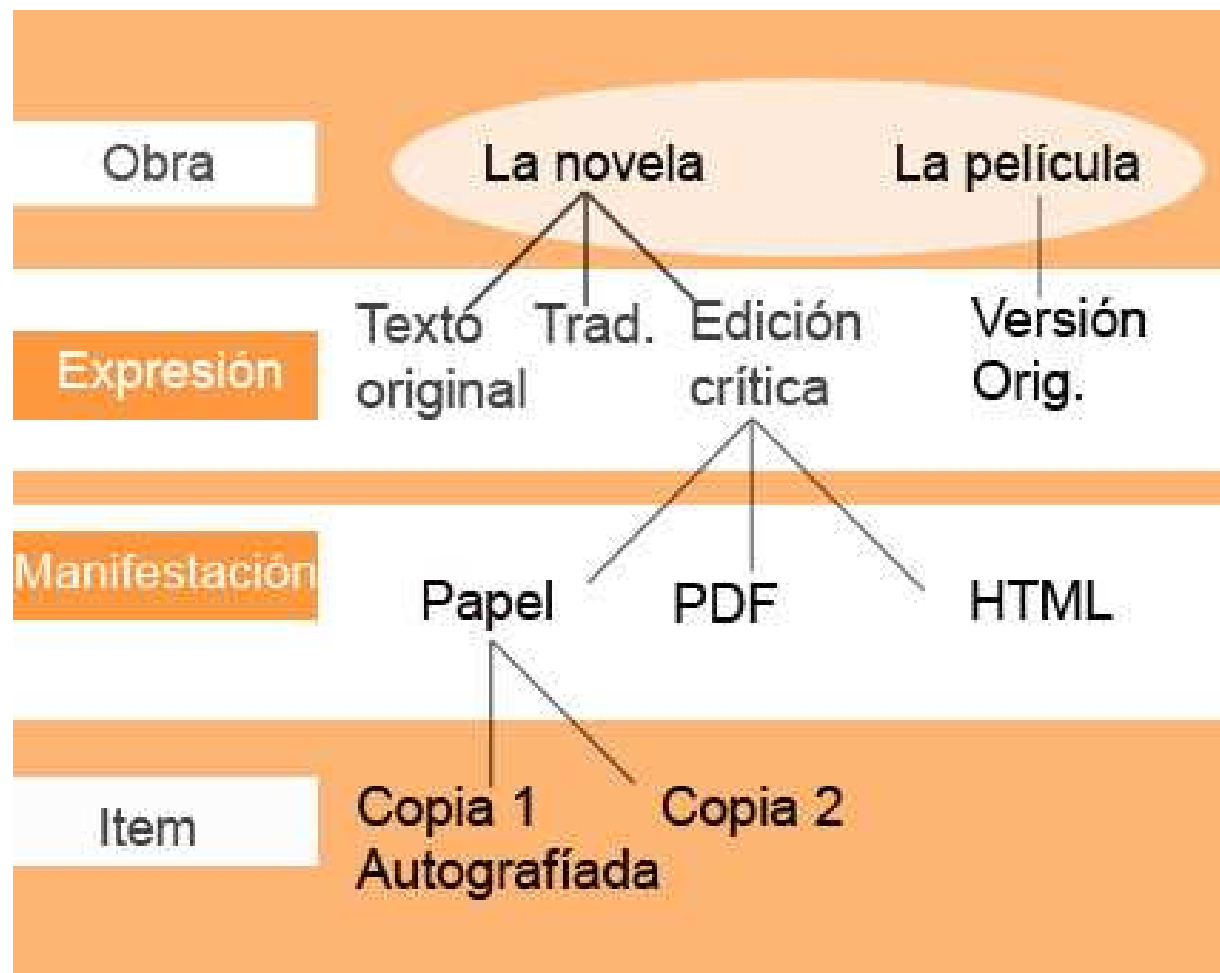
- El **Grupo 1** comprende los productos de los esfuerzos intelectuales o artísticos que se nombran o describen en los registros bibliográficos: obras, expresiones, manifestaciones e ítem. Forman la base del modelo y de su metodología.
- El **Grupo 2** abarca aquellas entidades responsables del contenido intelectual o artístico, de la producción física y de la difusión, así como de la conservación de dichos productos: personas y entidades corporativas.
- El **Grupo 3** incluye un conjunto adicional de entidades que sirven como materias a los esfuerzos intelectuales o artísticos: concepto, objeto, suceso y lugar. Estas entidades se emplean frecuentemente como materias de las obras. También, pueden referirse a una única obra o a multitud de ellas, y cada obra puede incluir varias entidades del tercer grupo.

FRBR



- Las entidades del Grupo 1 son la base del modelo:
- *Obra*: una creación intelectual o artística diferenciada, es una entidad abstracta; no hay un objeto material único que pueda denominarse la *obra*
- *Expresión*: la realización intelectual o artística de una obra en forma alfanumérica, musical, notación coreográfica, sonido, imagen, objeto, movimiento, etc., o cualquier combinación de dichas formas; es la forma específica recibe una obra cada vez que se “realiza”.
- *Manifestación*: la materialización física de la expresión de una obra.; abarca una amplia gama de materiales, que incluye manuscritos, libros, publicaciones periódicas, mapas, carteles, registros sonoros, películas, videograbaciones, CD-ROMs, kits multimedia, etc.
- *Ítem*: un ejemplar concreto de una manifestación, el objeto físico

FRBR



FRBR



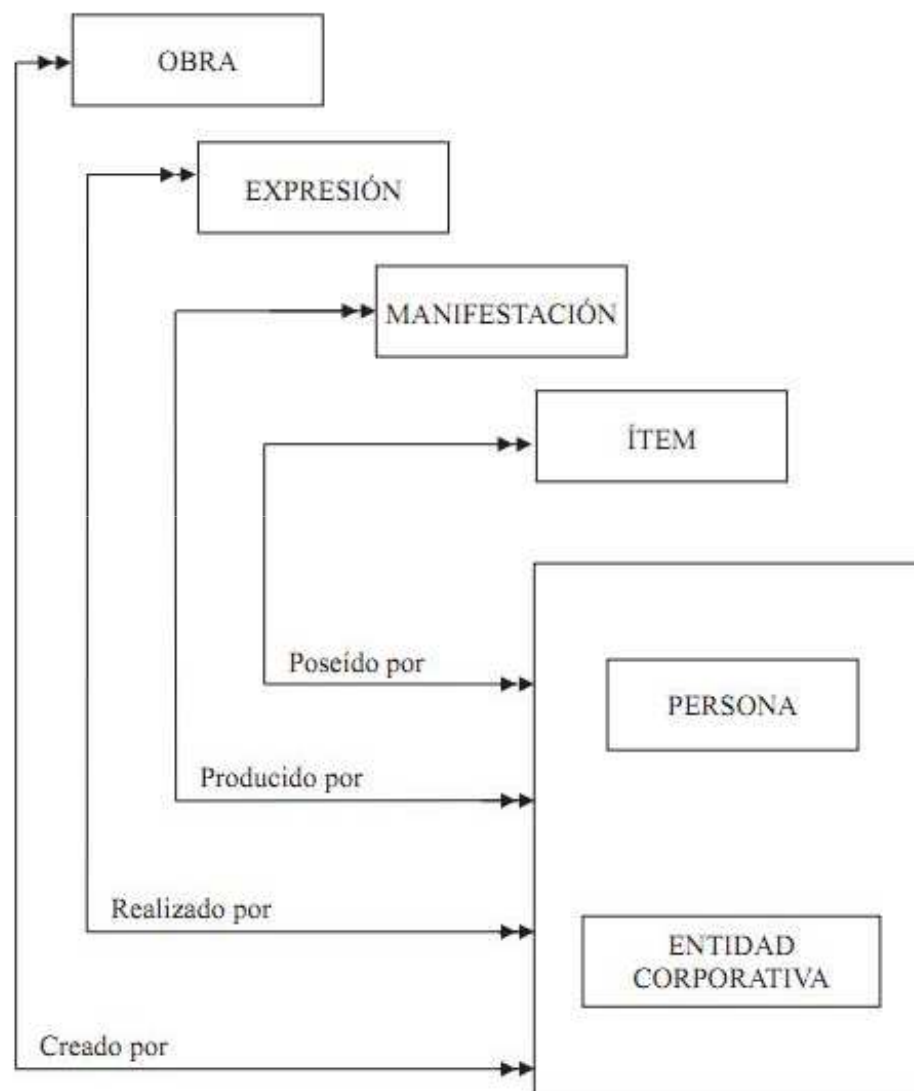
Gráfico 3.1. Grupo 1 Entidades y relaciones primarias



Fuente <http://archive.ifla.org/VII/s13/frbr/frbr-es.pdf>



Gráfico 3.2. Grupo 2 Entidades y relaciones de "responsabilidad"



Fuente
<http://archive.ifa.org/VII/s13/frbr/frbr-es.pdf>



RDA

- El proyecto RDA *Descripción y acceso al recurso* estipula que el universo documental está representado por recursos que adoptan las diferentes formas de comunicación en que el contenido puede ser expresado y comprendido: impreso, imagen, sonido y digital.
- Además establece lineamientos e instrucciones para la descripción y acceso a los recursos; cubre todos los tipos de contenido, medios y soportes físicos, y establece que los catálogos de bibliotecas u otros sistemas de información deben permitir a los usuarios el descubrimiento de recursos mediante las siguientes tareas: *encontrar, identificar, seleccionar y obtener recursos* apropiados para sus necesidades de información

RDA



- RDA es una nueva norma de catalogación que desde el 2004 se vienen estudiando y se dio a conocer en 2009.
- Supera al código de catalogación anterior (AACR2) dado que proporciona pautas para la catalogación de recursos digitales y pone un mayor énfasis en auxiliar al usuario a encontrar, identificar, seleccionar y obtener la información deseada. RDA favorece la agrupación de registros bibliográficos para mostrar las relaciones inherentes entre las obras y sus creadores (como las FRBR)



RDA

- Está desarrollado por el Joint Steering Comité (JSC) for Development of RDA, en conjunto con representantes de:
 - American Library Association
 - Australian Committee on Cataloguing
 - British Library
 - Canadian Committee on Cataloguing
 - Chartered Institute of Library and Information Professionals (CILIP)
 - Library of Congress

Indización – Análisis del contenido



La indización consiste:

- En la búsqueda, análisis y selección de los conceptos presentes en el documento.
- En la traducción de esos conceptos para pasar del lenguaje natural del autor al lenguaje documental.



Lenguaje documental

Un lenguaje documental es un sistema convencional de signos que permite representar el contenido de los documentos con el fin de encontrar aquellos pertinentes en respuesta a preguntas sobre un tema. Es sinónimo de lenguaje controlado, frente al lenguaje natural que es el que utiliza el discurso científico, técnico o literario.

Dicho discurso está cargado de metáforas, sinonimias y homonimias y necesita ser organizado y normalizado para permitir una recuperación eficaz del documento a salvo de los dos grandes problemas documentales: el ruido, -exceso de información encontrada no pertinente,- y el silencio,- ausencia de información pertinente que existe en la base de datos y no es recuperada.



Lenguaje documental: características

- Simplifica el lenguaje natural
- Evita la sinonimia y la polisemia
- Univoco: un único término y un único concepto
- Evitar la ambigüedad
- Algunos lenguajes documentales resuelven los problemas planteados por el multilingüismo (tesauros multilingües y lenguajes de clasificación numéricos)

Tipos de lenguajes documentales



- Lista de encabezamientos de materia:
- Son listas de términos (palabra o conjunto de palabras) ordenadas alfabéticamente. A estos términos se les denomina “encabezamientos de materia” y bajo cada uno de ellos se indican los términos sinónimos, genéricos o específicos con los cuales se relaciona.
- A partir de la determinación del tema principal de un documento se asigna el encabezamiento de materia que le corresponde para ese tema.
 - Library of Congress Subject Headings (LCSH)
 - LEMB (Lista de encabezamiento de materias para bibliotecas) ([demo](#))



Tesauros

- Los tesauros son listas de términos (palabra o conjunto de palabras) llamados “descriptores” que guardan entre sí relaciones semánticas de equivalencia, jerarquía o asociación . Estos términos generalmente se disponen conforme a dos presentaciones:
- **Parte alfabética:** los descriptores se ordenan alfabéticamente y bajo cada uno de ellos se indican los términos sinónimos, los términos genéricos y/o los términos específicos con los cuales se relaciona.
- **Parte sistemática:** los descriptores se agrupan jerárquicamente, siguiendo un esquema que va de lo general a lo particular.



Tesauros

- Los tesauros generalmente son elaborados por organismos internacionales y recogen los términos que se aplican en general para las distintas temáticas pero se excluyen términos regionales. Es por esta razón que surgen las listas de términos propuestos para complementar este lenguaje.
- Norma internacional ISO 2788 de 1986, para tesauros monolingües.
- Tesauros:
 - [DeCS](#)
 - [EUROVOC](#)

Tesaurus



EuroVoc Tesaurus multilingüe de la Unión Europea

Europa > Página inicial de EuroVoc > Campos temáticos & MT > 0406 marco político

Lengua del contenido:

(es) Español

Búsqueda

Búsqueda avanzada

Examinar

Examinar la presentación temática

Descargar

- Por campo temático
- Presentación alfabética permutada
- Lista multilingüe
- Índice alfabético
- SKOS/XML

Sus propuestas

- Contribuir
- Nuevos conceptos aprobados

0406 marco político

Estado

- NT1 Estado confederal
- NT1 Estado confesional
 - NT2 Estado islámico
 - RT Derecho musulmán [1206]
- NT1 Estado de Derecho
 - RT primacía del Derecho [1206]
- NT1 Estado del bienestar
 - RT bienestar social [2821]
 - RT política de intervención [1606]
- NT1 Estado federal
 - RT federalismo [0406]
- NT2 Estado federado
- NT1 Estado laico
 - RT laicismo [0406]
- NT1 Estado multiétnico
 - RT grupo étnico [2821]
- NT1 Estado regional
 - RT Parlamento regional [0421]
 - RT regionalismo [0406]
- NT1 Estado unitario
- NT1 símbolo del Estado
 - RT bandera [2831]

Fuente <http://eurovoc.europa.eu/>



Lista de descriptores libres

- *También llamados “Listas de términos propuestos”*: son listas de términos (palabra o conjunto de palabras) que elabora el bibliotecario según las necesidades de su biblioteca y de sus usuarios. Generalmente se utiliza para introducir términos o modismos de uso local o regional que no están contemplados en ningún tesoro editado.
- Por ej. : villas miseria, chicos de la calle, etc.



Palabras clave

- Vocablos extraídos del lenguaje natural
- Se extraen del título o del contenido del documento y carecen de sistematización



Sistemas de clasificación

Sistemas de clasificación decimal: son sistemas jerárquicos que dividen al conocimiento en 10 clases principales que a su vez se subdividen de 10 en 10 y así sucesivamente creándose notaciones más específicas. A partir de la determinación del tema principal de un documento se asigna el número que corresponde para ese tema. Las clases y sus divisiones se representan a través de una notación numérica.

Los sistemas de clasificación más difundidos son: Clasificación Decimal de Dewey (CDD) y Clasificación Decimal Universal (CDU).

CDD	
Notación	Temas que abarca
000	Obras generales
100	Filosofía
200	Religión
300	Ciencias Sociales
400	Lengua
500	Ciencias naturales y matemática
600	Ciencias aplicadas
700	Arte. Deporte. Ocio
800	Literatura
900	Geografía e historia

Así, por ejemplo, un libro sobre *Geología submarina* se clasificará en 551.460 8

Un libro sobre “Edificios e instalaciones oficiales de enseñanza media”



<i>Lenguaje documental usado</i>	<i>Traducción al lenguaje documental</i>
Sistema de clasificación decimal	371.6
Lista de encabezamientos de materias	ARQUITECTURA ESCOLAR
Tesauro	ESCUELAS ESPACIOS EDUCATIVOS PLANIFICACION DE LAS INSTALACIONES NORMAS DE CONSTRUCCION DISEÑO ARQUITECTONICO
Lista de términos libres	ESCUELAS-RANCHO



Puntos de acceso

Término o términos que posibilitan la búsqueda y localización de un documento o grupo de documentos en el catálogo.

Tres puntos de acceso principales:

- Autor (personal o entidad)
- Título
- Materia



Control de autoridades

- Para que un punto de acceso de cualquier tipo cumpla eficazmente su función, es preciso que sea sometido a lo que se denomina gestión o control de autoridades.
- Se puede definir como tal al conjunto de labores de normalización que determinan, de entre todas las variantes posibles, la forma con la cual un punto de acceso debe figurar en todo momento en el catálogo y que establecen, mediante la confección de referencias, reenvíos desde las variantes no aceptadas a la forma admitida y relaciones con otros puntos de acceso ya autorizados

Control de autoridades



Unifica los punto de acceso idénticos, pero que están expresados de forma distinta.

Cicerone, M. Tullio , Cicero Marcus Tullius,
Cicero, M.T., Cicerón M.T., Tullius Cicero,
Marcus, etc.

Cicerón, Marco Tulio

Permite diferenciar puntos de acceso distintos que se pueden encontrar de la misma manera (homónimos).

- Santoro, Daniel
 1. Santoro, Daniel, 1954- (artista plástico)
 2. Santoro, Daniel, 1955- (Periodista argentino, editor de la sección política del Clarín, especializado en casos de corrupción. Presidente del Foro de Periodismo Argentino (FOPEA))

Objetivos del control de autoridades



- Convertir el catálogo bibliográfico en una red de relaciones que permitan al usuario moverse con seguridad desde formas no aceptadas como autoridad, pero posibles, a las formas aceptadas, y desde formas aceptadas a otras aceptadas relacionadas, mediante el sistema de referencias
- Dirigir al usuario, mediante notas explicativas, a una clase, categoría o grupo de puntos de acceso que pueden hacerle mejorar su búsqueda de información en el catálogo

Catálogo de autoridades



Es un conjunto organizado de registros de autoridad que contienen los datos relativos a los puntos de acceso de un catálogo bibliográfico

Qué información contiene un catálogo de autoridades:

- Forma autorizada (autoridad)
- Uno o más formas no autorizadas (variantes).
- Fuentes de información sobre la forma autorizada (fuentes)
- Fuentes de información sobre las formas no autorizadas (fuentes)
- Notas de información hacia referencias (véase, véase además).
- Notas hechas por el catalogador durante el proceso de creación de la autoridad (datos biográficos, administrativos, etc.).

Catálogo de autoridades



- [Biblioteca Nacional de España](#)
- [Library of Congress Authorities](#)
- [Propuesta de manual de ayuda a los investigadores españoles para la normalización del nombre de autores e instituciones en las publicaciones científicas](#)



Manual de procedimientos

- Un manual de procedimientos tiene la función de servir de guía para el bibliotecario y para todos los que se interesen, en los futuros procedimientos inherentes.

En el manual se detallan acciones, previamente consensuadas, para unir criterios de trabajo. El manual plasma la política de la institución que lo produce, deja asentado la manera de trabajar, tanto a corto como a largo plazo.

- El manual debe ser revisado constantemente, para ajustarse a las necesidades cambiantes y seguir el paso de las nuevas innovaciones.

Manual de procedimientos



OBJETIVOS

- Procesar y diseminar la información recibida
- Actualizar permanentemente al personal profesional y administrativo
- Garantizar la uniformidad de tratamiento de las actividades periódicas
- Reducir los errores operativos
- Reducir el período de adiestramiento de los nuevos empleados
- Facilitar el mantenimiento de un buen nivel organizacional

Manual de procedimientos



Inicio Mi página Proyectos Ayuda Conectado como **silvia** Mi cuenta Desconexión

SeDiCI » SeDiCI-Docs

Búsqueda: SeDiCI-Docs

Vistazo Actividad Peticiones Nueva petición Noticias **Wiki** Foros

Manual de Procedimientos de SeDiCI actualizado a junio de 2011

Modificar Monitorizar Bloquear Renombrar Borrar Histórico

En él se encuentran los pasos necesarios para cargar cualquier tipo de recursos y los recaudos que deben tomarse en cada caso. Contiene capturas de pantalla de CelsiusDL para futuros sedicianos. También incluye una tabla con ejemplos reales de carga según el tipo de recurso.

En otro archivo adjunto se encuentran las Reglas de Ingreso de Nombres y Apellido citadas en el manual.

1. [Metadatos Obligatorios](#)
2. [Metadatos Especificos](#)
3. [Tutoriales](#)
4. [Traducciones](#)

MANUAL_DE_PROCEDIMIENTOS_SeDiCI_junio_2011.docx (7,5 MB) Ariel Matias Roberto Miten, 21 October 2011 11:51 AM

Reglas_de_Ingreso_para_Apellidos_y_Nombres.docx (16,6 KB) Ariel Matias Roberto Miten, 21 October 2011 11:51 AM

Nuevo fichero

Exportar a: HTML | TXT

Wiki

[Página principal](#)
[Índice por título](#)
[Índice por fecha](#)



Manual de procedimientos

Autor

Metadato	Ejemplo	Observaciones
Autor	Aplica a todos los documentos	Se cargará nombre y apellido del o de los autores. Los apellidos que comiencen con minúscula (como di Marco, Marcelo) serán cargados como Marco, Marcelo di . Se respetarán las <u>grafías</u> y los <u>acentos</u> . Se evitará siempre colocar tanto el nombre como el apellido en mayúsculas (es decir, RICHARD, DIEGO). Para una mejor carga, consultar las Reglas de Ingreso de apellidos y nombres en documento aparte.

Índice

Título del documento

Metadato	Ejemplo	Observaciones
Título del documento	Aplica a todos los documentos	Se copiará el título del documento tal como aparece en el mismo; se pondrá mayúscula al inicio y NO se colocará punto al final. En lo posible se evitará copiar y pegar para que no aparezcan enters. En todos los casos, cuidar la ortografía. No se usará en ningún caso el título a la norteamericana, es decir, Todo Con Mayúsculas. En el caso que deban colocarse subíndices o superíndices, o cualquier otro símbolo se recurrirá, en la medida de lo posible, al HTML .----- ----- En el siguiente link se encontrará una aplicación para evitar los dobles enters y respetar los símbolos que sea necesario: ☞ http://sedici.unlp.edu.ar/norm.php . Simplemente deberá pegarse el texto del título (también funciona para los abstracts) y cliquear en "enviar". <i>Estudio experimental y de cálculos de primeros principios del semiconductor Sc₂O₃ dopado con 111In → 111Cd, 181Hf → 181Ta y 44Ti → 44Sc</i> En la administración deberá cargarse como sigue: <i>Estudio experimental y de cálculos de primeros principios del semiconductor Sc<SUB>2</SUB>O<SUB>3</SUB> dopado con <SUP>111</SUP>In → <SUP>111</SUP>Cd, <SUP>181</SUP>Hf → <SUP>181</SUP>Ta y <SUP>44</SUP>Ti → <SUP>44</SUP>Sc</i>

Manual de procedimientos



Autor institucional	Facultad de Ciencias Exactas	Se pondrá siempre la unidad académica a la que pertenece el tesista.
fecha de Presentación	2009	Se usará siempre, para las tesis, "fecha de presentación" y no "fecha de publicación".
Director de tesis	Rentería, Mario	En caso de que se desconozca, se procurará obtener este dato. Si la búsqueda resultare infructuosa, se omitirá.
Co-director de la tesis	Errico, Leonardo Antonio	En caso de que se desconozca, se procurará obtener este dato. Si la búsqueda resultare infructuosa, se omitirá.
Miembro del jurado		En el caso de que la tesis declare quiénes fueron los jurados, se colocará en este campo.
Materias	Ciencias Exactas; Física	Se seleccionarán entre dos y tres materias (☐ grandes grupos temáticos seleccionables en el tesauro indicado con el número 5) que describan temáticamente el recurso en cuestión.
Descriptor	Materia condensada; Materiales; Cálculo	Se seleccionarán entre tres y cinco descriptores (términos controlados) abordados por la tesis. Para localizarlos se utilizarán los tesauros disponibles en SeDiCI, Eurovoc (número 6) y Decs (número 7). Sólo si estos no tuvieren ningún descriptor disponible se usará el tesauro 4. Se evitará el uso del tesauro de la UNESCO, sindicado con el número 1.

Bibliografía



Delgado, E. A., Organización documental mediante la catalogación y el análisis de información: entorno normativo y tecnológico. Revista Códice, 2007, Vol. 3 No. 2, p. 35-50. Disponible en: <http://redalyc.uaemex.mx/pdf/953/95330204.pdf>

Farfán Caudillo, M. A., Descripción y acceso al recurso: nuevo código de catalogación: Full draft of RDA. Boletín del Instituto de Investigaciones Bibliográficas, 2008, Vol 13, No 1-2, p. 363-371. Disponible en: <http://www.revistas.unam.mx/index.php/biib/article/view/24295/22829>

Gavilán, C. M., Temas de Biblioteconomía: El documento y sus clases: análisis documental: indización y resumen. 2009. En E-LIS: E-Prints in Library and Information Science. Disponible en <http://eprints.rclis.org/bitstream/10760/14605/1/tipdoc.pdf>

Gavilán, C. M., Temas de biblioteconomía: La normalización de la identificación bibliográfica: ISBD, ISBN, ISSN. 2009. En E-LIS: E-Prints in Library and Information Science. <http://eprints.rclis.org/bitstream/10760/14250/1/normaliza.pdf>

Bibliografía



Herrero-Pascual, C. (1999). El control de autoridades. Anales de Documentación, 1999, No 2, p. 121-136. Disponible en:

<http://redalyc.uaemex.mx/src/inicio/ArtPdfRed.jsp?iCve=63500208>

Murillo Madrigal, A., González del Valle, L.

Introducción bibliotecológica : procesos técnicos y soportes de información.

Buenos Aires : Ministerio de educación, ciencia y tecnología, Biblioteca nacional de maestros, 2004. Disponible en:

http://www.bnm.me.gov.ar/redes_federales/publicaciones/doc/cuadernillo_1.pdf

Requisitos Funcionales de los Registros Bibliográficos: informe final. 2004.

Disponible en: <http://archive.ifla.org/VII/s13/frbr/frbr-es.pdf>

Rodríguez García, A. A. Elementos para reflexionar sobre el nuevo estándar para la descripción y acceso a recursos. Biblioteca Universitaria, 2010, Vol. 13, No 1, p. 55-63. Disponible en: <http://redalyc.uaemex.mx/redalyc/pdf/285/28516679005.pdf>



SeDiCI SERVICIO DE DIFUSIÓN
DE LA CREACIÓN INTELECTUAL



Bibliotecas y Repositorios Digitales

Tecnología y Aplicaciones

<http://sedici.unlp.edu.ar>



Participantes del dictado

- Marisa De Giusti
- Nestor Oviedo
- Silvia Pelоче
- Matías Cánepa

Bibliotecas y repositorios digitales



Capítulo 6: La preservación en el repositorio. Modelo de Referencia para un Sistema Abierto de Archivo de Información: Norma ISO 14721. Los metadatos de preservación. El esquema PREMIS. Las herramientas del software que soporta el repositorio SeDiCI: DSPACE. Sus facilidades para implementar OAIS y PREMIS.



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**. ^{E1}
- 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.

Diapositiva 16

E1

Comunidad designada: un grupo identificado de consumidores potenciales que es capaz de comprender un conjunto particular de información.

ExpeUEW7, 18/09/2012



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 OAI

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

Conceptos en OAIS

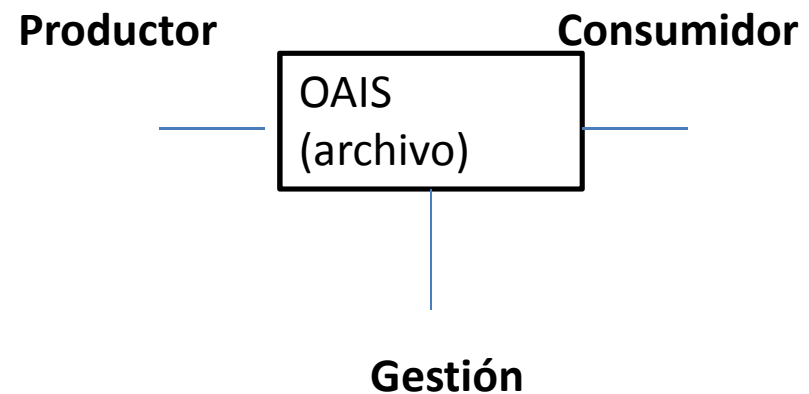


Medioambiente OAIS

- 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

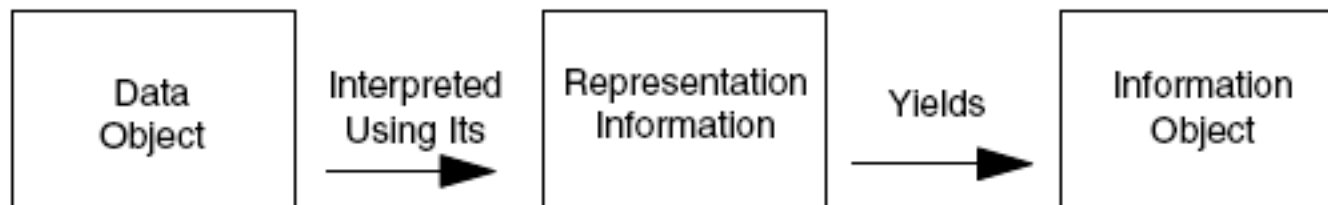


Fuente [ISO 14721](#)



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

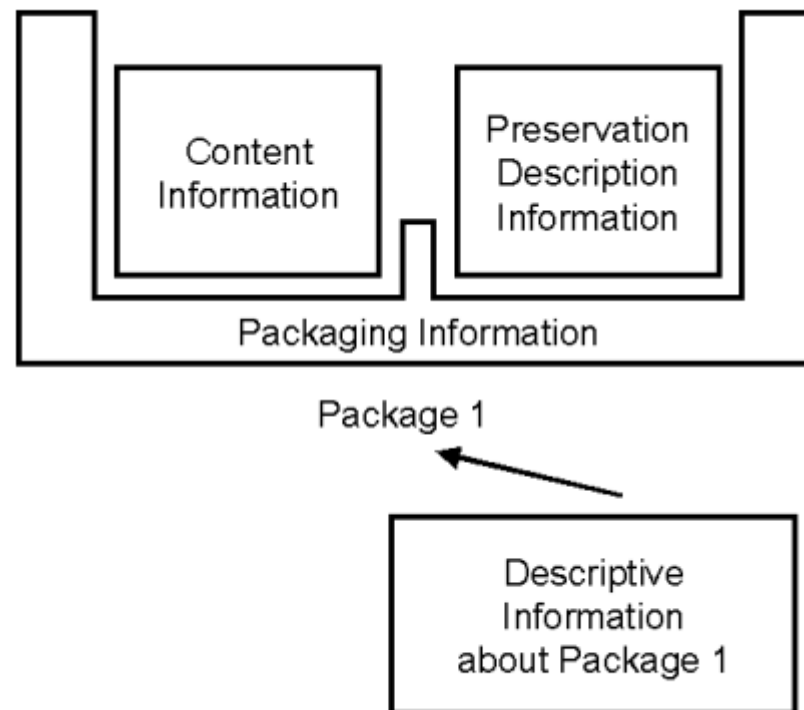


Fig 2-3: Paquete de información: conceptos y relaciones



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

Información de contenido (CI)

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

Información de
representación (RI)
Sobre los objetos de contenido

Información descriptiva de preservación (PDI)

Information de procedencia

Información de contexto

Información de referencia

Información de integridad

Información de empaquetado (PI)

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

Información descriptiva (DI)

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

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 /



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



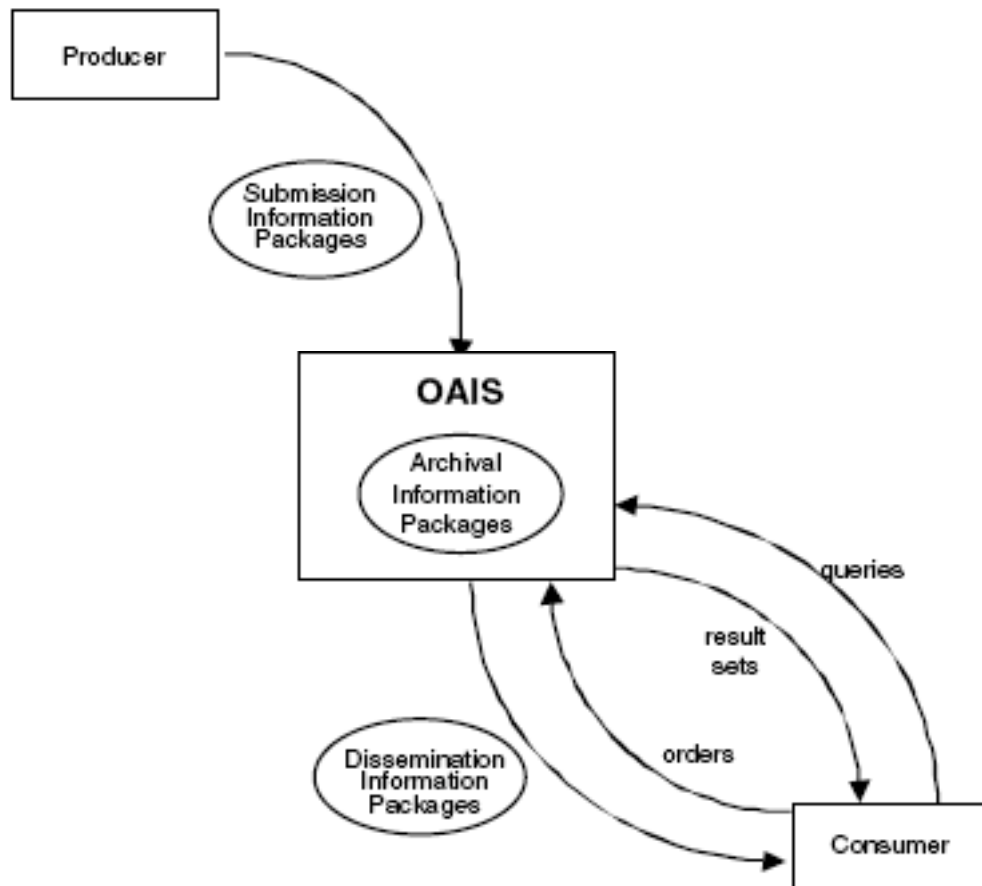
- Dissemination Information Package (DIP): es el paquete que se entrega a un consumidor en respuesta a una solicitud. La información de empaquetado toma muchas formas dado que los usos de OAIS son diversos, puede ser tan completo como los AIP a partir de los cuales se construye o ser sólo una breve descripción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



Fuente [ISO 14721](http://www.iso.org/iso/14721.html)

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 OAIS



- 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.

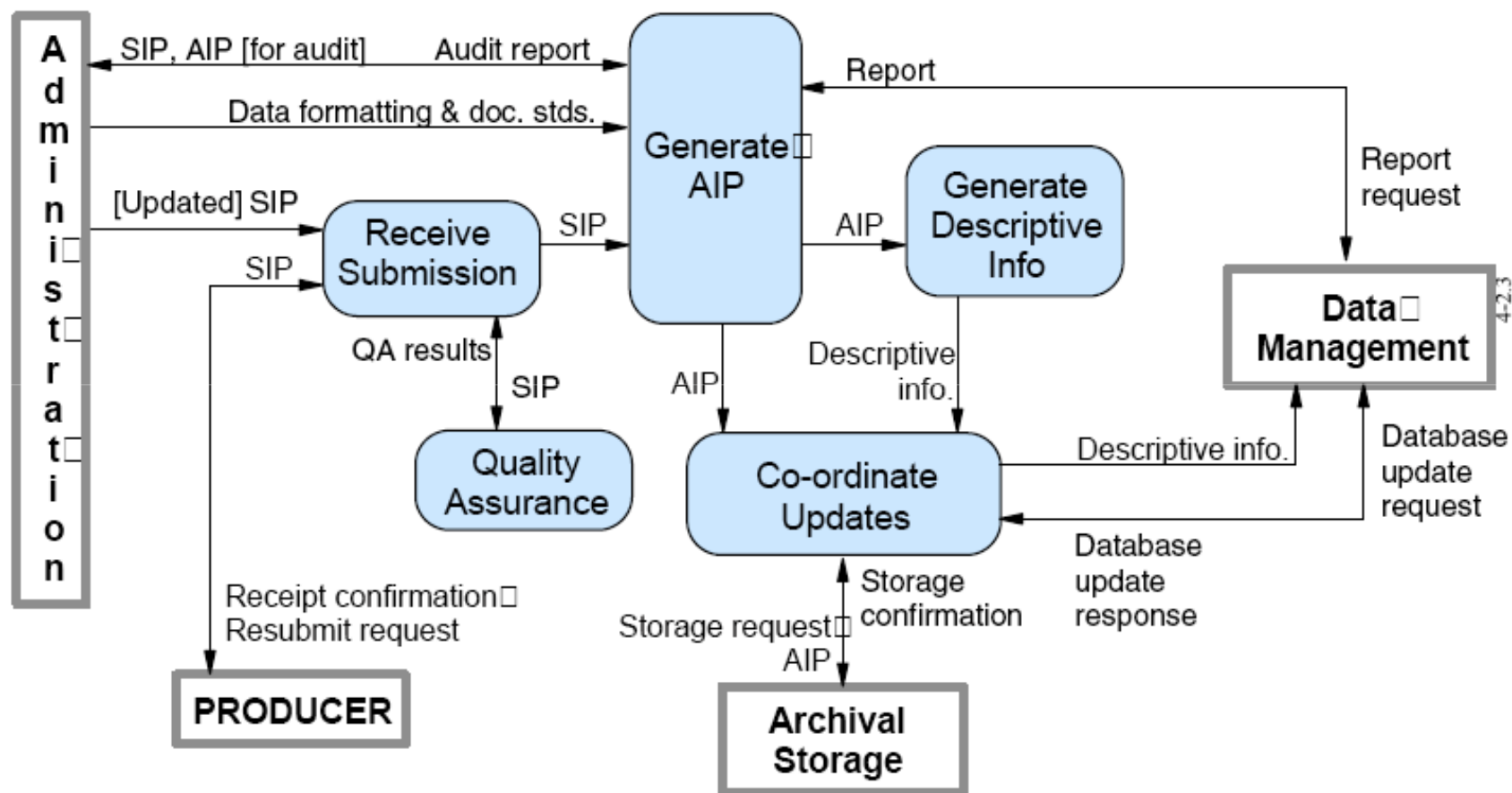
Ingesta/Ingest/presentación



- 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*



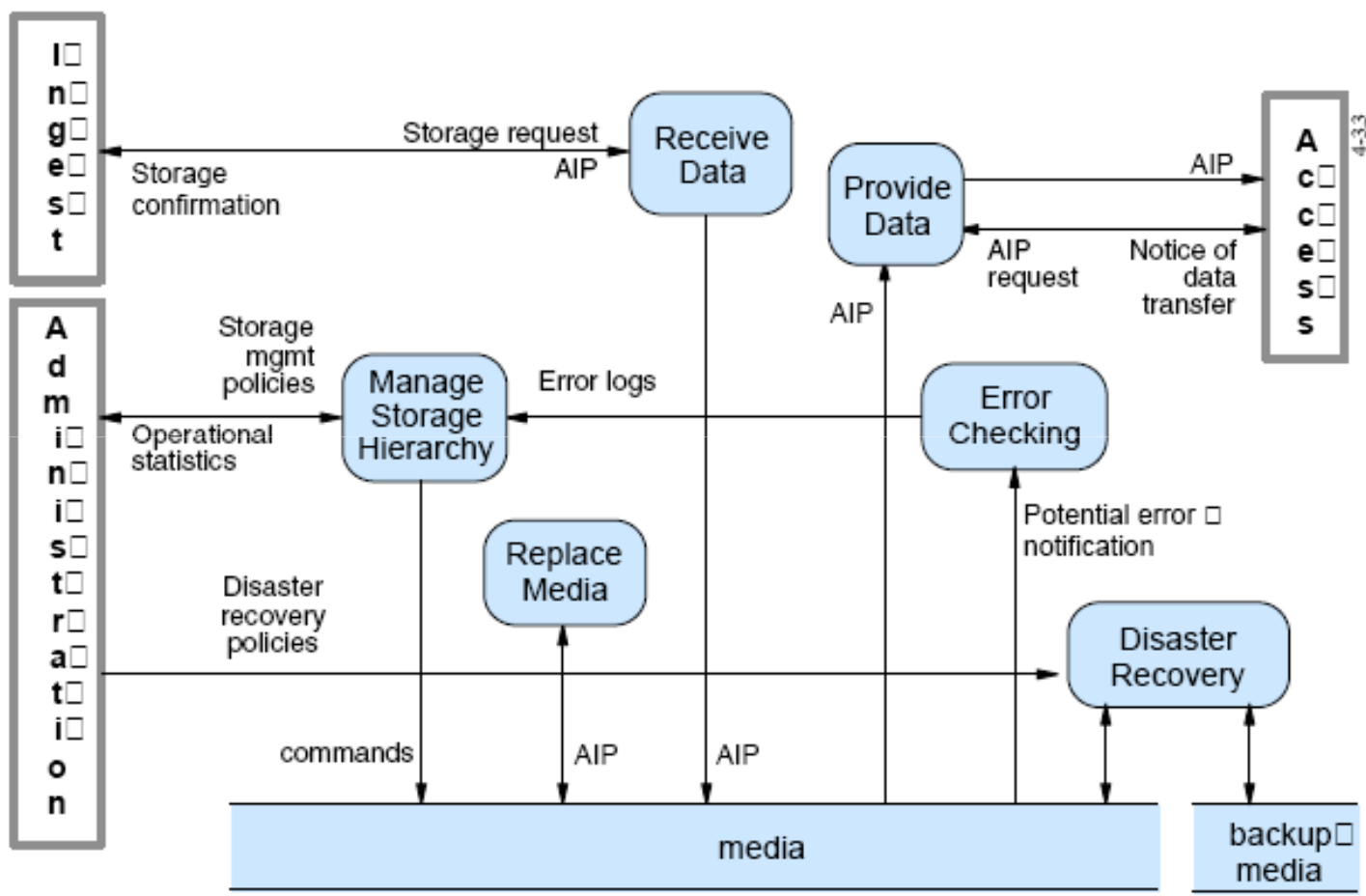
Entidad OAIS *Ingest*



- **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*



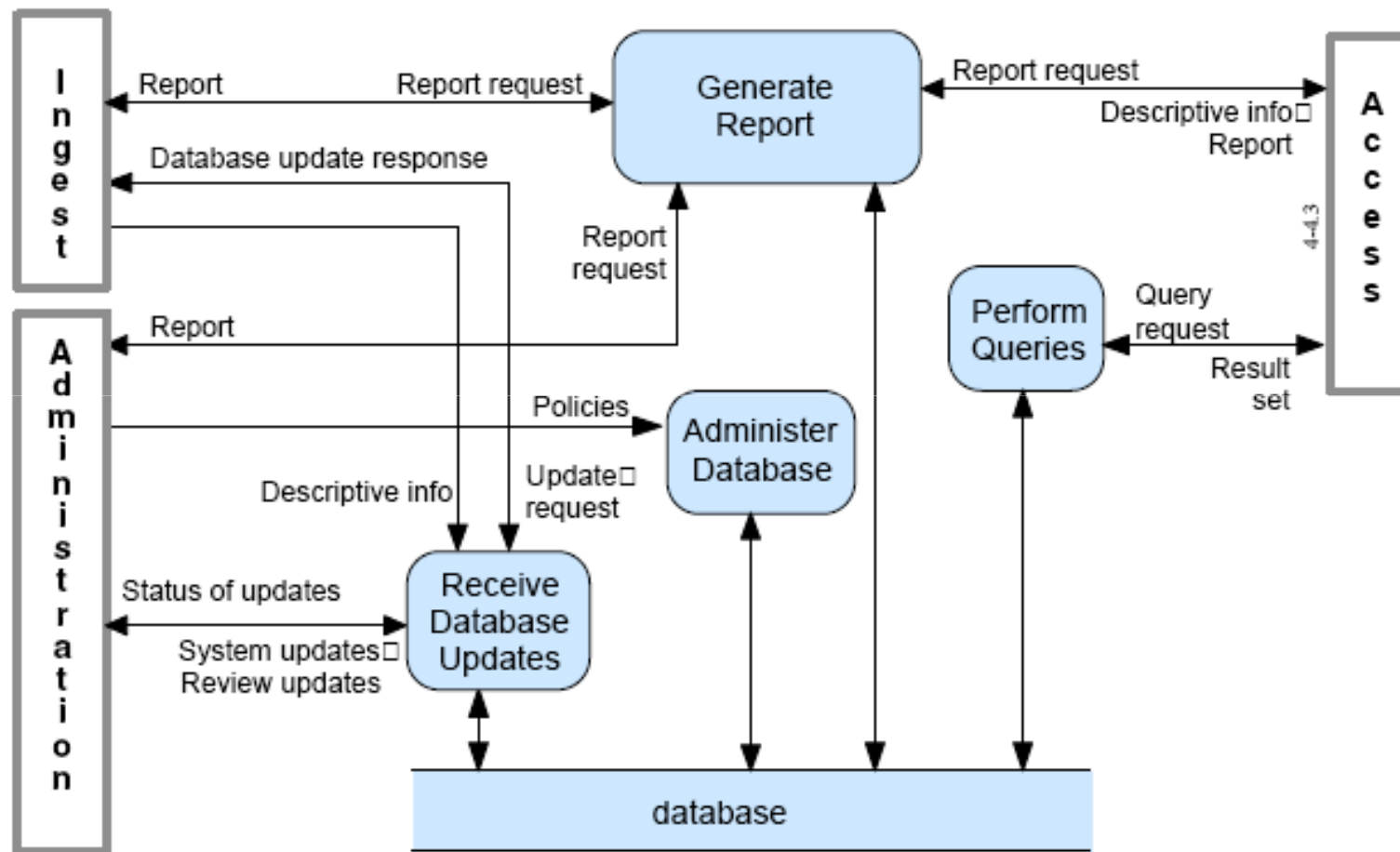
Fuente [ISO 14721](#)

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 una lugar físico separado. Provee copia de los AIPs almacenados a la entidad *access*.

Functions of *Data Management*



Fuente [ISO 14721](#)

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*.

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.

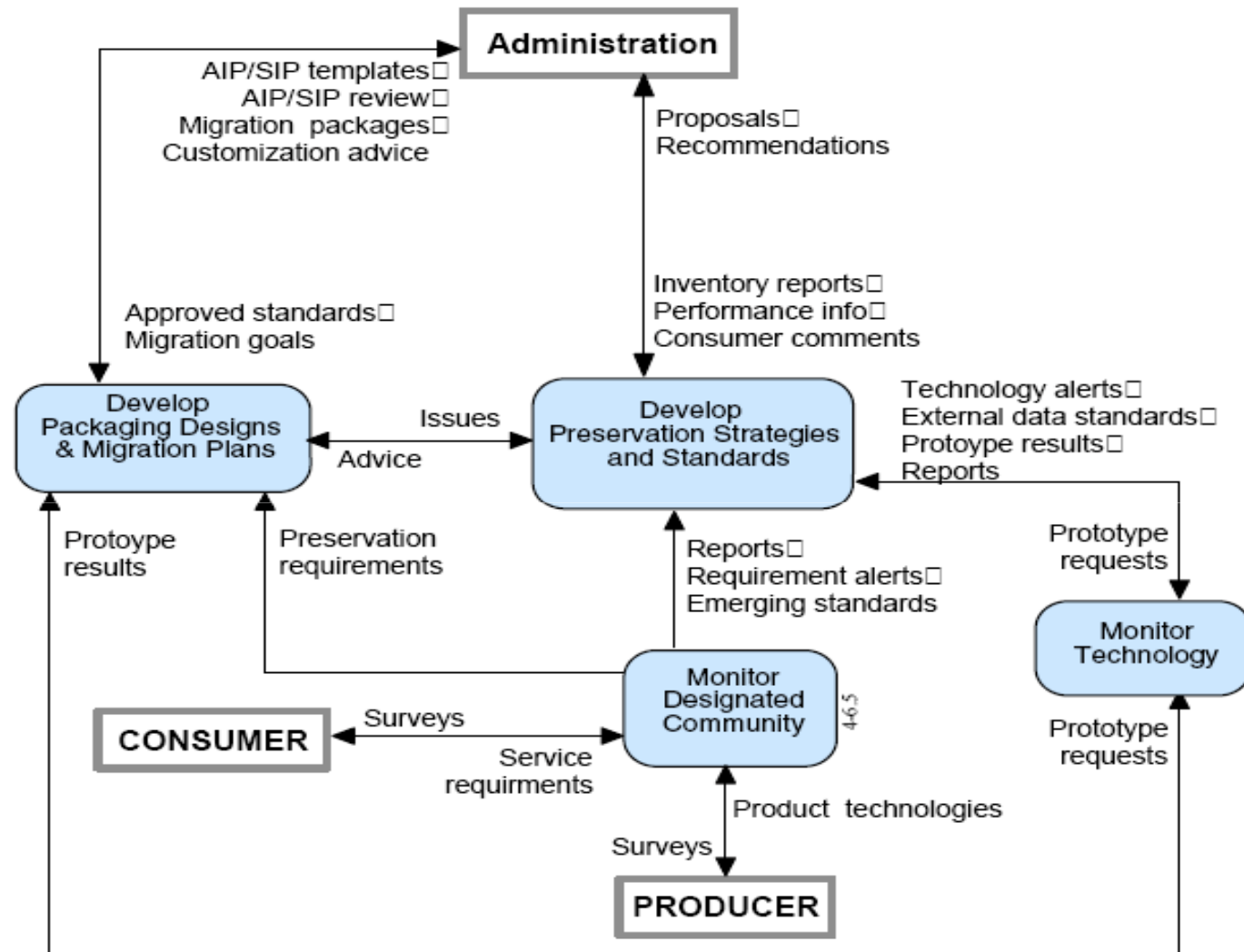
Es responsable de enviar información para establecer estándares y políticas. Desarrolla políticas de gestión de archivo por jerarquías, incluyendo políticas de migración. Es responsable de la recuperación ante desastres.

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 *Preservation Planning*



Fuente [ISO 14721](http://www.iso.org/iso/14721)

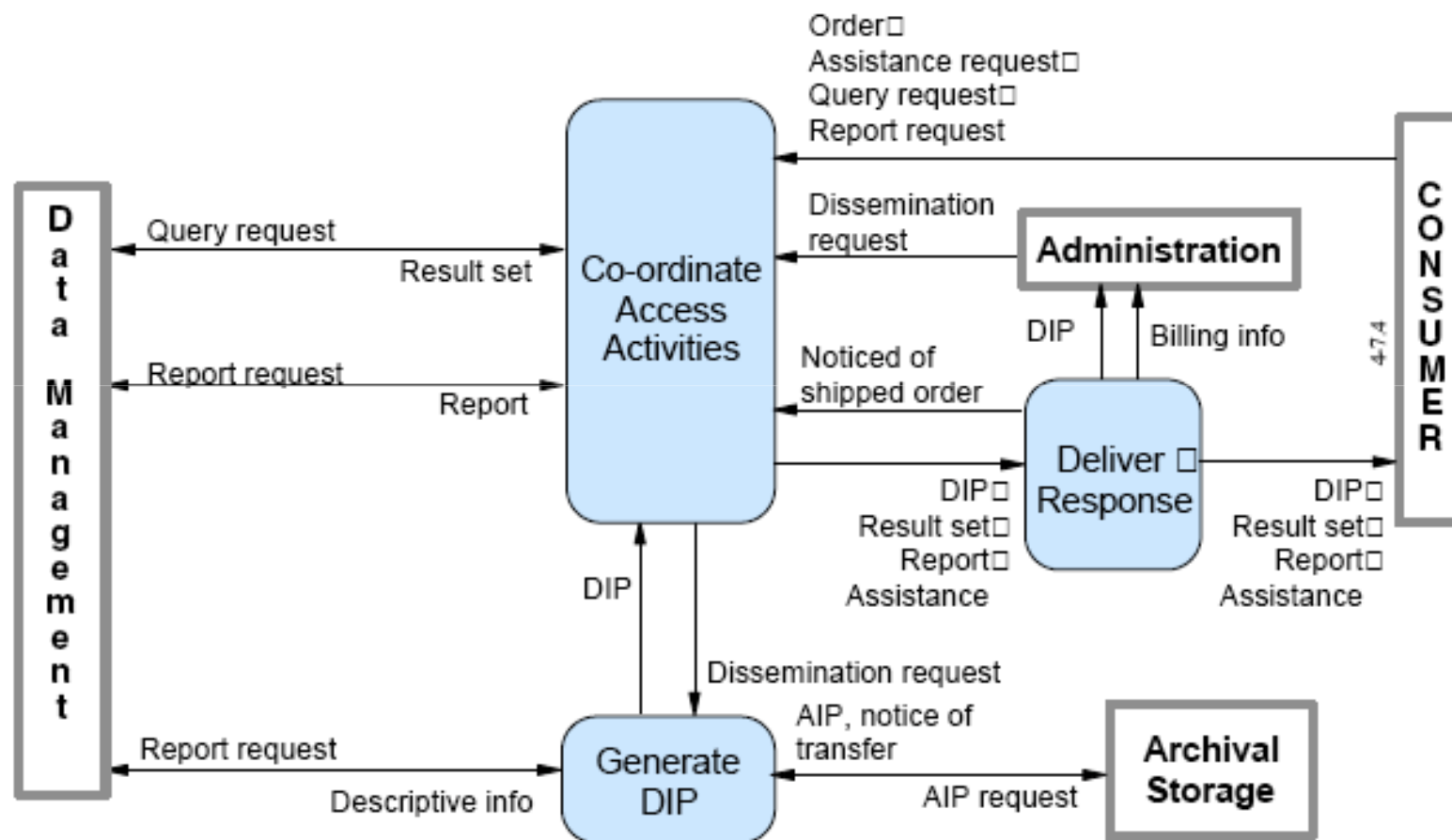
Entidad OAIS *preservation planning*



- **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



Fuente [ISO 14721](#)

Entidad OAIS Access



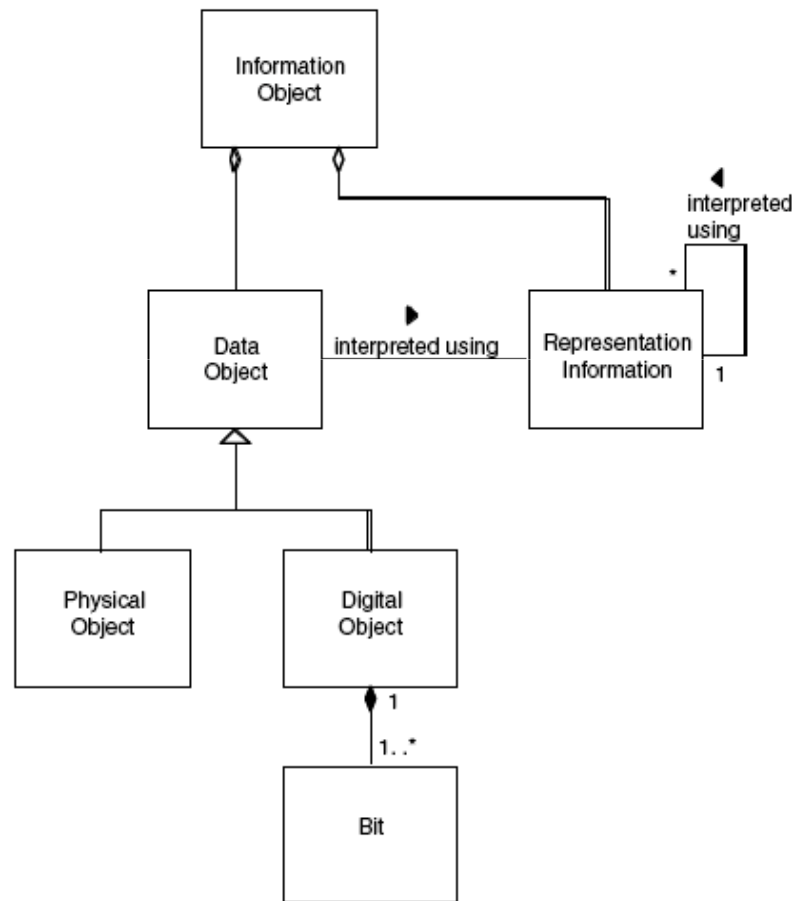
- **Descripción:** Proporciona una interfaz única de usuario para el acceso a la información de los archivos. Tiene 3 categorías, los *query requests*, los *result sets* y los *report requests*.
- Acepta los requerimientos de los paquetes de diseminación recuperados de los AIP de la entidad *archival 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 sólo preliminares

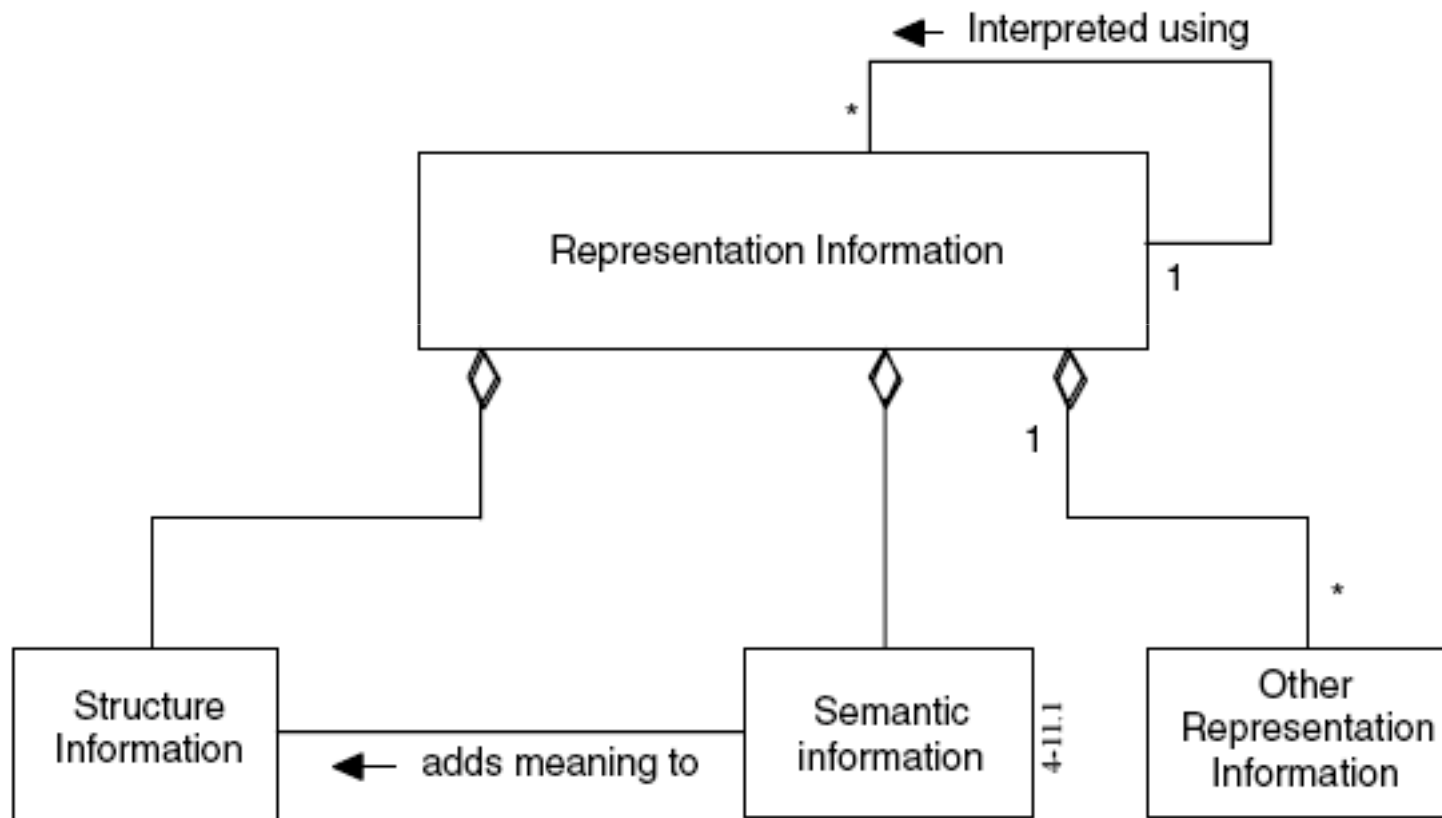
OAIS Objeto de información



- 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.

Fuente [ISO 14721](http://www.iso.org/iso/14721)

Representation Information Object

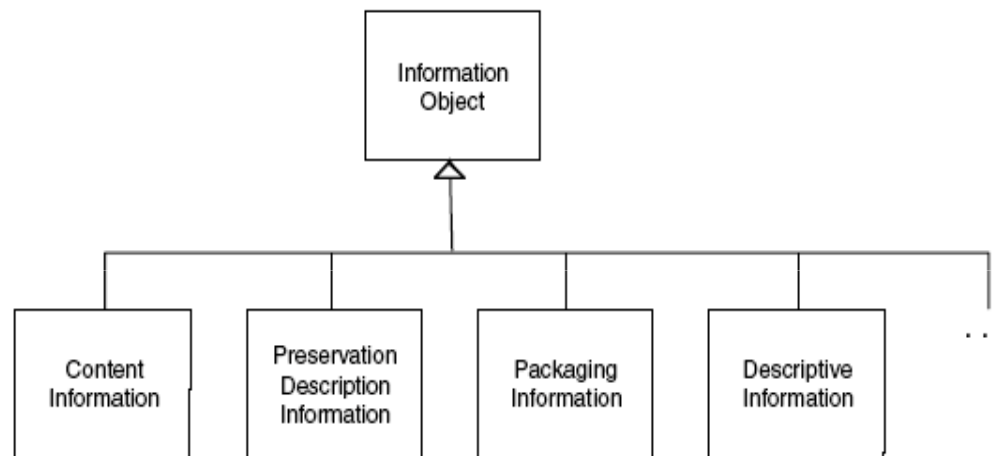


Fuente [ISO 14721](http://www.iso.org/iso/14721.html)

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.

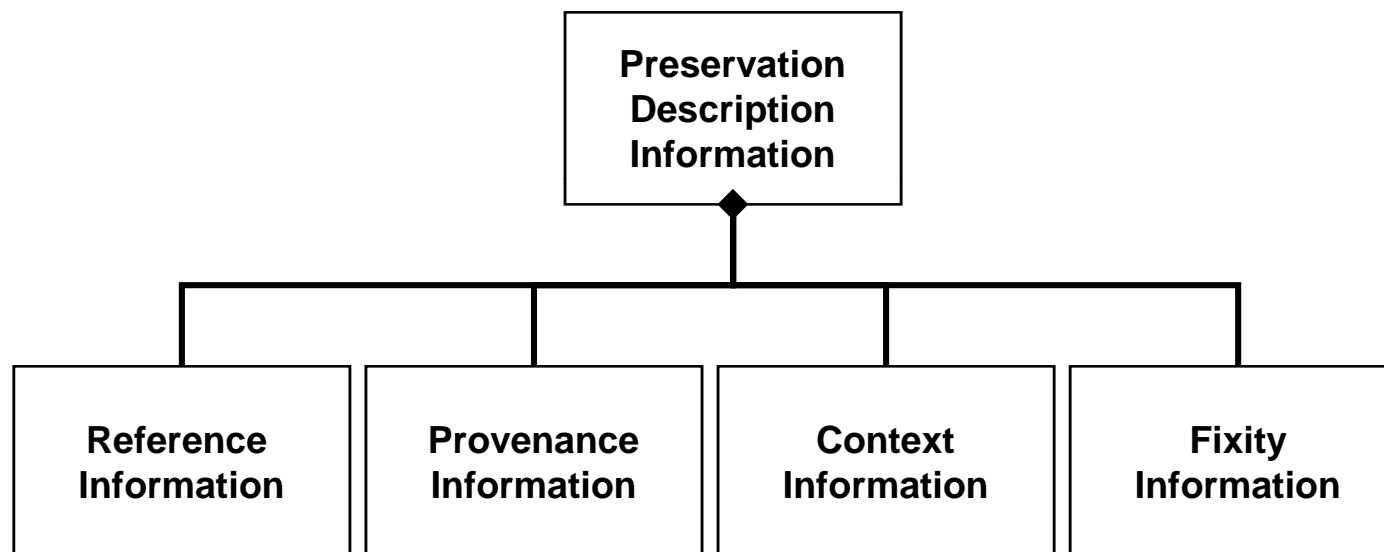


Información de contenido



- 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.

Preservation Description Information



PDI Preservation Description Information (Figure 4-16)

Fuente [ISO 14721](#)

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

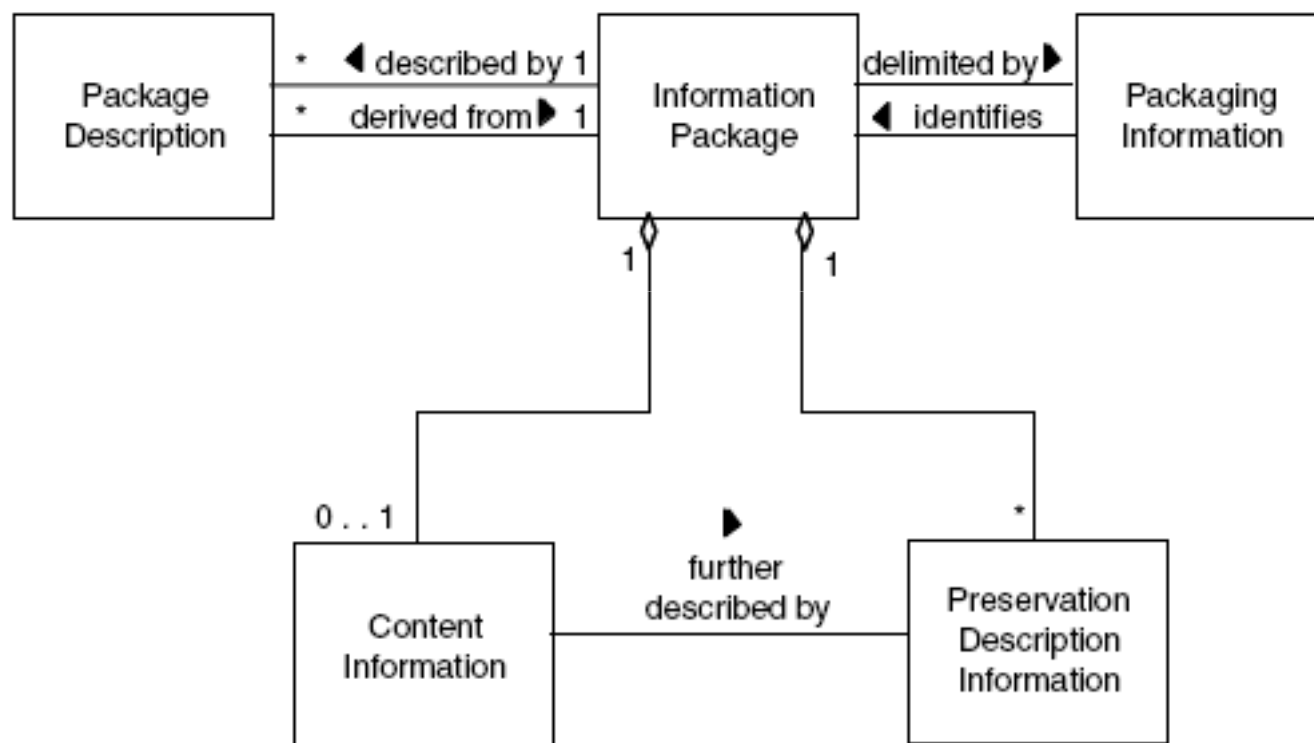


- Información de procedencia: los documentos de la historia de la información de contenido (origen o fuente, los cambios y la custodia) de procedencia puede ser visto como un tipo especial de información de contexto.
- Información de la fijeza: proporciona los controles de integridad de los datos o claves de validación usados para asegurar que la información de contenido no ha sido alterada.

Paquetes de información en OAIS

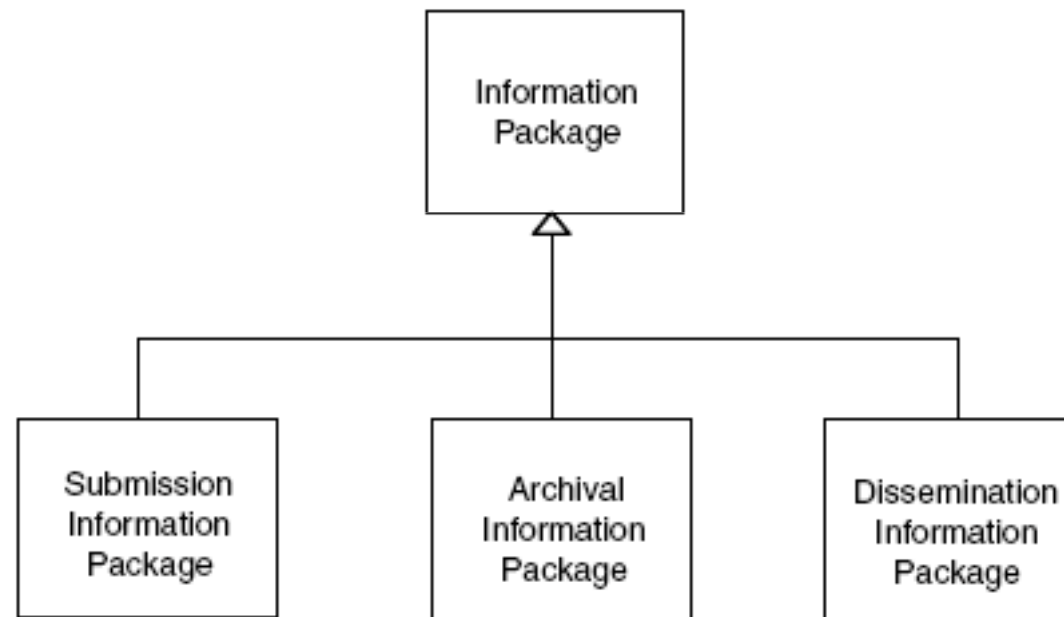
- Las estructuras de información conceptual necesarias para cumplir las funciones OAIS.
- Un paquete de información es un contenedor.
- Hay varios tipos de paquetes de información que se utilizan en el proceso de archivo. Estos paquetes de información pueden ser utilizados para:
 - estructurar y almacenar 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



Fuente [ISO 14721](http://www.iso.org/iso/14721)

Tipos de paquetes de información



Fuente [ISO 14721](http://www.iso.org/iso/14721.html)

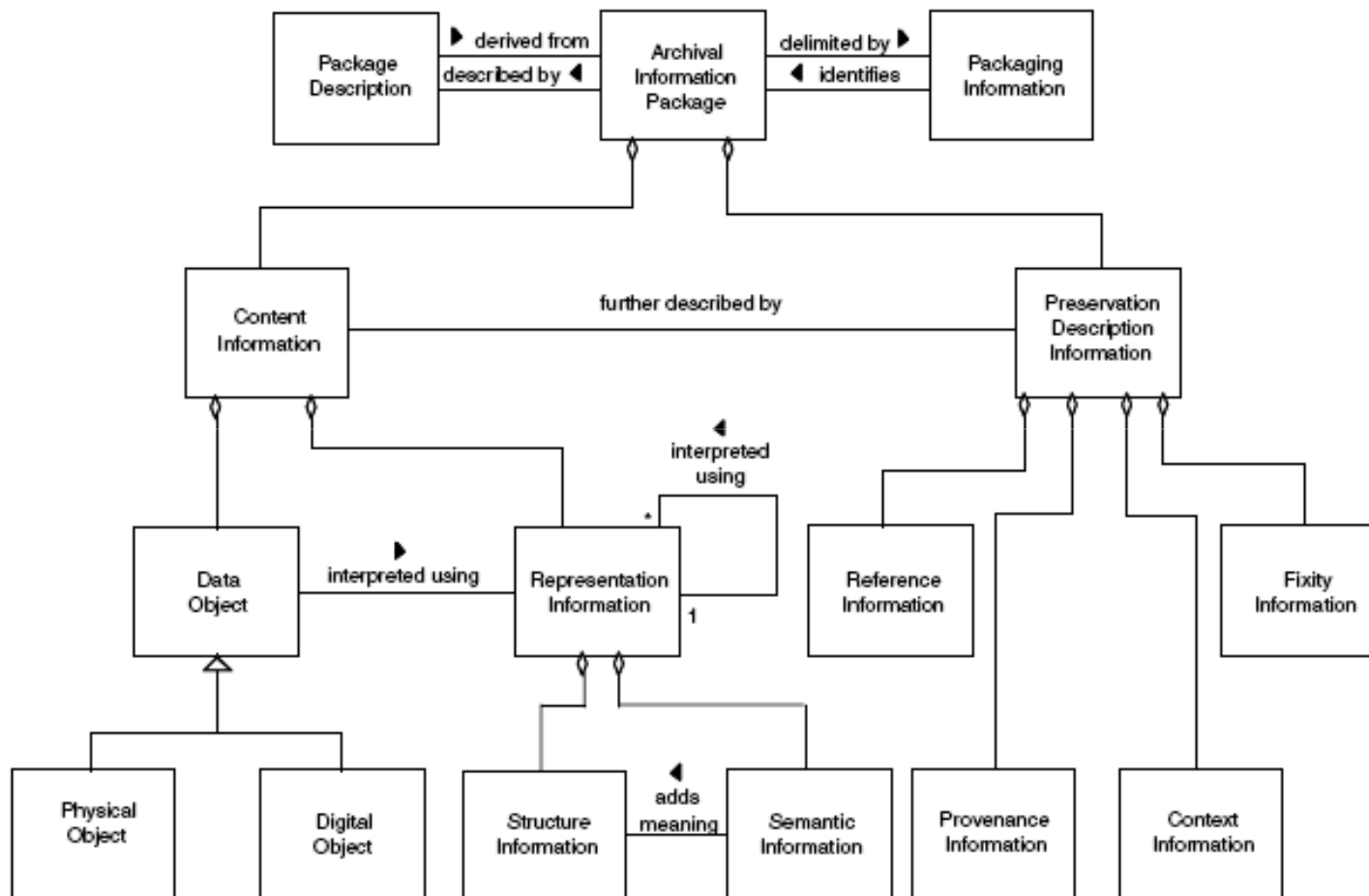
SIP



- 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.



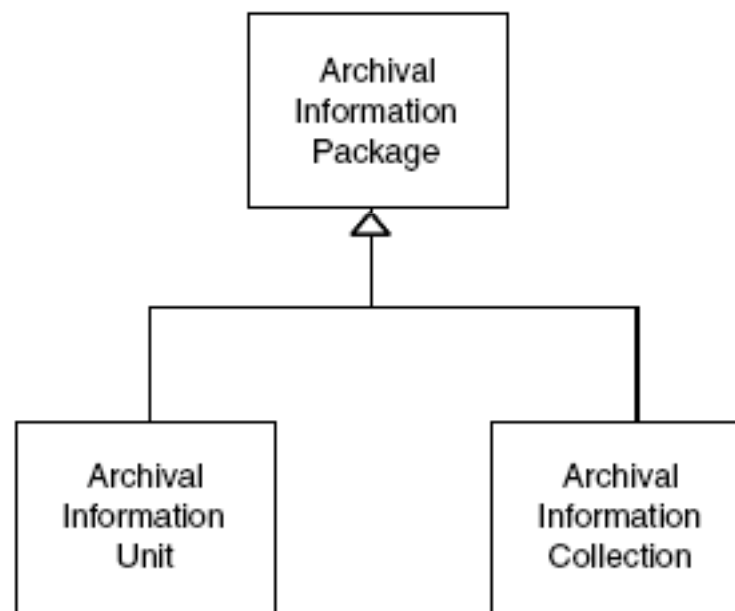
AIP



Fuente [ISO 14721](https://www.iso.org/standard/54467.html)



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





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.

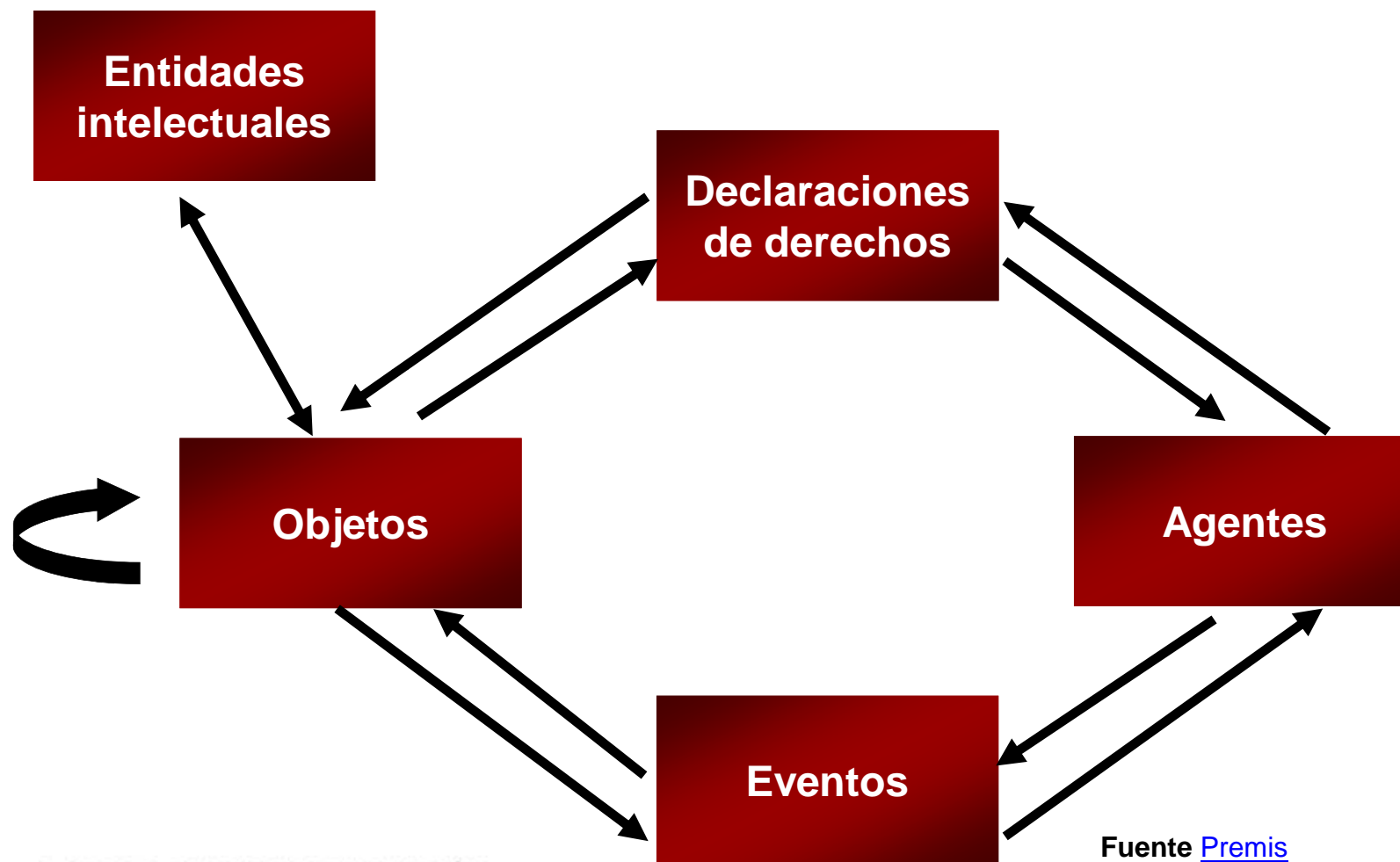
Modelo de Datos PREMIS



Las entidades que este modelo define se denominan:

- Entidades intelectuales
- Objetos
- Derechos
- Agentes
- Eventos

Modelo de datos

Fuente [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^a Méndez Rodríguez

emendez@bib.uc3m.es

Dpto. Biblioteconomía y Documentación



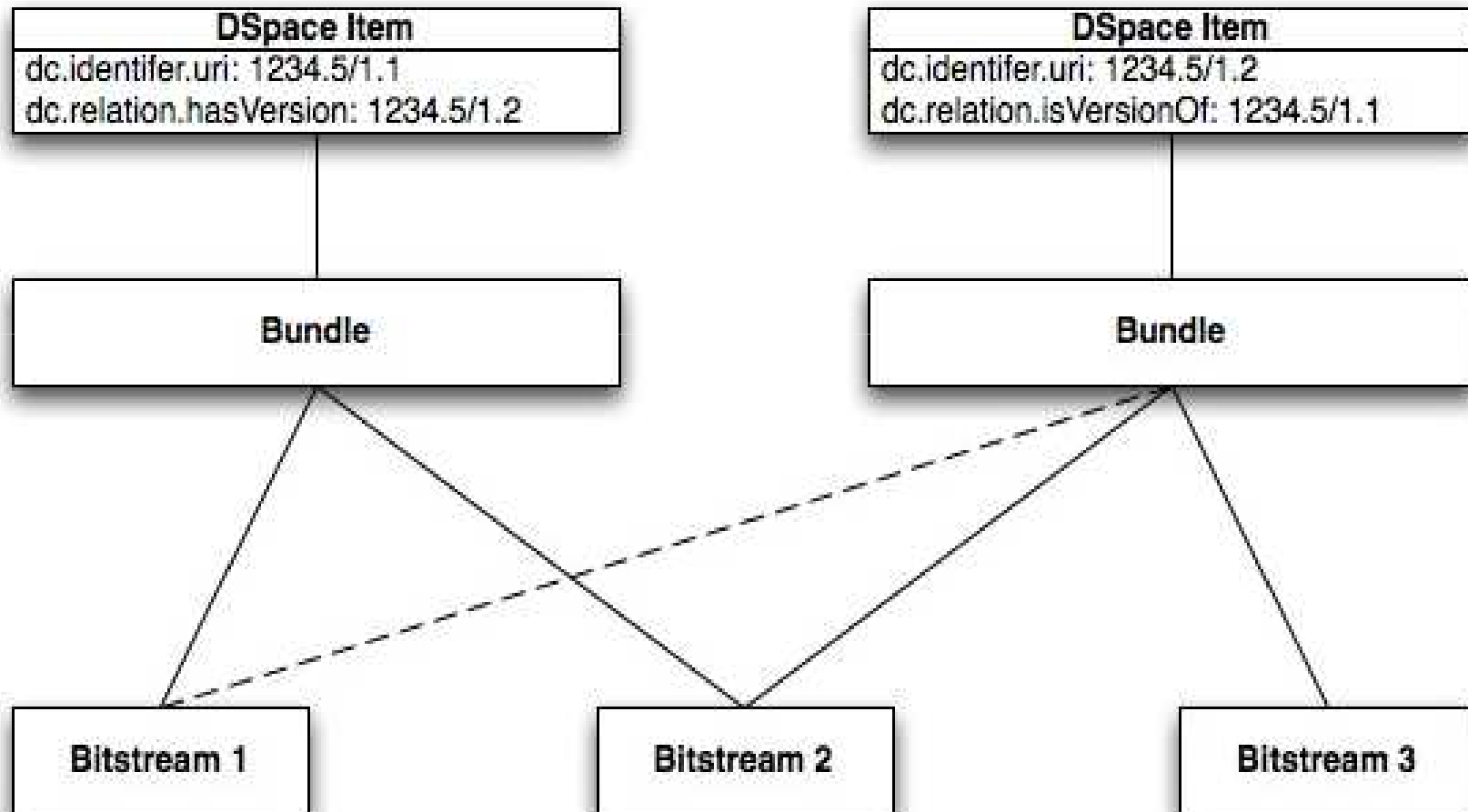
Universidad Carlos III de Madrid

SeDiCI SERVICIO DE DIFUSIÓN DE LA CREACIÓN INTELECTUAL



<http://sedici.unlp.edu.ar>

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 items.

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

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

OAIS en DSPACE



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

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.

Volviendo a la ISO 14721



- Sección 6: Interoperabilidad



SeDiCI SERVICIO DE DIFUSIÓN
DE LA CREACIÓN INTELECTUAL



Bibliotecas y Repositorios Digitales

Tecnología y Aplicaciones

<http://sedici.unlp.edu.ar>

Participantes del dictado



Marisa R. De Giusti

Nestor F. Oviedo

Silvia Pelоче

Matías Cánepa

Bibliotecas y repositorios digitales



Capítulo 7: Interoperabilidad: ventajas y dificultades. La recolección desde otros repositorios y la exposición por diversos protocolos. El protocolo OAI-PMH. Problemas derivados del volumen y heterogeneidad de los datos recolectados. Directrices de interoperabilidad.



Contenido

Introducción

Niveles de interoperabilidad

Formas de interoperar

Formatos de metadatos

OAI-PMH

Recolección de recursos

Directrices de interoperabilidad



Introducción



Introducción

¿Qué es la interoperabilidad?

Capacidad de los sistemas informáticos de interactuar a través del intercambio de información y servicios, para lograr un objetivo.



Introducción

¿Por qué es importante interoperar?

El intercambio de servicios y recursos ayuda a cumplir parte de los objetivos de un repositorio digital:

- Mayor visibilidad e impacto de los recursos propios
- Mayor cantidad de recursos ofrecidos a los usuarios
- Mayor cantidad y diversidad de servicios para ofrecer



Introducción

El contexto del Open Access

Los movimientos de Acceso Abierto y la tendencia mundial hacia estas políticas plantea un marco altamente propicio para la interoperabilidad entre repositorios digitales.

Introducción



Agregadores de recursos

Existen repositorios que se dedican exclusivamente a la recolección y exposición de recursos de terceros. Esto significa que no cuentan con producción propia.

Hispana: más de 3 millones de registros recolectados de entre más de 150 repositorios de España.

<http://hispana.mcu.es>

Europeana: más de 15 millones de registros recolectados de entre más de 1500 repositorios de Europa (específicamente de la Unión Europea). <http://www.europeana.eu>

OAister: más de 23 millones de recursos recolectados de entre más de 1100 repositorios de acceso abierto de todo el mundo. <http://www.oclc.org/oaister>



Introducción

Directrices de interoperabilidad

Son un conjunto de reglas y recomendaciones que buscan establecer un marco de trabajo a fin de que dos sistemas puedan interactuar de forma exitosa y confiable.



Niveles de interoperabilidad



Niveles de interoperabilidad

Dado que *interoperabilidad* es un término muy amplio (aplicable en muchas disciplinas), existen múltiples clasificaciones del mismo.

En lo que respecta a los repositorios digitales, interesa analizar una perspectiva mas bien tecnológica y acotada:

- Interoperabilidad Sintáctica
- Interoperabilidad Semántica

Niveles de interoperabilidad

Sintáctica



Hace referencia a todo lo necesario para que dos sistemas sean capaces de establecer una comunicación e intercambiar información.

Esto incluye:

- protocolos de comunicación y transferencia
- codificación de caracteres
- formatos de datos

Niveles de interoperabilidad

Sintáctica



Elementos que corresponden a la interoperabilidad sintáctica pueden ser, por ejemplo:

- protocolo TCP/IP
- protocolo HTTP
- protocolo OAI-PMH
- formato XML y esquemas XML (XSD)
- Directrices de interoperabilidad

Niveles de interoperabilidad

Semántica



Hace referencia a todo lo necesario para que el sistema receptor haga una correcta interpretación de la información recibida, de forma automática.

Se busca que el sistema receptor "**entienda**" los datos tal como los "**entiende**" el emisor.

Para contar con interoperabilidad semántica, primero debe asegurarse la interoperabilidad sintáctica

Niveles de interoperabilidad

Semántica



Entran en juego:

- Formatos de metadatos
- Vocabularios controlados:
 - Tesoros
 - Sistemas de clasificación
- Ontologías
- Directrices de interoperabilidad

Niveles de interoperabilidad

Estándares internacionales



La adopción de estándares internacionales aumenta las capacidades de interoperabilidad del repositorio.

Protocolos de transferencia: REST, Z39.50, etc

Formatos de archivos: XML, etc

Formatos de metadatos: DC, MODS, MARCXML, etc

Directrices: DRIVER, Lucis MODS, OpenAIRE, etc



Formas de interoperar



Formas de interoperar

En general, en el contexto de los repositorios digitales se habla de:

- Búsqueda remota
- Recolección de recursos
- Depósito remoto

Formas de interoperar

Búsqueda remota: Z39.50



- Definido en los estándares internacionales ANSI/NISO z39.50 e ISO 23950
- Protocolo cliente-servidor de búsqueda y recuperación desde bases de datos remotas.
- Ampliamente utilizado en sistemas integrados de bibliotecas (ILS - *Integrated Library Systems*) para la búsqueda remota y la gestión de préstamos interbibliotecarios (*Interlibrary Loan*).
- Sintaxis de consulta específica: PQF (*Prefix Query Format*)

Formas de interoperar

Búsqueda remota: Z39.50



```
Z> find @attr 1=1003 software
```

```
Sent searchRequest.
```

```
Received SearchResponse.
```

```
Search was a success.
```

```
Number of hits: 66, setno 1
```

```
records returned: 0
```

```
Elapsed: 0.267659
```

```
Z> show 1
```

```
Sent presentRequest (1+1).
```

```
Records: 1
```

```
[INNOPAC]Record type: USmarc
```

```
00770nam 2200193I 4500
```

```
001 547843
```

```
008 730130s1970 enkm a100 0 eng u
```

```
040 $c MIA $d m.c. $d IQU
```

```
049 $a IQUU
```

```
099 $a QA $a 76.6 $a S64 $a 1970
```

```
111 2 $a Software 70 Conference $d (1970 : $c University...)
```

```
245 10 $a Software 70: $b proceedings of a conference ...
```

```
260 $a Princeton, N. J., $b Auerbach, $c 1970.
```

```
300 $a 197 p. $b illus. $c 29 cm.
```

```
500 $a Includes bibliographical references.
```

```
650 0 $a Computer programming $v Congresses.
```

```
650 0 $a Programming languages (Electronic computers) $v Congresses.
```

```
700 1 $a Evans, David J.
```

```
710 2 $a Software World (Firm)
```

```
nextResultSetPosition = 2
```

```
Elapsed: 0.296679
```

```
Z>
```

Formas de interoperar

Búsqueda remota: Z39.50



Ventajas y desventajas

- Las consultas son abstractas respecto de la estructura de la base de datos que se está consultando
- Los mapeos de campos de búsqueda dependen de la implementación de cada servidor
- No aprovecha las ventajas de la web actual (protocolo REST)

Formas de interoperar

Búsqueda remota: SRU/SRW



SRU (*Search / Retrieve via URL*) y SRW (*Search / Retrieve via Web*) nacen como los sucesores del protocolo Z39.50, y se apoyan sobre tecnologías actuales y muy difundidas (HTTP, XML).

Al igual que Z39.50, la agencia responsable del mantenimiento de estos dos estándares es la Library of Congress

Ambos son considerados muy simples de entender e implementar

Formas de interoperar

Búsqueda remota: SRU



Se caracteriza por enviar la expresión de búsqueda (y cualquier otra indicación) dentro de una URL.

Esto es, todos los comandos necesarios para que el servidor entienda una petición y lleve a cabo las acciones pertinentes, se envían dentro de la URL misma de la petición.

<http://fedora.dlib.indiana.edu:8080/SRW/search/GSearch?query=dc.title=road>

Formas de interoperar

Búsqueda remota: SRW



Al igual que su *mellizo* SRU, trabaja sobre tecnologías actuales y muy difundidas: XML y HTTP, pero presenta una importante diferencia: el envío de la petición se realiza mediante un POST al servidor, en el que se envía un documento XML que contiene todas las instrucciones y datos correspondientes.

Esto es, la consulta al servidor se "empaqueta" en XML y se envía, recibiendo XML como respuesta (al igual que en el caso de SRU)

Formas de interoperar

Búsqueda remota: SRW



Las reglas y restricciones utilizadas para armar e interpretar el paquete XML están dadas por el protocolo **SOAP**.

SOAP fue creado y es mantenido por la W3C, en el área de los Web Services.

SOAP es un protocolo estándar y muy difundido.

Casi cualquier lenguaje de programación moderno tiene librerías para trabajar con SOAP.

Formas de interoperar

Búsqueda remota: SRW



Petición SRW

```
<SOAP:Envelope xmlns:SOAP="http://schemas.xmlsoap.org/soap/envelope/">
  <SOAP:Body>
    <SRW:searchRetrieveRequest xmlns:SRW="http://www.loc.gov/zing/srw/">
      <SRW:version>1.1</SRW:version>
      <SRW:query>(dc.author exact "jones" and dc.title >= "smith")</SRW:query>
      <SRW:startRecord>1</SRW:startRecord>
      <SRW:maximumRecords>10</SRW:maximumRecords>
      <SRW:recordSchema>info:srw/schema/1/mods-v3.0</SRW:recordSchema>
    </SRW:searchRetrieveRequest>
  </SOAP:Body>
</SOAP:Envelope>
```


Formas de interoperar

Búsqueda remota: SRW



Respuesta

```
<SOAP:Envelope xmlns:SOAP="http://schemas.xmlsoap.org/soap/envelope/">
  <SOAP:Body>
    <SRW:searchRetrieveResponse xmlns:SRW="http://www.loc.gov/zing/srw/"
      <SRW:version>1.1</SRW:version>
      <SRW:numberOfRecords>2</SRW:numberOfRecords>
      <SRW:resultSetId>8c527d60-c3b4-4cec-alde-1ff80a5932df</SRW:resultSetId>
      <SRW:resultSetIdleTime>600</SRW:resultSetIdleTime>
      <SRW:records>
        <SRW:record>
          <SRW:recordSchema>info:srw/schema/1/mods-v3.0</SRW:recordSchema>
          <SRW:recordPacking>string</SRW:recordPacking>
          <SRW:recordData> DATOS </SRW:recordData>
          <SRW:recordPosition>1</SRW:recordPosition>
        </SRW:record>
      </SRW:records>
    </SRW:searchRetrieveResponse>
  </SOAP:Body>
</SOAP:Envelope>
```

Formas de interoperar

Búsqueda remota: OpenSearch



Es un protocolo que extiende otros formatos para agregar la búsqueda remota.

Las peticiones se realizan vía GET

Proporciona **Autodiscovery**: permite que los navegadores detecten que el sitio soporta OpenSearch y así el sitio podrá seleccionarse como motor de búsquedas del navegador

Las respuestas se envían en RSS o ATOM, extendidos con elementos OpenSearch que agregan información sobre la búsqueda

Ejemplos: Youtube, SeDiCI, Facultad de Informática

Formas de interoperar

Recolección de recursos: OAI-PMH



Open Archives Initiative - Protocol for Metadata Harvesting

Establece un conjunto de reglas a partir de las cuales puede realizarse el intercambio de recursos de forma exitosa.

Se centra en la **transferencia** de metadatos de un extremo a otro, sin establecer restricciones en cuanto a los datos que se transfieren.

Formas de interoperar

Recolección de recursos: OAI-PMH



Define dos perfiles de trabajo

Data Provider: es aquél repositorio que ofrece sus recursos bajo el protocolo OAI-PMH, para que otros los recolecten mediante cosechas.

Service Provider: es aquél que recolecta recursos desde distintos Data Providers y brinda un servicio a una comunidad de usuarios en base a los recursos recolectados y el valor agregado aportado sobre los mismos (deduplicación, normalización, ordenamiento, búsquedas, etc).

Formas de interoperar

Depósito remoto: SWORD



Simple Web service Offering Repository Deposit

Protocolo basado en APP (Atom Publishing Protocol, a.k.a ATOMPUB)

Permite realizar el depósito de documentos de forma remota: desde otros sistemas.

Es un protocolo cliente-servidor

Formas de interoperar

Depósito remoto: SWORD



Múltiples usos potenciales

- Depósito simultáneo en múltiples repositorios
- Depósito automático por parte de equipamiento científico
- Depósito desde aplicaciones externas al repositorio (escritorio, OJS, etc)

Es un estándar que se limita a la transferencia de un objeto desde el cliente al servidor, sin imponer restricciones en cuanto a los objetos que se transportan.

Esto lo hace suficientemente flexible como para ser usado en cualquier tipo de repositorio.



Formatos de metadatos



Formatos de metadatos

Existen muchos estándares de formatos de metadatos

Cada repositorio decide que formato de metadatos usar
(incluso puede usar un formato propio)

Los repositorios que deciden interoperar deben estar de acuerdo en cuanto a un formato de metadatos que ambos puedan manejar



Formatos de metadatos

En todas las formas de interoperar presentadas existe un rol de proveedor de recursos y un rol de receptor de recursos.

¿Qué sucede cuando el proveedor de recursos utiliza un formato de metadatos que no es manejado por el receptor?

¿Como se gestiona este problema?



Formatos de metadatos

Algunas de las alternativas aplicables en cualquiera de los dos roles mencionados pueden ser:

- Se decide no interactuar con ese repositorio en particular
- Extender el software para así agregar soporte para un formato de metadatos en particular
- Realizar mapeos entre formatos de metadatos
 - También dependen de la flexibilidad del software

Formatos de metadatos

Mapeos entre formatos de metadatos



En algunos casos, las entidades responsables de un formato de metadatos recomiendan cómo deben realizarse los mapeos a otros formatos. Ejemplo de esto es MODS:

Conversión de DC (sin calificar) a MODS:

<http://www.loc.gov/standards/mods/dcsimple-mods.html>

Conversión de MODS a DC (sin calificar):

<http://www.loc.gov/standards/mods/mods-dcsimple.html>

Formatos de metadatos

Mapeos entre formatos de metadatos



Manual: es un trabajo muy costoso, ya que puede tratarse de miles de registros

Automático: la transformación desde un formato complejo/jerárquico a uno simple/plano implica pérdida de información. La transformación inversa puede generar recursos deficientes en cuanto a la descripción (campos incompletos, imposibilidad de uso de la especificidad de un formato complejo). No hay un humano tomando decisiones.



OAI-PMH

Open Archives Initiative
Protocol for Metadata Harvesting

OAI-PMH

Introducción



Protocolo para la recolección de metadatos

- Ampliamente adoptado por repositorios digitales en todo el mundo
- Es muy simple de entender y utilizar
- Funciona sobre XML y HTTP
- Se centra en establecer un marco de reglas para la transferencia eficiente de recursos
- No impone (*casi*) ninguna restricción en cuanto al contenido a transmitir

OAI-PMH

Introducción



Las peticiones al servidor se hacen por medio de un *verbo* y un conjunto de parámetros, codificados en una URL

http://host/oai?verb=ListRecords&metadataPrefix=oai_dc&from=2011-05-01&until=2011-10-01

<http://host/oai?verb=ListRecords&resumptionToken=1320093034051>

Un verbo es una *orden* que indica al servidor lo que se requiere, refinando algunos aspectos de ese requerimiento a través del uso de parámetros.

OAI-PMH

Introducción



La respuesta a una petición OAI-PMH es un documento XML.

Se compone de dos secciones:

- *Información de la petición:* fecha, hora, verbo y parámetros (común para cualquier verbo)
- *Cuerpo con la respuesta:* datos con una estructura acorde a la información solicitada (específico para cada verbo)

OAI-PMH

Funcionamiento



Los verbos disponibles son:

- Identify
- ListRecords
- ListMetadataFormats
- ListSets
- ListIdentifiers
- GetRecord

OAI-PMH

Funcionamiento



Verbo Identify

Retorna información del repositorio e información acerca de la implementación del OAI Data Provider.

No recibe parámetros.

<http://sedici.unlp.edu.ar/oai/request?verb=Identify>

<http://bdigital.uncu.edu.ar/OAI/index.php?verb=Identify>

OAI-PMH

Funcionamiento



Elementos importantes que se desprenden del *Identify*

- Fecha/hora de creación del recurso mas viejo
- Granularidad de las peticiones
- Gestión de registros eliminados
- Compresión de los datos a transferir
- OAI Friends
- Descripción del repositorio

OAI-PMH

Funcionamiento



Verbo *ListRecords*

- Retorna un listado de recursos que cumplen con los parámetros especificados en la petición:
 - metadataPrefix (*obligatorio*)
 - resumptionToken (*opcional*)
 - set (*opcional*)
 - from (*opcional*)
 - until (*opcional*)

http://sedici.unlp.edu.ar/oai/request?verb=ListRecords&metadataPrefix=oai_dc&from=2011-01-01

OAI-PMH

Funcionamiento

Cosechas incrementales

por fecha (from y until)

Información clasificada

por conjuntos (set)

Paginación de resultados

resumptionToken



OAI-PMH

Funcionamiento



Registro de respuesta

```
<header>
  <identifier>ARG-UNLP-TPG-0000000006</identifier>
  <timestamp>2010-07-14</timestamp>
</header>
<metadata>
  <oai_dc:dc xmlns:...>
    <dc:title>Simulación numérica de difusión ...</dc:title>
    <dc:creator>Zyserman, Fabio Iván</dc:creator>
    <dc:subject>Física</dc:subject>
    <dc:contributor>Plastino, Angel L.</dc:contributor>
    <dc:date>2000</dc:date>
    <dc:type>Tesis de Posgrado</dc:type>
  </oai_dc:dc>
</metadata>
<about>
  <rights/>
  <provenance/>
</about>
```

OAI-PMH

Funcionamiento



Verbo *ListMetadataFormats*

Lista todos los formatos de metadatos soportados por el repositorio.

OAI-PMH obliga a exportar, por lo menos, Dublin Core sin calificar.

Se indica el *prefix* que identifica el *namespace* del formato de metadatos.

Parámetro opcional *identifier*

<http://sedici.unlp.edu.ar/oai/request?verb=ListMetadataFormats>

OAI-PMH

Funcionamiento



Verbo *ListSets*

- Lista los distintos Sets soportados por el repositorio
- Son una forma de organizar la información dentro del repositorio
- Poseen un nombre y una clave que los identifica
- Parámetro opcional *resumptionToken*

sedici.unlp.edu.ar/oai/request?verb=ListSets

bdigital.uncu.edu.ar/OAI/index.php?verb=ListSets

OAI-PMH

Funcionamiento



Verbo *ListIdentifiers*

- Lista los encabezados de todos los registros que se corresponden con los parámetros especificados.
- Recibe los mismos parámetros que ListRecords
- Se suele usar para determinar la cantidad y estado de los registros (borrado o no) que coinciden con ciertos parámetros, sin necesidad de descargar sus metadatos

http://sedici.unlp.edu.ar/oai/request?verb=ListIdentifiers&metadataPrefix=oai_dc&from=2011-11-01

OAI-PMH

Funcionamiento



Verbo *GetRecord*

Retorna el registro completo (encabezado y metadatos) de un recurso específico.

Recibe los parámetros
identifier

metadataPrefix



Recolección de recursos

Utilizando OAI-PMH



Recolección de recursos

Cuando se recolectan recursos desde múltiples repositorios, se presentan varios problemas.

- Políticas de catalogación independientes
- Diferencia de formatos de metadatos (y por lo tanto de especificidad de la información)
- Múltiples términos para el mismo concepto (ej.: idiomas)
- Uso de múltiples vocabularios controlados (tesauros, sistemas de clasificación, etc)
- **La gran mayoría expone sus recursos sólo en Dublin Core sin calificar**

Recolección de recursos

Problemas a solucionar



Formatos de metadatos

Mapeos a un formato común

- ¿cuál?

Diferencias en la codificación de caracteres

Presencia de caracteres inválidos:

- ¿se descarta el caracter inválido?
- ¿se descarta el documento completo?
- ¿se utiliza un caracter de reemplazo?

Recolección de recursos



Problemas a solucionar

Autores

- Distinción entre apellido y nombres (considerar el uso de iniciales)
- Muchas veces se incluye a la institución como autor
- Unificación de autores

Instituciones

- Identificación de instituciones (generalmente aparecen junto con personas)
- Unificación de instituciones

Recolección de recursos

Problemas a solucionar



Idiomas

Identificación del idioma: eng, en, en_US

Muchas veces no se indica el idioma (se necesita aplicar una detección automática)

Unificación de idiomas

Tipología documental

Múltiples formas de referenciar el mismo tipo de recurso

Artículo, ART, Article

Unificación de tipologías documentales

Recolección de recursos

Problemas a solucionar



Acceso al PDF o a los metadatos

Muchos casos en los que la URL apunta a una *jump-page* desde donde se accede al PDF

Otros casos, la *jump-page* no presenta ningún link al PDF

Validación de la URL de acceso al recurso

Muchas veces el enlace de acceso al recurso no funciona (o deja de funcionar un tiempo después)

¿Cómo detectar esos casos y cómo actuar? ¿se descarta el recurso?



Directrices de interoperabilidad

Directrices de interoperabilidad



Son un conjunto de recomendaciones que buscan maximizar la interoperabilidad entre los repositorios.

DRIVER 2.0 es la mas difundida en Europa y la base de muchas otras directrices en el mundo (ej.: LUCIS-MODS, OpenAIRE)

DRIVER 2.0 establece recomendaciones tanto a nivel **sintáctico** y como a nivel **semántico**.

Directrices de interoperabilidad

DRIVER 2.0



Extracto del documento de DRIVER 2.0

Para la comunicación en general es importante que la persona B sea capaz de comprender lo que la persona A está diciendo. Para este entendimiento mutuo, se necesita una base común, un léxico básico con una comprensión del significado de las cosas. A partir de este punto, ya se puede comenzar el razonamiento. Para respaldar la comunicación científica con el uso de repositorios, éstos deberían hablar el mismo idioma y por tanto es fundamental crear una base común.

Directrices de interoperabilidad

DRIVER 2.0: características generales



Diseñado sólo para:

- Protocolo OAI-PMH
- Recursos textuales
- Documentos a texto completo
- Documentos en Acceso Abierto
- *Dublin Core sin calificar* como formato de metadatos

Directrices de interoperabilidad

DRIVER 2.0: características generales



Sobre el uso de OAI-PMH

- Se reserva el prefijo *oai_dc* para identificar el formato de metadatos *DC Sin Calificar*
- Los datestamp (tanto en las solicitudes como en las respuestas) debe respetar el formato ISO8601, expresadas en UTC: AAAA-MM-DDThh:mm:ssZ
- La política de registros eliminados debe ser por lo menos *transient* (aunque se recomienda *persistent*).

Directrices de interoperabilidad

DRIVER 2.0: características generales



Sobre el uso de OAI-PMH

Se recomienda que el `resumptionToken` se mantenga activo por lo menos por 24 horas.

El tamaño del lote debe ubicarse entre 100 y 500 registros.

Si se utiliza un set específico para DRIVER, se recomienda usar *driver* como `setSpec`.

Es obligatorio indicar un mail de contacto (campo *adminEmail* de la respuesta del verbo *Identify*)

Directrices de interoperabilidad

DRIVER 2.0: características generales



Sobre el uso de Dublin Core

Es obligatorio usar codificación Unicode.

El contenido de los metadatos no puede incluir lenguaje de marcado (HTML ni XML).

Se recomienda que el contenido de los metadatos se encuentre en inglés.

El metadato *dc:creator* debe respetar el estilo bibliográfico APA: *apellido, iniciales (nombre)*

Directrices de interoperabilidad

DRIVER 2.0: características generales



Sobre el uso de Dublin Core

Se recomienda que el metadato *dc:description* contenga un resumen del documento (el abstract).

El metadato *dc:date* debe repetir el formato de fecha ISO8601. Se recomienda que contenga la fecha de publicación del documento.

Directrices de interoperabilidad

DRIVER 2.0: características generales



Sobre el uso de Dublin Core

El metadato *dc:type* debe pertenecer a un vocabulario definido en un esquema URI (info:eu-repo/semantic)

info:eu-repo/semantics/article

info:eu-repo/semantics/book

info:eu-repo/semantics/bachelorThesis

info:eu-repo/semantics/masterThesis

info:eu-repo/semantics/doctoralThesis

info:eu-repo/semantics/preprint

Directrices de interoperabilidad

DRIVER 2.0: características generales



Sobre el uso de Dublin Core

Se recomienda que el metadato *dc:format* sea un MIME-Type incluido en IANA. Ej.: application/pdf

El metadato *dc:identifier* debe respetar un esquema URI, y vincular a:

Identificador persistente (DOI, Handle, etc)

Documento a texto completo (ej.: PDF)

Página de transición (jump-page)