



UNIVERSIDAD  
NACIONAL  
DE LA PLATA

**SEDICI**

REPOSITORIO INSTITUCIONAL DE LA UNLP



# Curso de Repositorios Digitales 2017

Prof. a cargo: Dra. Marisa R. De Giusti

**PREBI**  
prebi.unlp.edu.ar



Esta obra está bajo una [Licencia Creative Commons Atribución-CompartirIgual 4.0 Internacional](#).

Este material ha sido elaborado en conjunto con el personal del repositorio SEDICI

# Objetivos 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.



# Orden del Curso de repositorios 2017

**Clase 1:** Historia, acceso abierto, repositorios, la ley, implementación.

**Clase 2:** Aspectos tecnológicos.

**Clase 3:** Derechos, licencias, vías de publicación y visibilidad.

**Clase 4:** Administración.

**Clase 5:** Preservación y digitalización. Visita al **SEDICI**.



# ¿Cómo se aprueba el curso?

Entregando un trabajo monográfico **ANTES** de culminar 2017.

La carátula del trabajo debe llevar los datos del curso, la edición, la facultad, la universidad y los datos del alumno y del docente.

La temática del trabajo debe estar dentro de los temas tratados en el curso.

Debe ser un trabajo para el curso, no una adaptación ni copia de otros trabajos localizados en la web porque se pretende una demostración de lo aprendido en el curso.

En lo posible diagramar qué se va a realizar previo a la finalización del curso.



# Clase 1

## Acceso abierto y repositorios





# Acceso Abierto



# Acceso Abierto

El Movimiento de Acceso Abierto tiene como fin asegurar el **acceso libre y abierto** más amplio posible a la producción científica.

Una de las formas de lograr ese objetivo es por medio de la creación de **repositorios institucionales** en los que 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.

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**
- ❑ Los **repositorios digitales**:
  - ❑ *temáticos e institucionales.*

# El sistema de comunicación científica y la crisis de su entorno

A partir de la década del 80, desde todos los ámbitos (consultores, científicos, bibliotecarios y editores) se mantuvo el señalamiento de que el sistema tradicional de comunicación científica estaba en crisis porque no se cumplía con sus objetivos primarios: favorecer la diseminación y el intercambio de los resultados científicos para lograr avanzar en ciencia y obtener mayor progreso para la sociedad.



# Factores de la crisis

- ❑ Fusiones de grandes empresas editoriales y desaparición de las más pequeñas: un mercado sin competencia.
- ❑ Restricciones a los derechos de autor para dar acceso y diseminación de la información científica, desvirtuando los objetivos primarios de la comunicación científica y del propio derecho de autor, y focalizando la publicación en revistas “de impacto” en desmedro de la amplia diseminación de los resultados científicos.

# Factores de la crisis

- ❑ El incremento de los precios de las revistas científicas, sobre todo en las áreas de ciencia, tecnología y medicina.
- ❑ El impacto en áreas científico-técnicas de constante cambio, como es el caso de la informática, con una necesidad de mayor agilidad entre la extensión del período entre envío de un artículo y su publicación.



**Hitos** ArXiv fue desarrollado originalmente por [Paul Ginsparg](#) y comenzó en 1991 como un archivo para borradores de física.

Cornell University Library

We gratefully acknowledge support from the Simons Foundation and UNC- Chapel Hill Library

arXiv.org

Search or Article ID All papers

([Help](#) | [Advanced search](#))

Open access to 1,303,895 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics

Subject search and browse:

11 Sep 2017: Steinn Sigurdsson Appointed as arXiv Scientific Director  
 11 Sep 2017: arXiv awarded grant from Heising-Simons Foundation  
 08 Aug 2017: A survey for users accessing arXiv programmatically  
 20 Apr 2017: Applied Physics subject area added to arXiv  
 10 Mar 2017: New members join arXiv Member Advisory Board  
 See cumulative "What's New" pages. Read robots beware before attempting any automated download

### Physics

- Astrophysics ([astro-ph new](#), [recent](#), [find](#))  
 Includes: Astrophysics of Galaxies; Cosmology and Nongalactic Astrophysics; Earth and Planetary 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](#))
- Nonlinear Sciences ([nlin new](#), [recent](#), [find](#))  
 includes: Adaptation and Self-Organizing Systems; Cellular Automata and Lattice Gases; Chaotic Dynamics; Exactly Solvable and Integrable Systems; Pattern Formation and Solitons

# Arxiv hoy

- ❑ Cubre los campos de la física, matemáticas y ciencias de la computación, biología, finanzas y estadística y los artículos eran y son depositados por los autores antes de su envío a las revistas especializadas, hay una revisión de pares.
- ❑ Durante todos estos años se ha discutido el modelo de ArXiv para otros campos del saber.
- ❑ Lo previo está ligado a las diferentes formas de comunicar según el área del saber.



# Hitos

**1993:** 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 AA, poniendo a su disposición software libre, cumpliendo tareas de repositorio con sus propios servidores para los artículos de investigación y alojando las iniciativas europeas en el ámbito.



# Hitos

En **1994**, Stevan Harnad lanza la iniciativa por el **autoarchivo** (*self-archiving*): el propio autor deposita su trabajo en un lugar confiable.



## Harnard: 1994


- ❑ No cuestionaba la utilidad del proceso de publicación formal, sino su capacidad y rapidez comunicacional.
- ❑ La publicación académica tiene dos funciones que vale la pena distinguir: la práctica comunicativa, cuya prioridad es ser leído, y la publicación funcionalista, que sirve para dar status, autoría y recibir subsidios.
- ❑ La distribución de *preprints* obedecía a la primera.



# Harnard: 1994

- ❑ Recoge la iniciativa ArXiv y la desarrolla para su aplicación en otros campos y la piensa ya no sólo para *preprints* sino para trabajos con revisión.
- ❑ Analiza el impacto del depósito de artículos científicos en un archivo de acceso abierto (en aquel entonces, un sitio FTP).
- ❑ Suscitó una discusión sobre todo el sistema de comunicación científica y constituyó uno de los primeros reclamos por hacer más abierta la ciencia.

# Hitos: 1996 ND LTD



Search this site ▾

[News](#) [Videos](#) [Community](#) [Thesis Resources](#) [Global ETD Search](#)

**About**

- Mission, Goals, and History
- Financial Information
- Official Documents
- FAQ
- List of Members
- ETD Metadata Providers

**Contact Us**

- Directory
- Board of Directors
- Committees & Working Groups

**Membership**

- List of Members
- Membership Benefits

**NDLTD Celebrates its 20th Anniversary**

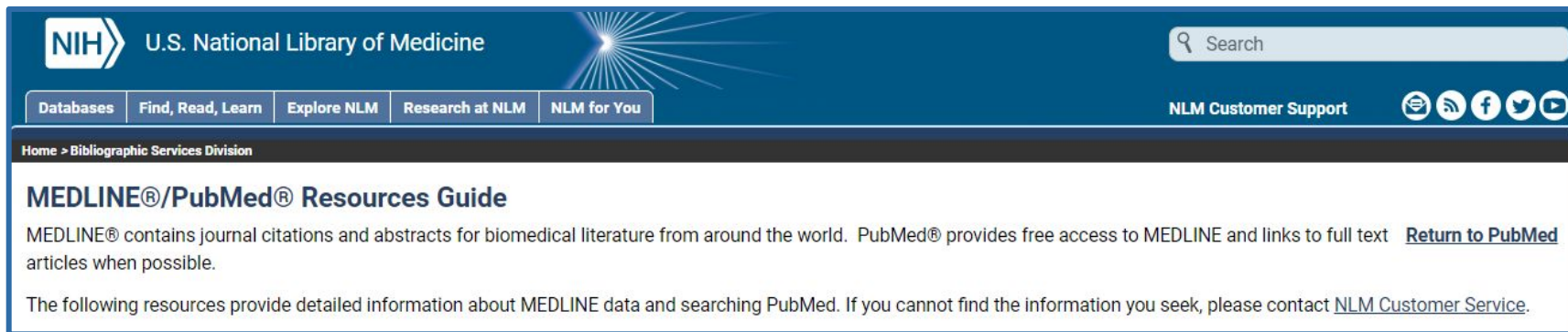
Edward A. Fox  
Executive Director and Chairman of the Board, NDLTD  
August 2, 2017

Almost 30 years ago, in November 1987, a first meeting was held to explore using SGML (Standard Generalized Markup Language) for dissertations as part of the Electronic Manuscript Project. This event and others in the first decade of activities related to the Networked Digital Library of Theses and Dissertations, NDLTD (<http://ndltd.org>) are described in <http://fox.cs.vt.edu/pub/ETD/ETD.html>. Many early key events had been in the USA, including the U.S. Department of Education award by 9/1/1996 to "Improve Graduate Education with a National Digital Library of Theses and Dissertations (NDLTD)." Later in 1996, Virginia Tech decided to require electronic theses and dissertations (ETDs), and set a start date for this mandate, 1/1/1997. This marked the first instance (with hundreds to follow) of requiring ETDs at colleges and universities around the world. See a



# Hitos: 1997

En 1997 se inicia también el acceso libre a MEDLINE a través de 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.



The screenshot shows the NIH U.S. National Library of Medicine website. The header includes the NIH logo and a search bar. Below the header are navigation tabs for Databases, Find, Read, Learn, Explore NLM, Research at NLM, and NLM for You. The main content area is titled "MEDLINE®/PubMed® Resources Guide" and contains the following text:

Home > Bibliographic Services Division

## MEDLINE®/PubMed® Resources Guide

MEDLINE® contains journal citations and abstracts for biomedical literature from around the world. PubMed® provides free access to MEDLINE and links to full text [Return to PubMed](#) articles when possible.

The following resources provide detailed information about MEDLINE data and searching PubMed. If you cannot find the information you seek, please contact [NLM Customer Service](#).

# Hitos: 1998

- ❑ 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”.*



## Después de 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 los trabajos y lanzan revistas de AA por cuenta propia, apoyados por SPARC y su propuesta “Declaración de Independencia” (2001), que fue una guía para crear publicaciones científicas controladas por los mismos académicos.

# Problemas del crecimiento

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



## Hitos: 1999

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



# OAI-PMH

La iniciativa 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.



# OAI-PMH

El protocolo OAI-PMH propone unos lineamientos generales tanto para listar y recuperar (cosechar) metadatos de un repositorio (**OAI Service Providers**), como también para exponer recursos (**OAI Data Providers**) para que puedan ser cosechados por aplicaciones externas.

Estos lineamientos proponen la organización de los recursos en conjuntos (*sets*), el uso del estándar XML para la representación y transporte de recursos (vía HTTP), y un conjunto de seis verbos necesarios para interactuar, como por ejemplo identificar el repositorio, listar conjuntos, listar formatos de metadatos soportados u obtener registros.

# 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.

Pros y contras...



# OAI-PMH

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

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

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

# Convención de Santa Fe, Albuquerque

La convención consideró como cruciales los siguientes componentes:

- ❑ 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).



# Hitos: 2000 PUBMED Central

NCBI Resources How To Sign in to NCBI

PMC  
US National Library of Medicine  
National Institutes of Health

Journal List Advanced Help

Search

**PMC**

PubMed Central® (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).

**PubReader**

A whole new way to read scientific literature at PubMed Central

**Get Started**

- [PMC Overview](#)
- [Users' Guide](#)
- [Journal List](#)
- [PMC FAQs](#)
- [PMC Copyright Notice](#)

**Participate**

- [Information for Publishers](#)
- [Add a Journal to PMC](#)
- [Participation Agreements](#)
- [File Submission Specifications](#)
- [File Validation Tools](#)

**Keep Up to Date**

- [New in PMC | RSS](#)
- [PMC Announce Mail List](#)
- [Utilities Announce Mail List](#)
- [Tagging Guidelines Mail List](#)

**Other Resources**

- [PMC International](#)
- [Text Mining Collections](#)

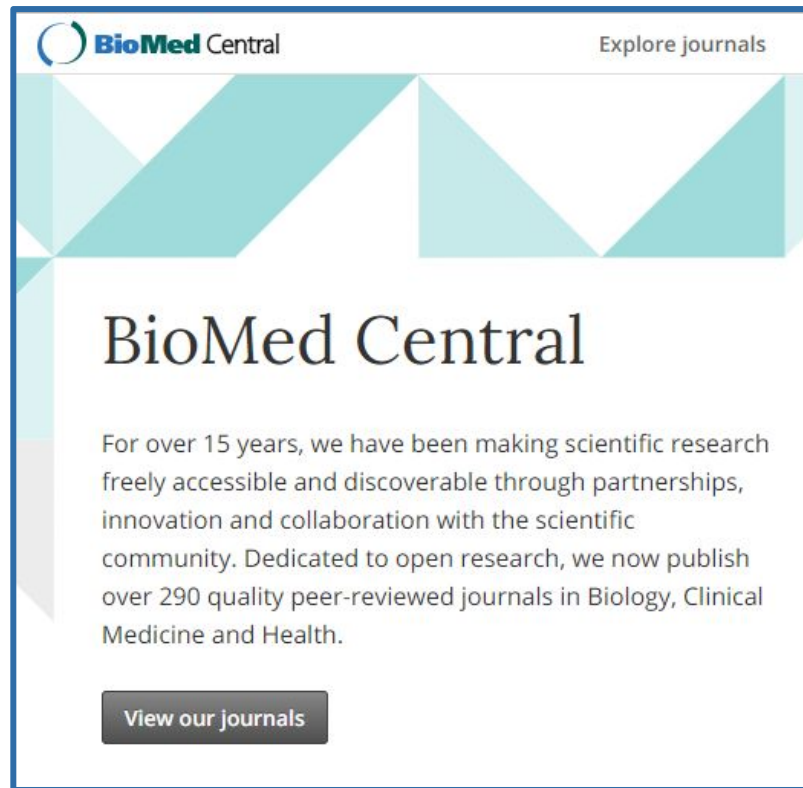
**4.5 MILLION Articles**  
are archived in PMC.

**Public Access**

- [Fundors and PMC](#)
- [How Papers Get Into PMC](#)

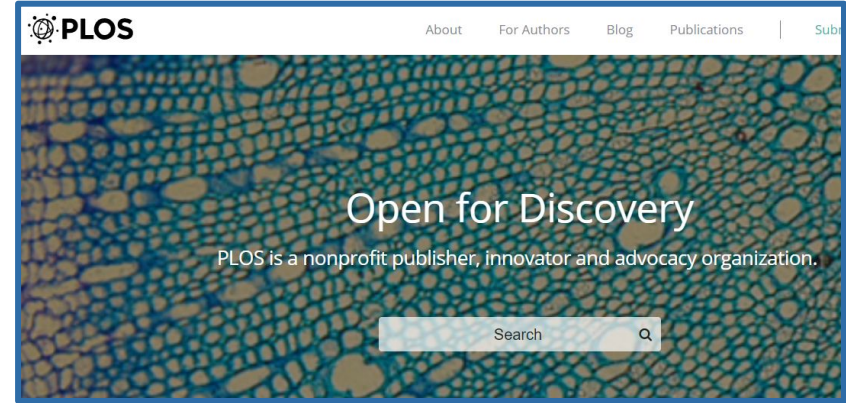
## Hitos: 2000

Lanzamiento de la primera iniciativa de un editor privado: **BiomedCentral**. En la actualidad, edita un importante número de revistas accesibles a texto completo y que pone en marcha un sistema de APC.



# Hitos: 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.



Publication fees vary by journal and are payable upon article acceptance.



***PLOS ONE***

\$1,495 USD



***PLOS Biology***

\$2,900 USD



***PLOS Medicine***

\$2,900 USD



***PLOS Computational Biology***

\$2,250 USD



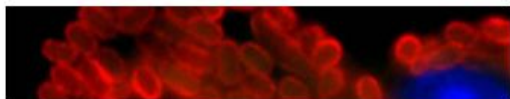
***PLOS Genetics***

\$2,250 USD



***PLOS Neglected Tropical Diseases***

\$2,250 USD



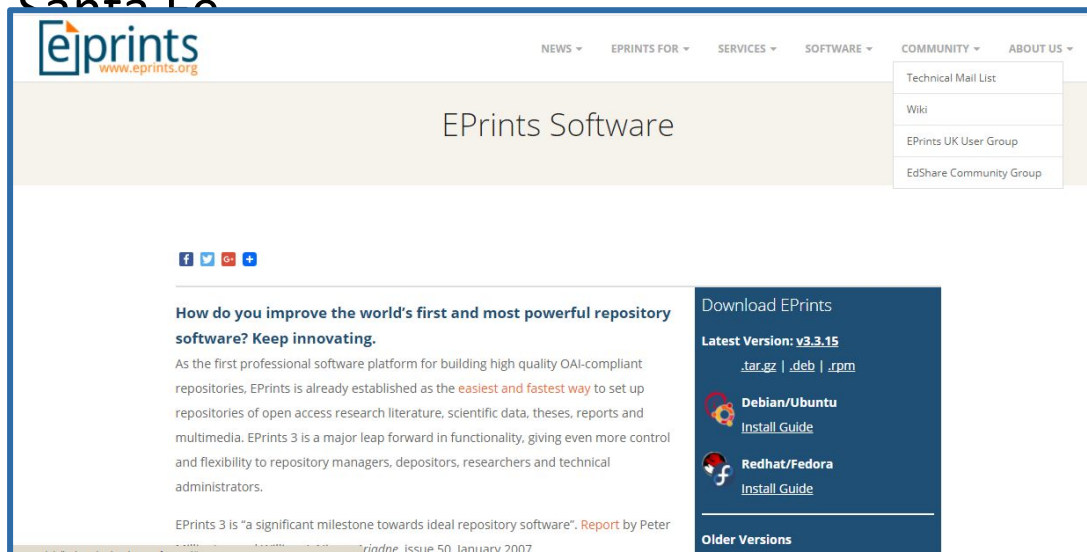
***PLOS Pathogens***

\$2,250 USD



# Hitos: 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



The screenshot shows the Eprints Software website. The header includes the Eprints logo (www.eprints.org) and navigation menus for NEWS, EPRINTS FOR, SERVICES, SOFTWARE, COMMUNITY, and ABOUT US. A dropdown menu for COMMUNITY is open, showing links to Technical Mail List, Wiki, Eprints UK User Group, and EdShare Community Group. The main content area features a section titled "How do you improve the world's first and most powerful repository software? Keep innovating." with a sub-heading "As the first professional software platform for building high quality OAL-compliant repositories, Eprints is already established as the easiest and fastest way to set up repositories of open access research literature, scientific data, theses, reports and multimedia. Eprints 3 is a major leap forward in functionality, giving even more control and flexibility to repository managers, depositors, researchers and technical administrators." Below this is a quote: "Eprints 3 is 'a significant milestone towards ideal repository software'. Report by Peter ...". To the right, there is a "Download Eprints" section with the "Latest Version: v3.3.15" and download links for .tar.gz, .deb, and .rpm. It also includes "Debian/Ubuntu Install Guide" and "Redhat/Fedora Install Guide" sections, and a link for "Older Versions".



# Acceso Abierto: declaraciones

**Las tres B:** 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** (febrero de 2002).
- ❑ **Bethesda Statement on Open Access Publishing** (junio de 2003).
- ❑ **Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities** (diciembre de 2003).

# Budapest OA Initiative – Open Access

En ella se recomiendan las modalidades de publicación en:

- ❑ revistas de **acceso abierto**
- ❑ **autoarchivo** en archivos electrónicos abiertos



# Declaración de Budapest (2002)

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

# Declaración de Bethesda (2003)

- ❑ Derecho libre, irrevocable, universal y perpetuo de **acceso**
- ❑ **Licencia de uso**
- ❑ **Apropiada atribución de la autoría**
- ❑ Depósito en **repositorios** en línea apoyado por una **institución académica**
- ❑ **Interoperabilidad** y el **archivado a largo plazo**

# Declaración de Bethesda (2003)

La Declaración de Bethesda sobre Publicación de Acceso Abierto añade que para que un trabajo sea de acceso abierto, tiene que cumplir dos condiciones:

- 1.** El/los autor/es y el/los propietario/s de los derechos de propiedad intelectual otorgan a los usuarios un derecho libre, irrevocable, universal y perpetuo de **acceso** y una **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.



## Declaración de Bethesda (2003)

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).

# Declaración de Berlín (2003)

Ratifica lo anterior y afirma:

- ❑ **Diseminación del conocimiento**
- ❑ **Acceso abierto por medio de Internet**



# Declaración de Berlín (2003)

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



Tabla 1 . Iniciativas conjuntas sobre el acceso abierto

Iniciativa	Objetivo	Restricciones	Estrategias tecnológicas	Beneficios	Participantes
Declaración de Budapest (2002)	Distribución electrónica mundial de la literatura publicada en revistas científicas	El derecho del autor con relación al control sobre la integridad de su trabajo y el derecho de ser reconocido y citado correctamente	Archivos y publicaciones periódicas estandarizados	Acelera la investigación, enriquece la educación, comparte el aprendizaje	Académicos, fundaciones, gobiernos, universidades, laboratorios, y fideicomisos
Declaración de Bethesda (2003)	Proporcionar acceso abierto a la literatura científica primaria	Se sujeta a las atribuciones adecuadas de autoría	Cualquier medio digital	Disemina los resultados de investigación inmediata, amplia y eficazmente	Organizaciones académicas y científicas, científicos, editores, agencias gubernamentales y bibliotecarios
Declaración de Berlín (2003)	Garantizar a todos los usuarios por igual el acceso a un trabajo erudito	Sujeto al reconocimiento de autoría	Formato electrónico estándar	Crea una amplia fuente de conocimiento humano y patrimonio cultural	Todos los productores de conocimiento y poseedores de herencia cultural

Fuente: Rodríguez Gallardo, A. (2008). “Elementos que fundamentan el Acceso Abierto”. *Investigación Bibliotecológica*, 22(44), 161-182.



- ❑ Reconocimiento creciente de que buena parte de la investigación es financiada con fondos públicos y que para maximizar la disseminación de sus resultados, deben estar disponibles en acceso abierto.
- ❑ Las TIC facilitando la creación de distintas plataformas.
- ❑ 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.



# Estrategias en AA

- 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**.

## Estructura



# Ventajas del acceso abierto

- Mayor accesibilidad, uso, anticipación en el tiempo de publicación y calidad.
- Preservación a largo plazo.
- Acceso a otros servicios: identificadores persistentes, estadísticas de uso, sistemas de búsqueda avanzados ,etc.
- Aumento del impacto de los artículos a través de mayor cantidad de citas.
- Acelera el circuito de la producción de conocimiento
- Los artículos en abierto se descargan y se leen 2 veces más que los que no se difunden en abierto.

# ¿Qué es un repositorio?



# Repositorios: una posible definición y características

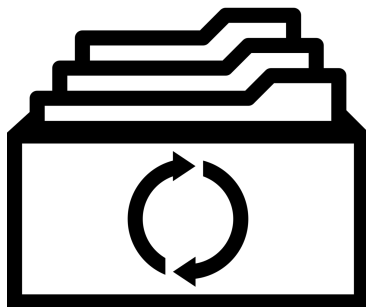
Los repositorios son archivos digitales provistos de un conjunto de servicios web centralizados, creados para organizar, gestionar, preservar y ofrecer acceso libre a la producción científica, académica o de cualquier otra naturaleza cultural, en soporte digital, generada por los miembros de una organización.

- Los repositorios deben tener una política definida y asegurar:
  - a. El **autoarchivo**
  - b. La **interoperabilidad**
  - c. El **libre acceso**
  - d. La **preservación a largo plazo**



# Autoarchivo

El autoarchivo es la forma más fácil y rápida de colaborar con el repositorio. En 1994 Stevan Harnad, lanza la iniciativa por el autoarchivo (self-archiving): **el propio autor deposita su trabajo en un lugar confiable**. En el caso de SEDICI y mediante un formulario de cinco pasos, cualquier alumno, docente o investigador de la UNLP puede subir su material, que luego será catalogado y depositado en la colección que corresponda.



# Repositorios: características

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.



# Las ventajas del repositorio

- Se ofrece un punto de acceso uniforme a la información de la institución y del autor.
- El material es catalogado y descrito en profundidad, maximizando su accesibilidad.
- Se preserva la integridad digital de las obras.
- Trabaja bajo estándares que permiten la exposición automática de los registros a otros portales académicos.
- El material se distribuye junto al detalle de sus derechos de uso, notificando a los lectores los usos permitidos.

# Ejemplos de Repositorios



# Tipos de repositorios

## Institucionales

Almacenar, preservar y dar acceso a los materiales de una institución

- **SEDICI** - Universidad Nacional de La Plata - Argentina  
[sedici.unlp.edu.ar](http://sedici.unlp.edu.ar)
- **Digital CSIC** - CSIC - España  
[digital.csic.es](http://digital.csic.es)

## Huérfanos

Creados para autores investigadores que no pueden publicar en ningún otro repositorio institucional

- **DEPOT** (Reino Unido)  
<http://opendepot.org/>
- **UAR Politics** (Portugal)
- **ZENODO** (UE, OpenAire)  
<https://www.zenodo.org/>

## Temáticos

Almacenar, preservar y dar acceso a los materiales según un tema o una disciplina

- **PubMed Central**- Biology and Biochemistry; Chemistry and Chemical Technology; Health and Medicine  
[www.ncbi.nih.gov/prnc/](http://www.ncbi.nih.gov/prnc/)
- **RePEc (Research Papers in Economics)** - Business Economics  
[repec.org](http://repec.org)

## Agregadores

Recolectan contenidos de otros repositorios por temas (o por tipo de documento o geografía)

- **Base**  
[www.base-search.net](http://www.base-search.net)
- **OATD**  
[www.oatd.org](http://www.oatd.org)
- **BDU2**  
[bdu.siu.edu.ar/cgi-bin/repoprpt.pl](http://bdu.siu.edu.ar/cgi-bin/repoprpt.pl)

## De datos

Almacenar y preservar datos científicos

- **ODISEA**  
[odisea.ciepi.org](http://odisea.ciepi.org)



# Modelos de Repositorios Institucionales

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.

- UNCUIYO <http://bdigital.uncu.edu.ar/>
- Universidad Autónoma de México <http://www.rad.unam.mx/index.php/index>
- Universidad del Pacífico <http://repositorio.up.edu.pe/>
- Universidad Central del Ecuador <http://www.dspace.uce.edu.ec/>



# Pasos para crear un repositorio institucional

- ¿Cómo se establece un repositorio institucional?
- ¿Cuáles son los aspectos importantes de su definición?
- Análisis previo.
- Reflexiones institucionales importantes.
- Organización del servicio.
- Determinación de políticas.



# Pasos para crear un repositorio institucional

Aprender examinando otros Repositorios Institucionales.

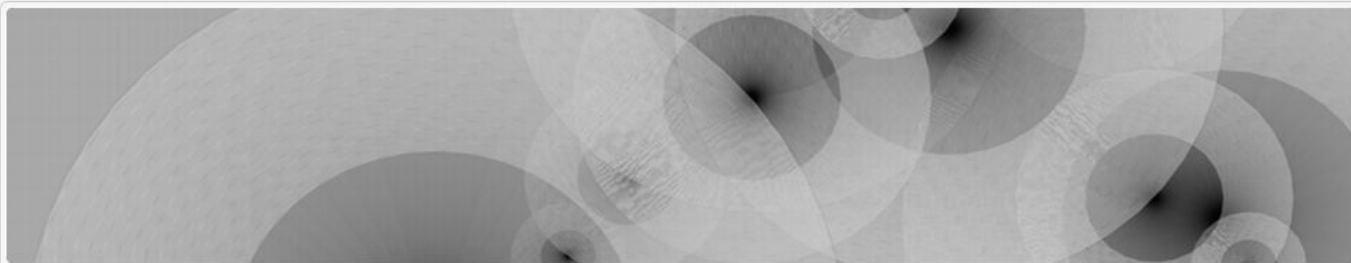
Desarrollar una definición y un plan de servicio:

- Realizar una evaluación de las necesidades
- Desarrollar un modelo de coste basado en este plan
- Crear una planificación y un horario en relación a los recursos humanos y costes



- Desarrollar políticas de actuación que gestionen la recopilación de contenidos, su distribución y mantenimiento
- Formar el equipo
- Tecnología: elegir e instalar el software
- Marketing: Difundir el servicio
- Poner en funcionamiento





## Cómo abordar un proyecto de Repositorio Institucional?

Enviado por Emilio Lorenzo el 2011/01/15

La implementación de un Repositorio en una organización, Biblioteca, Archivo o Museo, exige una considerable planificación y esfuerzo. Como pasos clave para tener éxito en la creación un Repositorio Institucional podríamos apuntar los siguientes:

- Evaluación y priorización de las necesidades de la institución.
- Desarrollar una visión a alto nivel del Repositorio.
- Identificación de los factores y agentes clave de la implantación.
- Planificación detallada del proyecto y elaboración del modelo de costes y sostenibilidad.
- Consideración de los aspectos relevantes de propiedad intelectual.
- Diseño de políticas que gestionen la recopilación de contenidos, su distribución y mantenimiento.
- Instalaciones HW y SW. Actuaciones tecnológicas necesarias para la correcta integración del repositorio en la infraestructura existente.
- Desarrollos específicos y extensiones DSpace.
- Carga de colecciones digitalizadas preexistentes.
- Formación a usuarios generales, gestores del repositorio y personal técnico.
- Marketing y difusión del repositorio institucional.
- Puesta en funcionamiento del mismo y soporte.



DSpace Registered Service Provider

### Contenidos

3.0 ant autenticación authority control  
 barra lateral Batch Metadata Editing CLI Control de  
 Autoridades correo curation cursos

discovery dspace easydeposit  
 eventos facebook filter-media GUI handle  
 import javaAPI jpg jspui language  
 packs ldap Lucene maven  
 messages\_es.xml mirage OAI  
 OR2010 pdf redes sociales repositorio  
 institucional stackable sword swordapp  
 thumbnail tomcat6 version 4 version 5  
 version 6 xmlui xpdf xsl

### Síguenos en twitter



Arvo Consultores

sedici.unlp.edu.ar  
 prebi.unlp.edu.ar



# Cómo crear un Repositorio Institucional

## Manual LEADIRS II

Por Mary R. Barton, MIT Libraries, [at] mit.edu  
y Margaret M. Waters, asesora, [at] aya.yale.edu  
patrocinado por The Cambridge-MIT Institute (CMI).

Copyright © 2004-2005 MIT Libraries

Learning about Digital Institutional Repositories: Creating an Institutional Repository: LEADIRS Workbook. 2004. Disponible en:  
[https://dspace.mit.edu/bitstream/handle/1721.1/26698/Barton\\_2004\\_Creating.pdf?sequence=1](https://dspace.mit.edu/bitstream/handle/1721.1/26698/Barton_2004_Creating.pdf?sequence=1)



## Cuestiones a tener en cuenta

- Colecciones de tamaño manejable en lugar de grandísimas colecciones.
- Opinión institucional vinculada a las colecciones.
- 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.



# ¿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.

# Cómo crear una definición de servicio

Cada Universidad que desarrolla un Repositorio Institucional en línea necesita investigar y especificar políticas de actuación y regulaciones para sus colecciones.

- Política de contenidos
- Política de acceso a datos
- Política de depósito
- Política de preservación
- Política de Metadatos
- Política de Servicios

# 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 pero éste no es el único **principio organizativo**:

- 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: Emergencia Hídrica.
- 3. Comunidades de Interés:** Un grupo ad hoc, de departamentos interfacultades. La organización de las comunidades de contenido depende de los intereses y acuerdos de sus integrantes. Ejemplo: ANAV



# Interoperabilidad y SNRD



# Interoperabilidad

- Es la capacidad del repositorio de interoperar con otros sistemas mediante el cumplimiento de estándares internacionales de exposición de recursos (OAI-PMH y OpenSearch).
- Los distintos protocolos existentes en la actualidad sirven al repositorio tanto para generar servicios para sus usuarios y/o su institución como a terceros y otorgan mayor alcance e impacto a los recursos del repositorio.

- El cumplimiento de estándares de interoperabilidad no asegura por sí mismo una correcta disseminación de la producción institucional.

Debe garantizarse:

- La calidad en la catalogación de los recursos
  - El control de calidad de los objetos digitales que se almacenan
  - Las actividades relativas a la preservación digital
- La interoperabilidad con el repositorio puede realizarse desde el entorno del repositorio hacia el mismo repositorio, o desde el repositorio hacia su entorno.



# Interoperabilidad Entorno Repositorio

Esto ocurre cuando aplicaciones o servicios externos al repositorio desean enviarle sus recursos digitales para que sean alojados, preservados y difundidos allí. Ejemplo: revistas científicas y académicas de la propia institución que se quieran agregar automáticamente.

Existe un protocolo de comunicación llamado SWORD (Simple Web-service Offering Repository Deposit) que permite realizar el depósito remoto de objetos digitales en el repositorio.

# SWORD PKP DSPACE

Muchos sistemas son compatibles con dicho protocolo: en particular, el software de PKP **Open Journals System** (OJS) posee un módulo que permite el envío de trabajos por este protocolo, y a su vez, **DSpace**, el software utilizado para gestionar el repositorio SEDICI, posee un módulo que permite la ingesta de recursos a partir de **SWORD**.

En el caso de la UNLP, por ejemplo, el **Portal de Revistas** utiliza OJS, y posee hasta la fecha 20 revistas periódicas; con cada nuevo número de cada revista, se realiza el depósito remoto en el repositorio de todos los trabajos, lo que toma menos de un minuto y asegura el arribo al repositorio de los mismos metadatos cargados en cada artículo de la revista.



# Interoperabilidad Repositorio Entorno

La interoperabilidad desde el repositorio hacia el entorno es una de sus funciones centrales, existen normas y directrices internacionales aseguran la presencia del repositorio en índices, directorios globales y agregadores de recursos.

La iniciativa de Open Archives con su protocolo para la cosecha de metadatos conocido como **OAI PMH** (Open Archive Initiative-Protocol for Metadata Harvesting), que sienta las bases para que los repositorios digitales compartan los registros correspondientes a los recursos que almacenan.

# OpenSearch

Existen otras alternativas que permiten recuperar recursos del repositorio para alojarlos y/o exponerlos en otros ámbitos. Un caso interesante es el uso del conjunto de herramientas **OpenSearch**, que permite codificar en una URL los términos y filtros de una búsqueda.

En el caso del repositorio, estos términos se conforman con las palabras por las que se realiza la búsqueda, y los filtros corresponden al ámbito de búsqueda (por ejemplo, una colección particular), a la cantidad de recursos a recuperar o los facets, descriptos previamente, que se aplicarán (tipo de documento, fecha de publicación, etcétera).

# Servicios adicionales: Interoperabilidad OpenSearch

## Producción personal / institucional

### Dra. Marisa Raquel De Giusti



Marisa De Giusti es Dra. en Ciencias Informáticas, Ingeniera en Telecomunicaciones y Profesora en Letras todos títulos de la UNLP, asimismo se desempeña como Investigadora de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires. Es Profesora de Grado y Posgrado en la Facultad de Informática de la UNLP donde integra y coordina la CAH en Tecnologías. Desde 1990 trabaja con el Consorcio Iberoamericano para Educación en Ciencia y Tecnología (ISTEC) y es actualmente la Directora de Investigación y Desarrollo del Proyecto Library linkage (LibLink) de ISTEC nominación que obtuvo tras presentarse al Concurso Internacional abierto para el mencionado cargo en el presente año 2007.

- SEBICI Repositorio SEDICI
- Google Citations
- ResearchGate
- ResearcherID
- ORCID
- Twitter
- Email

A fines de 1996 presentó a las autoridades de la Secretaría de Ciencia y Técnica de la UNLP el Proyecto de Enlace de Bibliotecas (PrEBI) y el mismo fue creado en Mayo de 1997, siendo desde entonces su Directora. De modo similar en el año 2003 presentó el Proyecto Servicio de Difusión de la Creación Intelectual (SeDICI) el cual se incorporó a PrEBI y el cual dirige desde su creación.

En su actividad profesional ha incursionado por muy diversas áreas científicas y cuenta con más de 50 publicaciones internacionales, en los últimos 2 años las mismas están dedicadas a las áreas de la información y bibliotecas.

#### Publicaciones:

#### Artículo

- Fostering the institutional repository through policies and interoperability with online services: the case of La Plata National University  
Autores: Villarreal, Gonzalo Luján-Terruzzi, Franco Agustín-Lira, Ariel Jorge-De Giusti, Marisa Raquel-Texier, José  
Compartir: [f](#) [t](#) [s](#) [in](#)
- El desarrollo de software dirigido por modelos en los repositorios institucionales  
Autores: Texier, José-De Giusti, Marisa Raquel-Gordillo, Silvia Ethel  
Compartir: [f](#) [t](#) [s](#) [in](#)
- DSpace como herramienta para un repositorio de documentos administrativos en la Universidad Nacional Experimental del Táchira

The screenshot shows the SEDICI website interface. At the top, there is a header with the logo 'CESGI Centro Propio de Servicios en Gestión de Información' and 'CIC'. Below the header is a navigation menu with items: INICIO, INSTITUCIONAL, GRUPOS DE TRABAJO, SERVICIOS, PUBLICACIONES, BLOG, and CONTACTO. The 'PUBLICACIONES' menu is expanded, showing options: Artículos, Congresos, Objetos de conferencias, Tesis (highlighted), and Cursos. Capacitaciones. The main content area is titled 'Tesis' and lists several entries:

- La representación de recursos usando la metodología del Desarrollo Dirigido institucional: Caso de estudio: SEDICI  
Autores: Texier, José  
Tipo de documento: Tesis de doctorado  
Compartir: [f](#) [t](#) [s](#) [in](#)
- Una metodología de evaluación de repositorios digitales para asegurar la preservación en el tiempo y el acceso a los contenidos  
Autores: De Giusti, Marisa Raquel  
Tipo de documento: Tesis de doctorado  
Compartir: [f](#) [t](#) [s](#) [in](#)
- Un entorno de aprendizaje y una propuesta de enseñanza de Simulación de Eventos Discretos con GPSS  
Autores: Villarreal, Gonzalo Luján  
Tipo de documento: Tesis de doctorado  
Compartir: [f](#) [t](#) [s](#) [in](#)
- Herramienta de validación aplicada a las tareas de gestión de calidad en un repositorio digital  
Autores: Terruzzi, Franco Agustín  
Tipo de documento: Tesis de grado  
Compartir: [f](#) [t](#) [s](#) [in](#)

On the right side of the page, there is a search bar, a section for 'Entradas recientes' (Recent entries) with a link to 'Tercer Congreso Internacional de Ciencia y Tecnología', and a 'SEDICI BLOG' section with a link to '12 de agosto: día de la UNLP El Bicentenario de la Independencia 25 aniversario del IRI Versión final.doc, disponible en el repositorio Estudios e investigaciones sobre gripe A'. At the bottom right, there is a 'Noticias CIC' section.

# Resoluciones UNLP - Secretaría de Ciencia y Técnica

Resoluciones UNLP

ecyt.presi.unlp.edu.ar/WordPress/?p=96

UNLP Investigación

Buscador de proyectos

Centros, Lab. e Institutos

Rev. Científicas de la UNLP

Indicadores CyT

Banco Evaluadores

Categorización

2014-Convocatoria

2014-Guías y Manuales

UNIVERSIDAD NACIONAL DE LA PLATA

Institucional Enseñanza Investigación Extensión Posgrado Gestión RSS

← Resolución MINCYT  
Manual de Procedimientos del Programa de Incentivos →

**Resoluciones UNLP**  
Marzo 26th, 2009 - Sin comentarios

**Res. N°413 / 2010** – 6 mayo 2010  
Becas Otorgadas 2010

**Res. N°357 / 2010** – 23 Abril 2010  
Prórroga 2010 Becas

**Res. N°871 / 09** – 04 noviembre 2009  
Programa de Retención de Recursos Humanos/ Otorgadas 2009

**Res. N°692 / 09** – 10 septiembre 2009

Buscar en SciCYT

Buscar...

Suscribirse por mail

Email:

Suscribir

Accesos Rápidos

SIGEVA UNLP

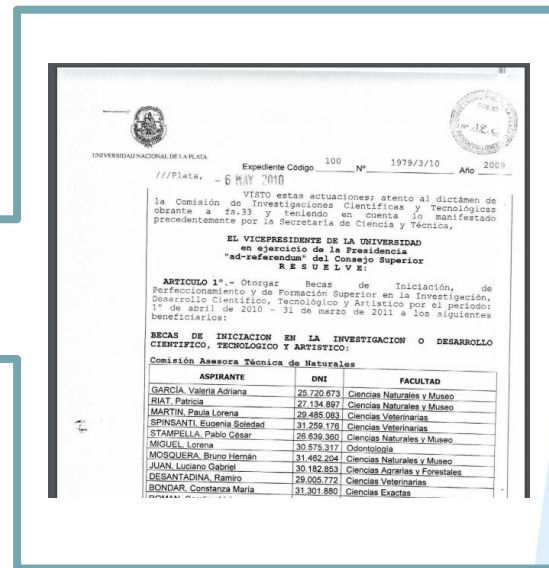
WEBAcreditación

WEBJovenes

WEBProyectos

WEBViajes

SIGEVA Instructivos



# Directrices DRIVER

- 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



# OAI-PMH Driver OpenAIRE

Como puede observarse, este protocolo no fuerza el uso de una norma de catalogación única, y es aquí donde surgen las distintas directrices internacionales, como por ejemplo **OpenAIRE** y **Driver**, que aseguran que todos aquellos que quieran participar de las redes de intercambio nacionales e internacionales «hablen el mismo idioma» cuando de registros de catalogación se trate.



# ▶ OPEN SCIENCE FAIR ▶

Workflows for Open Science

**ATHENS**

September 6-8, 2017



OpenAIRE calls on the European Parliament to halt potentially harmful copyright reform



Registration now open

## RESEARCHERS

Why Open Access. How to comply. What services to use.

## DATA PROVIDERS

How to make your content more visible. What to do to increase quality. How to join.



[OpenAIRE Guidelines for Literature Repositories](#)

[OpenAIRE Guidelines for Data Archives](#)

[OpenAIRE Guidelines for CRIS Managers based on CERIF-XML](#)

[Contributing](#)

[How the Horizon 2020 Open Access requirements are met](#)

## OpenAIRE Guidelines

Welcome to the OpenAIRE Guidelines. The intention of this is to provide a public space to share OpenAIREs work on interoperability and to engage with the community. The OpenAIRE Guidelines helps repository managers expose publications, datasets and CRIS metadata via the OAI-PMH protocol in order to integrate with OpenAIRE infrastructure.



OpenAIRE Guidelines consists of three guidelines for publication repositories, data archives and CRIS systems respectively:

- [OpenAIRE Guidelines for Literature Repositories](#)
- [OpenAIRE Guidelines for Data Archives](#)
- [OpenAIRE Guidelines for CRIS Managers based on CERIF-XML](#)

The guidelines specifically provide guidance on how to specify:

- Access right
- Funding information
- Related publication and datasets.

### Participate

You are invited to participate by commenting or editing the content. See our guide for how to get started:

- [Contributing](#)

### OpenAIRE Validator





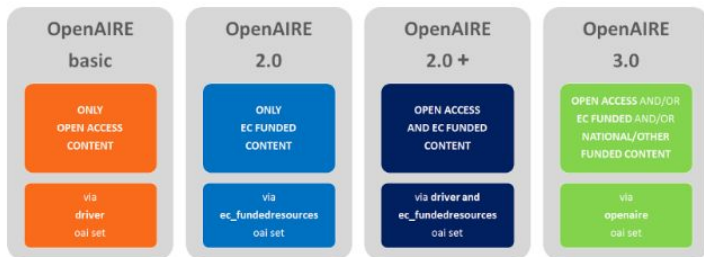
OpenAIRE has developed into an open access publications infrastructure, as well as making visible funded research outcomes (at present EC-FP7 and Wellcome Trust). We now aggregate from 450 data providers, including many institutional repositories.

### WHY IS MY REPOSITORY REGISTERED IN OPENAIRE?

When your repository is harvested by OpenAIRE, you will comply with the European Commission's policy on Open Access, as well as the upcoming Horizon2020 mandate. All open access records, and records (closed or open) with funding information attached will be made more visible via OpenAIRE. This will assist the monitoring of OA policies.

### SO WHAT HAS CHANGED?

- 1 - During 2014, all the DRIVER Open Access repositories were brought into OpenAIRE. This merging of OA data with FP7 materials will form a large, more comprehensive knowledge-space.
- 2 - OpenAIRE is also starting to support a wider range of scholarly resources and research information (see the [content acquisition policy](#)). To this extent, OpenAIRE now harvests CRIS systems and research data sets. Links to the OpenAIRE guidelines can be found below.
- 3 - For literature repositories, OpenAIRE now specifies different status of compatibility depending on the guidelines version and the exposed content via OAI-Sets.



### WHAT CAN I DO TO ENSURE AND IMPROVE OPENAIRE COMPATIBILITY?

Please carefully analyse the proper version of the OpenAIRE guidelines, especially the metadata and vocabulary application profile, OAI-PMH usage instructions



## OpenAIRE Guidelines for Literature Repositories

- Introduction
- Use of OAI-PMH
- Use of OAI-DC
- Application Profile Overview
- 1. Title (M)
- 2. Creator (M)
- 3. Project Identifier (MA)
- 4. Access Level (M)
- 5. License Condition (R)
- 6. Embargo End Date (MA)
- 7. Alternative Identifier (R)
- 8. Publication Reference (R)
- 9. Dataset Reference (R)
- 10. Subject (MA)
- 11. Description (MA)
- 12. Publisher (MA)
- 13. Contributor (R)
- 14. Publication Date (M)
- 15. Publication Type (M)
- 16. Publication Version (R)
- 17. Format (R)
- 18. Resource Identifier (M)
- 19. Source (R)
- 20. Language (R)
- 21. Relation (O)
- 22. Coverage (R)

# OpenAIRE Guidelines for Literature Repositories

- Introduction
- Use of OAI-PMH
- Use of OAI-DC
- Application Profile Overview

### Application Profile:

- 1. Title (M)
- 2. Creator (M)
- 3. Project Identifier (MA)
- 4. Access Level (M)
- 5. License Condition (R)
- 6. Embargo End Date (MA)
- 7. Alternative Identifier (R)
- 8. Publication Reference (R)
- 9. Dataset Reference (R)
- 10. Subject (MA)
- 11. Description (MA)
- 12. Publisher (MA)
- 13. Contributor (R)
- 14. Publication Date (M)
- 15. Publication Type (M)
- 16. Publication Version (R)
- 17. Format (R)
- 18. Resource Identifier (M)
- 19. Source (R)
- 20. Language (R)
- 21. Relation (O)
- 22. Coverage (R)
- 23. Audience (R)

[Previous](#)[Next](#)

OpenAIRE Guidelines for Literature Repositories

Introduction  
Use of OAI-PMH  
Use of OAI-DC

**Application Profile Overview**

1. Title (M)  
2. Creator (M)  
3. Project Identifier (MA)  
4. Access Level (M)  
5. License Condition (R)  
6. Embargo End Date (MA)  
7. Alternative Identifier (R)  
8. Publication Reference (R)  
9. Dataset Reference (R)  
10. Subject (MA)  
11. Description (MA)  
12. Publisher (MA)  
13. Contributor (R)  
14. Publication Date (M)  
15. Publication Type (M)  
16. Publication Version (R)  
17. Format (R)  
18. Resource Identifier (M)  
19. Source (R)  
20. Language (R)  
21. Relation (O)  
22. Coverage (R)  
23. Audience (R)

OpenAIRE Guidelines for Data Archives

OpenAIRE Guidelines for CRIS  
Managers based on CRIIE-XML

# Application Profile Overview

This documentation uses the following namespace abbreviation:

- `dc` : <http://purl.org/dc/elements/1.1/>

OpenAIRE-Field	OAI-DC Element	Refinement by vocabulary
Title (M)	dc:title	
Creator (M)	dc:creator	
Project Identifier (MA)	dc:relation	<a href="http://info:eu-repo/grantAgreement/">info:eu-repo/grantAgreement/</a>
Access Level (M)	dc:rights	<a href="http://info:eu-repo/semantics/">info:eu-repo/semantics/</a>
License Condition (R)	dc:rights	
Embargo End Date (MA)	dc:date	<a href="http://info:eu-repo/date/embargoEnd/">info:eu-repo/date/embargoEnd/</a>
Alternative Identifier (R)	dc:relation	<a href="http://info:eu-repo/semantics/altIdentifier/">info:eu-repo/semantics/altIdentifier/</a>
Publication Reference (R)	dc:relation	<a href="http://info:eu-repo/semantics/reference/">info:eu-repo/semantics/reference/</a>
Dataset Reference (R)	dc:relation	<a href="http://info:eu-repo/semantics/dataset/">info:eu-repo/semantics/dataset/</a>
Subject (MA)	dc:subject	
Description (MA)	dc:description	
Publisher (MA)	dc:publisher	
Contributor (R)	dc:contributor	
Publication Date (M)	dc:date	
Publication Type (M)	dc:type	<a href="http://info:eu-repo/semantics/">info:eu-repo/semantics/</a>
Publication Version (R)	dc:type	<a href="http://info:eu-repo/semantics/">info:eu-repo/semantics/</a>
Format (R)	dc:format	
Resource Identifier (M)	dc:identifier	
Source (R)	dc:source	
Language (R)	dc:language	
Relation (O)	dc:relation	
Coverage (R)	dc:coverage	
Audience (R)	dc:audience	



# Interoperabilidad – Normalización - Directrices- Antecedentes de las directrices del SNRD

- En 2008 cuando comienza a pensarse el SNRD existía a nivel latinoamericano **BDCOL** (Biblioteca digital colombiana) creó sus propias directrices tomando como base la propuesta europea



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)



# Directrices SNRD 2012

Relación con directrices internacionales:

- Las Directrices SNRD se basan en las “Directrices DRIVER 2.0. Directrices para proveedores de contenido - Exposición de recursos textuales con el protocolo OAI-PMH”.
- También se analizaron las Directrices OpenAIRE 1.12, las directrices de la Biblioteca Digital Colombiana “Modelo de Interoperabilidad para BDCOL” y las “Directrices LUCIS/MODS” propuesta de la Biblioteca Digital de la Universidad Nacional de Cuyo.



# Sistema Nacional de Repositorios Digitales (SNRD)

En este sentido, los repositorios del Sistema Nacional de Ciencia, Tecnología e Innovación (SNCTI) de la República Argentina deben cumplir con las directrices propuestas por el **Sistema Nacional de Repositorios Digitales (SNRD)**, que evalúa su cumplimiento, cosecha los recursos de los repositorios, provee una interfaz de búsqueda centralizada de los recursos recuperados de todos los repositorios que participan y provee estadísticas sobre tamaño, cosechas y tipos de registros recolectados. A su vez, el SNRD comparte los recursos cosechados con la **Red Federada de Repositorios Institucionales de Publicaciones Científicas (LA Referencia)**



# Sistema Nacional de Repositorios Digitales (SNRD)

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 Sistema Nacional de Repositorios Digitales Repositorios Digitales (SNRD) el que se crea por la **Resolución MINCYT N° 469/11**.



# Sistema Nacional de Repositorios Digitales (SNRD)

## 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.



# Sistema Nacional de Repositorios Digitales (SNRD)

## OBJETIVOS

- d) Delinear estrategias dirigidas a garantizar el respeto por los derechos de los autores de los objetos digitales incluidos en los repositorios.
- e) Definir estándares generales para el correcto funcionamiento del Sistema.
- f) Dotar de proyección internacional a la producción científicotecnológica producida en el país a través de su difusión en redes virtuales y su interoperabilidad con repositorios internacionales.

# Sistema Nacional de Repositorios Digitales (SNRD)

## OBJETIVOS

- 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.

## Directrices SNRD

Directrices para proveedores de contenido  
del Sistema Nacional de Repositorios Digitales

Ministerio de Ciencia, Tecnología e Innovación Productiva

**Buenos Aires, junio de 2015**



## Clase 2

### Aspectos tecnológicos



# Aspectos tecnológicos

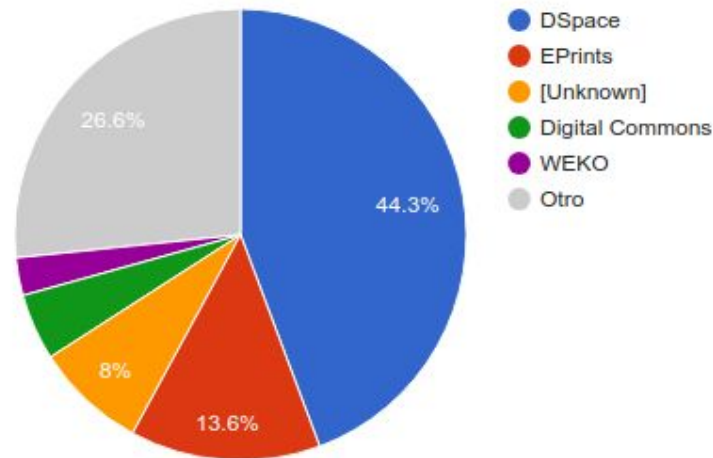
- Aspectos generales sobre software de repositorios
  - servicios
  - licencias
  - tecnologías
  - curva de aprendizaje
  - requerimientos físicos
- Representación y almacenamiento de recursos
  - almacenamiento de datos y metadatos
  - almacenamiento de objetos digitales
  - modelo de datos de DSpace
  - gestión de autoridades
- Ampliación y adecuación del software
  - Extensión
  - Interfaz web. JSP y XMLUI
- Vías de ingesta y circuitos de carga
  - Formularios en DSpace
  - Identificadores persistentes
- Localización de contenidos
- Interoperabilidad:
  - OAI PMH, OpenSearch, SRU/SRW
  - REST, SWORD
- Preservación digital
- Estadísticas



# Software de repositorios de Acceso Abierto (~180)

DSpace	1524 (+120)
EPrints	469 (+32)
DigitalCommons	164 (+16)
OPUS	84 (+3)
CONTENTdm, dLibra, Fedora, Greenstone, Weko	50~60
Drupal, OJS, Mediawiki, OCS, OMP, Wordpress	<50

Usage of Open Access Repository Software - Worldwide



Total = 3440 repositories

OpenDOAR - 18-Sep-2017

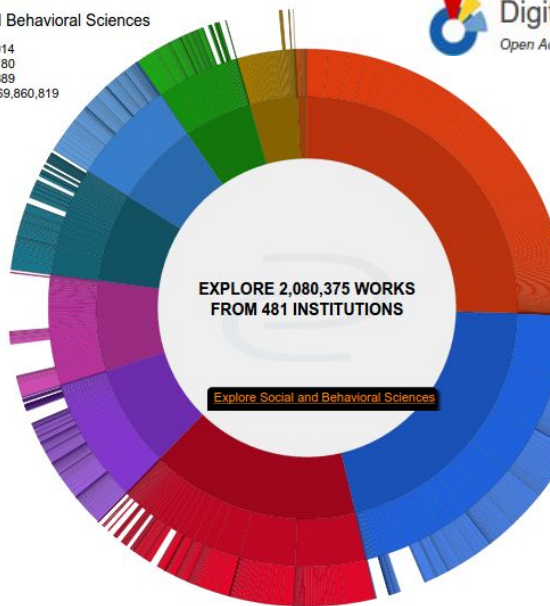


# Software del repositorio - gestión

## Servicio tercerizado o gestión propia?

Social and Behavioral Sciences

Works: 314,014  
Disciplines: 180  
Institutions: 389  
Downloads: 69,860,819



Digital Commons Network™  
Open Access. Powered by Scholars. Published by Universities.®

DSpace 5.x Documentation / DSpace 5.x Documentation

### Installing DSpace

© Creado por Mark Diggory, modificado por última vez por Tim Donohue el ago

- 1 For the Impatient
- 2 Hardware Recommendations
- 3 Prerequisite Software
  - 3.1 UNIX-like OS or Microsoft Windows
  - 3.2 Oracle Java JDK 7 or OpenJDK 7
    - 3.2.1 Elasticsearch requirements for Java
  - 3.3 Apache Maven 3.0.5+ (Java build tool)
    - 3.3.1 Configuring a Proxy
  - 3.4 Apache Ant 1.8 or later (Java build tool)
  - 3.5 Relational Database: (PostgreSQL or Oracle)
  - 3.6 Servlet Engine (Apache Tomcat 7 or later, Jetty, Caucho Resin)
  - 3.7 Perl (only required for [dspace]/bin/dspace-info.pl)
- 4 Installation Instructions
  - 4.1 Overview of Install Options
  - 4.2 Overview of DSpace Directories
  - 4.3 Installation
- 5 Advanced Installation
  - 5.1 'cron' jobs / scheduled tasks
  - 5.2 Multilingual Installation
  - 5.3 DSpace over HTTPS
    - 5.3.1 Enabling the HTTPS support in Tomcat itself (running

**DSpace Direct**  
The hosted repository solution for low cost discovery, access, archiving, and preservation.

OVERVIEW FEATURES SPOTLIGHTS PRICING

**A managed DSpace repository service offered by the DuraSpace not-for-profit organization**  
repository solution for preservation.

Fedora 4.0 Documentation  
© Creado por Andrew Woods, modificado por última vez en nov 15, 2014

- Quick Start
- Feature Tour
- Features
- Administrator Guide
- Developer Guide
- FAQs and How-Tos
- Glossary

**Component Stack**

**Integration Architecture**

# Software del repositorio

- **Lenguaje de programación:**
  - PHP, Perl, Java, Ruby?
    - ¿compilado vs interpretado? actualizaciones, estabilidad
  - ¿servidor web, contenedor web?
  - dependencia con frameworks: Zend , Rails, Cocoon
- **¿Quién es responsable del desarrollo?**
  - Empresa ⇒ Ej. DuraSpace
  - Universidad ⇒ Ej. Universität Stuttgart
  - ONG ⇒ Ej. Public Knowledge Project



# Software del repositorio - Licencia

- Licencia abierta (software libre) o software cerrado?
- no es sólo una cuestión económica
  - Modificaciones, mejoras, personalización
  - Actualizaciones y nuevas versiones
  - Múltiples instalaciones (ej. prod, dev y testing)

## Open System Architecture

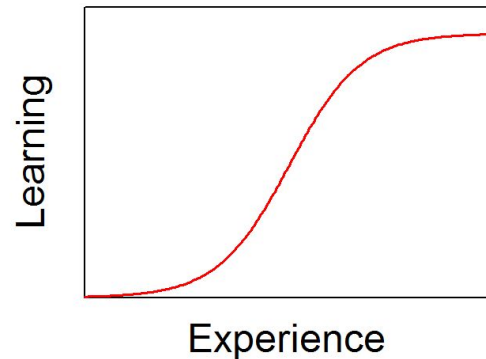
Knowvation™ is built using J2EE to provide an open Knowvation™'s Application Programming Interface and future third party applications. Using the API, it solutions, Workflow and Business Process Management easily performed.



# Curva de aprendizaje

- para usuarios
  - búsqueda / exploración
  - autoarchivo
  - otros servicios (RSS, DSI)
- para administradores
- para desarrolladores y sysadmins

S-Curve (Sigmoid)



Drawn with R<sup>®</sup> using R-studio  
© Alan Fletcher 2013 This file is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported

Cuánta documentación existe?

Tutoriales, cursos, webinars, talleres, congresos

La comunidad de usuarios es muy importante



# Requerimientos físicos

- **Conectividad**
  - usuarios
  - administradores
  - aplicaciones y motores de búsqueda
- **Almacenamiento**
  - espejado de datos
  - centralizado/distribuido
  - control de errores desde el hardware
  - considerar que se almacenará más de una versión por cada OD (original, optimizada, preservable...)



# Requerimientos físicos

- Procesamiento  $\Rightarrow$  memoria y procesador
  - base de datos
  - aplicación
- Algunas recomendaciones
  - separar servicios en diferentes hosts (BD, aplicación, storage...)
  - la virtualización ayuda mucho
  - utilizar hardware de calidad y optimizado
    - Almacenamiento en RAID (por soft o hard?, 0,1,0+1,5,6 ?)
    - Memorias para servers (ECC, DualChannel)
    - Procesadores para servidores (ej. Xeon)
  - planificar el espacio físico: energía, temperatura, seguridad



# Modelo de contenidos

Representación, almacenamiento y gestión de datos  
y metadatos



# Representación de recursos

- La representación interna puede afectar a la catalogación, a la exposición web y a la interoperabilidad
  - formato fijo ⇒ OJS, OCS, OMP
  - formato configurable, ampliable
- Perfil de metadatos basado en esquema propio, estándar o mixto
  - Ej: subconjunto de DC, uso de *application profiles* (esquema de metadatos + restricciones de uso) específicos.
- Vocabularios controlados
- Entidades concretas vs entidades abstractas

# Almacenamiento de datos y metadatos

Qué tecnología de base de datos utiliza?

- BD relacional, XML, NO SQL, RDF

Para empezar, debemos considerar la carga del sistema para procesar las consultas

- desde la web pública
- desde el sistema de administración
- desde otros sistemas (OAI, OpenSearch, SRW)



# Almacenamiento de datos y metadatos

## Control de calidad de los datos

- datos relacionados
- restricciones (desde el esquema): campos obligatorios, tipos de datos
- servicios
  - versionado de datos
  - índices de búsqueda
  - transacciones
  - seguridad, encriptación, control de acceso
  - relaciones semánticas? (RDF)



# Almacenamiento de datos y metadatos

Afecta la elaboración de listados, reportes y consultas estadísticas

EJEMPLOS:

DSpace: PostgreSQL (SQL) + Lucene/SolR --> JAVA

OPUS: MySQL (SQL) + Lucene/SolR --> PHP

Fedora: RDF (SPARQL) + Lucene/SolR --> JAVA

EPrints MySQL (SQL) --> PERL (paquetes rpm y deb)



# Almacenamiento de Objetos Digitales

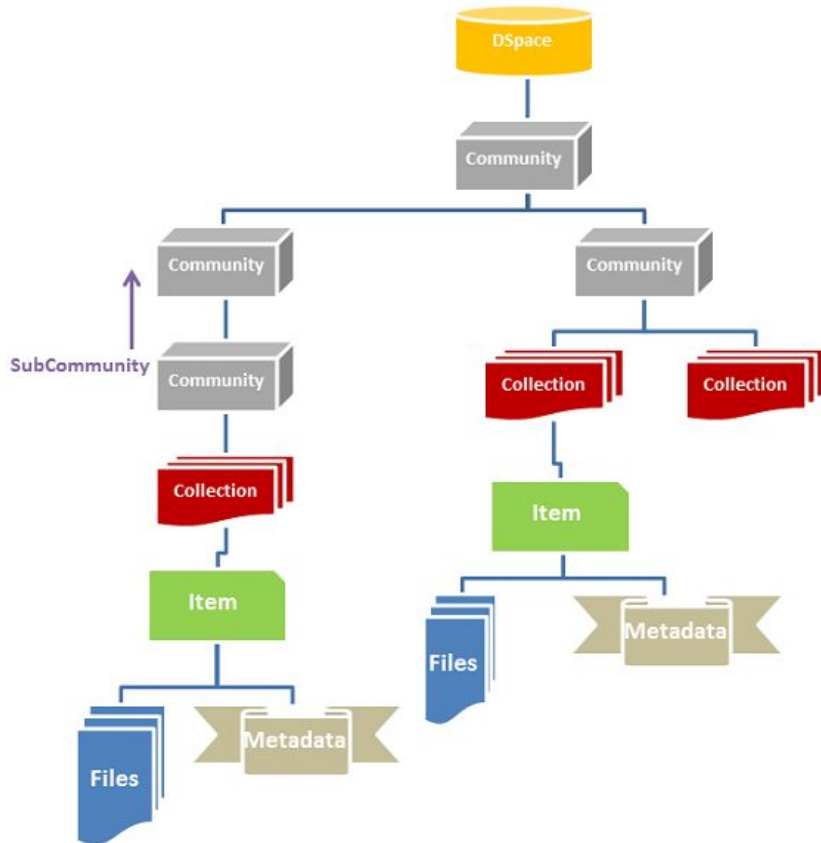
- ¿Filesystem? (archivos y directorios)
- ¿Base de datos? (ej. serialización, GridFS)

Permite operaciones sobre los OD? Checksum, diff, versioning, snapshots

Cómo se hacen los backups? soporta mirroring? almacenamiento distribuido?



# Modelo de contenidos - Estructura típica



1. El repositorio se organiza en una o más **comunidades** de nivel base que se organizan jerárquicamente en subcomunidades.
  - Son como espacios de trabajo
2. Las **colecciones** son los “estantes” dentro de las comunidades, que agrupan contenido relacionado.
3. Los ítems son las obras que van en los estantes y que se pretende que el público encuentre.
4. Los metadatos describen al recurso
5. Los bitstreams son la representación digital del recurso.



# Modelo de contenidos - Comunidad y Colecciones

**CIC**DIGITALRepositorio Institucional  
Comisión de Investigaciones Científicas[Inicio](#) [Explorar ▾](#) [Aportar Material](#) [Mas información ▾](#) [Contacto](#)[Mi cuenta ▾](#) [ES ▾](#)

## Comunidades en DSpace

Elija una comunidad para listar sus colecciones

### ▼ Centros [2053]

Centros de la Comisión de Investigaciones Científicas

#### ▶ [CIDCA](#) [1]

Centro de Investigación y Desarrollo en Criotecnología de Alimentos

#### ▶ [CESGI](#) [1]

Centro de Servicios en Gestión de Información

### ▼ LEMIT [806]

Laboratorio de Entrenamiento Multidisciplinario para la Investigación Tecnológica

Tesis [10]

Artículos y presentaciones en Congresos [20]

Libros [2]

#### ▼ [Revista Ciencia y Tecnología de los Materiales](#) [32]

Número 01 [7]

Número 02 [7]

Número 03 [6]

Número 04 [6]

Número 05 [6]

#### ▶ [Informes de Investigación](#) [34]

#### ▶ [Jornadas de Técnicas de Reparación y Conservación del Patrimonio](#) [419]

#### ▶ [Revista Ciencia y Tecnología del Hormigón](#) [46]

#### ▶ [Anales LEMIT. Serie II](#) [237]

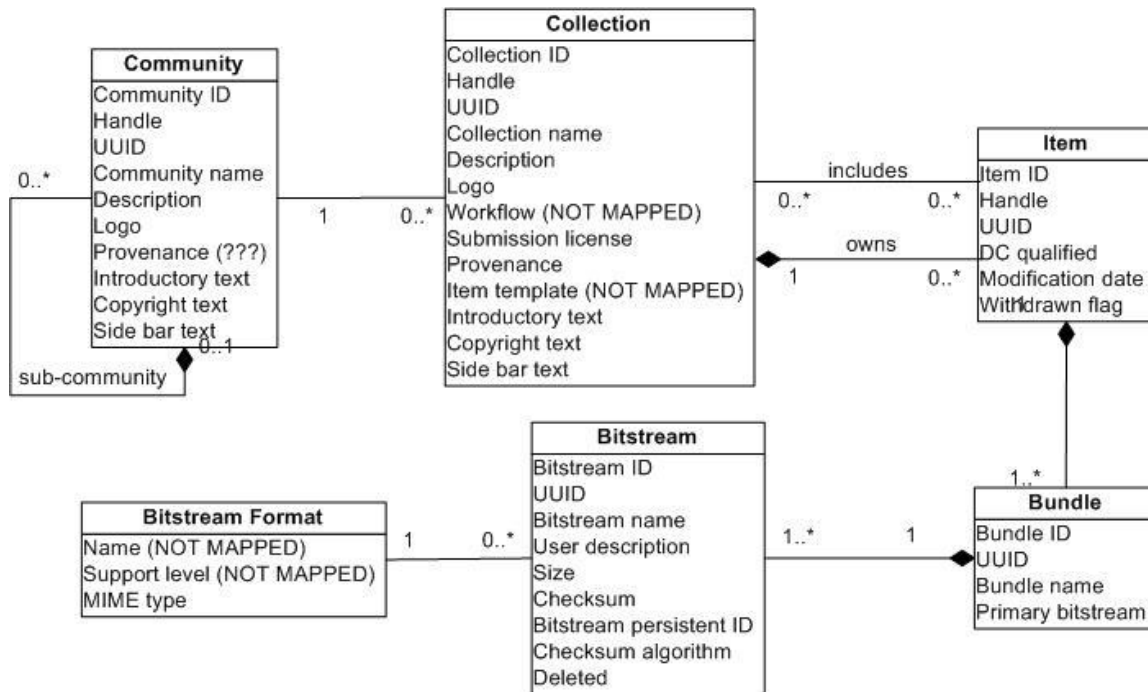
#### ▶ [Anales LEMIT. Serie III](#) [6]

#### ▶ [CIOp](#) [198]

Centro de Investigaciones Ópticas

# DSpace Data Model

- El repositorio se organiza en comunidades y subcomunidades
- Las comunidades poseen colecciones
- Cada colección posee ítems
- Los ítems se componen de bundles
- Cada bundle posee los bitstreams (archivos)



# Modelo de contenidos - Bundles

Los ítems están vinculados a sus bitstreams a través de entidades llamadas Bundles. Los **bundles** agrupan bitstreams bajo determinados criterios:

- **ORIGINAL**: contiene los bitstreams a publicar.
- **THUMBNAILS**: son archivos con thumbnails extraídos a partir de los bitstreams originales (p.e. thumbnails de PDFs, imágenes, etc.).
- **TEXT**: contiene el texto completo (full-text) de otros bitstreams. Se genera a partir de la extracción automática de texto sobre otros bitstreams y se usa durante la indexación para mejorar los resultados de búsqueda.
- **LICENSE**: contiene la licencia que el usuario aceptó al depositar el contenido.
- **CC\_LICENSE**: contiene la [Licencia CC](#) (si es que existe) seleccionada por el usuario durante la carga.



# Modelo de contenidos - Bitstreams

Los ítems del repositorio se componen de **objetos digitales** que representan la obra en sí que se quiere publicar. Los objetos digitales pueden ser:

- Audios
- PDFs
- Documentos de texto (.doc, .odt., etc)
- Diapositivas de presentaciones
- Planillas de cálculo
- Imágenes
- Videos
- etc.

En el lenguaje de DSpace, estos objetos digitales reciben el nombre de **BITSTREAMS**.

The screenshot shows the CIE DIGITAL Institutional Repository interface. The header includes the logo and the text 'Repositorio Institucional Comisión de Investigaciones Científicas'. The navigation menu contains 'Inicio', 'Explorar', 'Aportar Material', 'Mas información', and 'Contacto'. On the right, there are links for 'Mi cuenta', 'Estadísticas', and 'ES'. The main content area displays a document titled 'El Falcon Verde como lugar simbólico y la posibilidad de su activación para un uso turístico responsable' by 'Sahores Avalis, Virginia'. A 'Resumen:' section is visible, followed by a 'Descargas' section with a 'Documento completo' PDF file (5.525Mb) highlighted by a red box. A 'Año 2013' badge is also present.

# Modelo de contenidos - BitStore

Los bitstreams son alojados en un **Bitstore**, que puede estar alojado de forma local o en la nube (según la configuración en [dspace/config/spring/api/bitstore.xml](#)):

- **ASSETSTORE:** si tenemos esta forma de almacenamiento, los bitstreams se alojan en un directorio local en el servidor:

```
{dspace-install-dir}/assetstore/
```

- **S3Store:** si tenemos esta forma de almacenamiento, los bitstreams se alojarán en la nube, más específicamente en el Cloud Storage de [Amazon S3](#). Para que esto funcione debemos configurar diferentes parámetros como, p.e, una *accessKey* al sistema de Amazon.



# Modelo de contenidos - Metadatos

dc.format.extent	116 p.	es
dc.language	es	es
dc.title	Un metalenguaje de programación orientado al diseño de interfaces gráficas	es
dc.type	Tesis	es
sedici.creator.person	Santana, Carlos Ariel	es
sedici.creator.person	Coni, Cintia Vanesa	es
sedici.subject.materias	Ciencias Informáticas	es
sedici.subject.keyword	arquitectura dirigida por modelos	es
sedici.subject.keyword	transformaciones de modelos	es
sedici.subject.keyword	interfaz de usuario	es
sedici.description.fulltext	true	es
mods.originInfo.place	Facultad de Informática	es
sedici.subtype	Tesis de grado	es
sedici.rights.license	Creative Commons Attribution 4.0 International (CC BY 4.0)	
sedici.rights.uri	<a href="http://creativecommons.org/licenses/by/4.0/">http://creativecommons.org/licenses/by/4.0/</a>	
sedici.contributor.director	Pons, Claudia	es
thesis.degree.name	Licenciado en Sistemas	es
thesis.degree.grantor	Facultad de Informática	es

## Los metadatos

- son datos estructurados que describen otros datos
- son datos sobre datos

← Ej: representación del recurso

<http://sedici.unlp.edu.ar/handle/10915/47008>



# Representación - Metadatos



## Esquemas:

- **Dublin Core Element Set (DCES)** es uno de los esquema estandarizados de metadatos más utilizados <http://dublincore.org/documents/dces/>
  - **Título** → dc.title
- **DCMI Terms (dc extrendido o calificado)** es una extensión al esquema básico de DC <http://dublincore.org/documents/dcmi-terms/>
  - **Título alternativo** → dcterms.alternative
- Un repositorio puede utilizar un *perfil de metadatos* basado en esquema propio, estándar y/o mixto
  - Ej: subconjunto de DC, uso de *application profiles* (esquema de metadatos + restricciones de uso) específicos.



# Representación - Metadatos

Ejemplo: SEDICI utiliza un perfil de metadatos propio basado en:

- **DC simple**
  - ◆ dc.title
- **DC calificado**
  - ◆ dc.date.accessioned
- **ETD**
  - ◆ thesis.degree.name
- **MODS**
  - ◆ mods.location
- **Esquema de metadatos SEDICI**
  - ◆ sedici.subtype

dc.date.issued	2009	
dc.identifier.uri	http://hdl.handle.net/10915/1102	
dc.description.abstract	Tomo I: Memoria descriptiva; Informe medioambiental; Trazado; Proyecto de arquitectura la Estación Plaza Moreno. Tomo II: Proyecto estructural de la Estación Plaza Moreno. Tomo III: Plan de trabajos; Evaluación económica	es
dc.format.extent	3 vol. + planos	es
dc.language	es	es
dc.title	Línea "A" de subte - La Plata	es
dc.type	Tesis	es
sedici.creator.person	Liaudat, Joaquín	es
mods.location	http://biblio.ing.unlp.edu.ar/cgi-bin/koha/opac-detail.pl?bib=INGC-MON-017361	es
sedici.subject.materias	Ingeniería	es
sedici.subject.materias	Ingeniería Civil	es
sedici.subject.eurovoc	transporte subterráneo	es
sedici.subject.eurovoc	medio de transporte	es
sedici.description.fulltext	false	es
mods.originInfo.place	Departamento de Construcciones	es
sedici.subtype	Tesis de grado	es
sedici.contributor.director	Venier, Leonardo P.	es
thesis.degree.name	Ingeniero Civil	es
thesis.degree.grantor	Facultad de Ingeniería	es

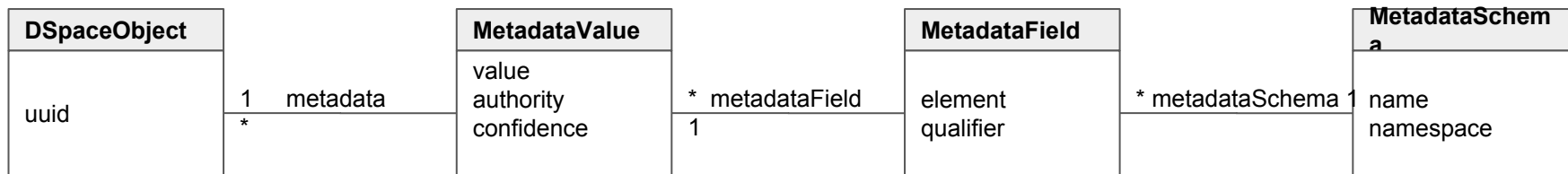


# Modelo de contenidos - Metadatos

DSpace permite definir **perfiles de metadatos** a partir de

1. combinación de elementos de diferentes schemas
2. usa de calificadores:
  - schema.element.qualifier

DC Simple (estándar DCES)	DC simple (perfil en DSpace)
dc.date	dc.date.accessioned dc.date.issued dc.date.created dc.date.available dc.date.submitted



# Personalización del software

## Arquitectura, librerías, lenguajes



# Mejoras y nuevas funciones

¿Para qué se extiende el software?

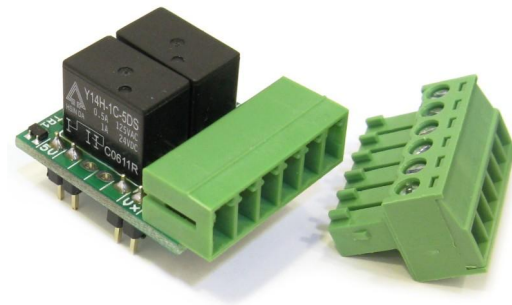
- cambios estéticos
- optimización de procesos / circuitos de trabajo
- detección de errores
- nuevos requerimientos
- integración con otros sistemas (ej. autoridades externas)



# Mejoras y nuevas funciones

¿Cómo se extiende el software del repositorio?

- modelo de clases documentado
- sistema de plugins o módulos
- integración con otras aplicaciones (ej. REST)

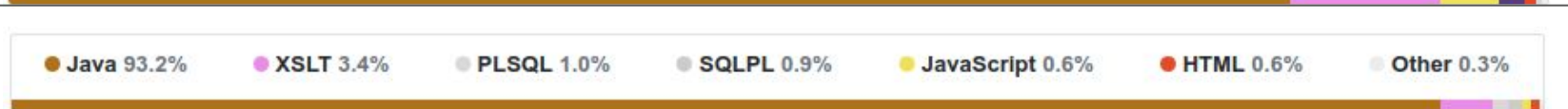


La arquitectura, los lenguajes, las librerías, las herramientas y las licencias que componen el software determinarán la facilidad (y viabilidad) de expender, personalizar o mejorar el software (del repositorio, y cualquier otro software)



# Características generales de DSpace

- Software libre y open-source
- Proyecto escrito mayoritariamente en lenguaje **JAVA**.
- Desarrollado y mantenido por una amplia **comunidad** de usuarios en todo el mundo
- Altamente **configurable**
- Puede ser fácilmente **extendido**
- Dispone de una gran cantidad de funcionalidades
- Posee un **modelo** de datos simple, con metadatos no jerárquicos e independencia de los formatos de archivos.



# Características generales de DSpace - Stack

DSpace es una familia de aplicaciones java, algunas de las cuales funcionan en modo CLI (línea de comandos) mientras otras lo hacen como Aplicación Web JAVA.

- Para la ejecución **cli** se requiere una jre7 o jre8
- Para la ejecución de las **aplicaciones web** requiere un web container como: Jetty, Tomcat, entre otros.

En ambos casos se requiere una base de datos relacional para persistir el modelo. Se puede usar:

- PostgreSQL, Oracle u otras soportadas por Hibernate.



# Características generales de DSpace - Stack

También se utiliza un indexador de texto para

- registros de acceso (estadísticas)
- índice de búsqueda (metadatos y fulltext)
- otros usos.

Solr:

- está optimizado para búsquedas: permite definir relevancia, filtrado, **pesos** en los campos, y más.
- permite gestionar grandes volúmenes de datos



Más información: Solr in DSpace <https://wiki.duraspace.org/display/DSPACE/Solr>



# Características generales de DSpace - Dependencias

**Spring es un** framework que implementa el modelo de **Dependency Injection** (también llamado IoC) sobre objetos POJO. Se compone de

- un bean container que administra el ciclo de vida de beans o POJO: spring-core.
- módulos específicos para
  - aplicaciones web mvc: [spring mvc](#)
  - autenticación y autorización: [spring-security](#)
  - entre otros.



DSpace utiliza Spring para administrar:

- Los **servicios** de la Capa de la Lógica de Negocios ([core-services.xml](#))
  - *MetadataValueService, ItemService, CollectionService, etc.*
- Las **extensiones** al core de DSpace ( configuraciones en [spring/api](#), mayormente)
  - *discovery.xml, workflow-actions.xml, bitstore.xml, etc.*



# Características generales de DSpace - Dependencias

## Ejemplo de archivo de beans de spring

Archivo `discovery.xml`. donde se declaran los beans que se utilizarán en el módulo.

```

<!--Bean that is used for mapping communities/collections to certain discovery
configurations.-->
<bean id="org.dspace.discovery.configuration.DiscoveryConfigurationService"
      class="org.dspace.discovery.configuration.DiscoveryConfigurationService">
  <property name="map">
    <map>
      <!--The default entry, DO NOT REMOVE the system requires this-->
      <entry key="default" value-ref="defaultConfiguration" />
    </map>
  </property>
</bean>

<!-- ... .. -->

<!--The default configuration settings for discovery-->
<bean id="defaultConfiguration"
      class="org.dspace.discovery.configuration.DiscoveryConfiguration"
      scope="prototype">
  <!--Which sidebar facets are to be displayed-->
  <property name="sidebarFacets">
    <list>
      <ref bean="searchFilterType" />
    </list>
  </property>
</bean>

```

## Clase DiscoveryConfigurationService

```

public class DiscoveryConfigurationService {
    private Map<String, DiscoveryConfiguration> map;
    private Map<Integer, List<String>> toIgnoreMetadataFields = new HashMap<>();

    public Map<String, DiscoveryConfiguration> getMap() {
        return map;
    }

    public void setMap(Map<String, DiscoveryConfiguration> map) {
        this.map = map;
    }
    ///...
}

```

## Clase DiscoveryConfiguration

```

public class DiscoveryConfiguration implements InitializingBean{
    /** The configuration for the sidebar facets */
    private List<DiscoverySearchFilterFacet> sidebarFacets = new
    ArrayList<DiscoverySearchFilterFacet>();

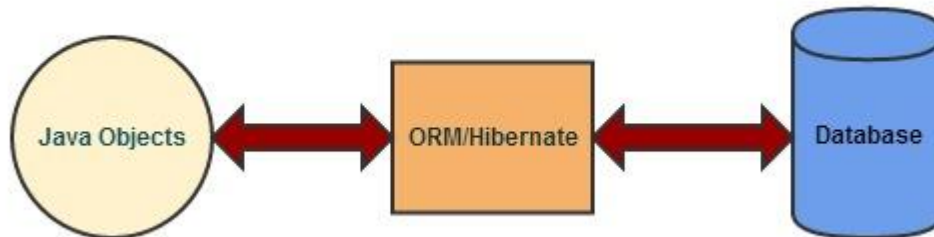
    public List<DiscoverySearchFilterFacet> getSidebarFacets() {
        return sidebarFacets;
    }

    @Required
    public void setSidebarFacets(List<DiscoverySearchFilterFacet> sidebarFacets) {
        this.sidebarFacets = sidebarFacets;
    }
    /// ...
}

```

# Características generales de DSpace - Dependencias

**Hibernate** ORM es un framework para persistencia a partir del mapeo Objeto a Relaciones.



- permite **abstraer** a la aplicación de la BD utilizada
- **mapea automáticamente** los datos de objetos al momento de leer de/persistir en la BD
- permite usar consultas abstractas independientes de la BD usando lenguaje **HQL**
- puede utilizarse en combinación con otras librerías como *HibernateValidation*
- además provee:
  - caching de objetos,
  - transacciones de alto nivel
  - entre otros.

A partir de DSpace 6.x, se comenzó a utilizar **Hibernate** como parte de la capa de almacenamiento.



# Interfaz web

La interfaz web es la cara visible del repositorio ante el mundo

- Debe ser simple para el usuario final
- Debe ser intuitiva (nadie debería necesitar un tutorial)
- Debe ser internacionalizable
- Sería interesante que provea
  - accesibilidad
  - *responsiveness*



# Interfaz web

El software que se utiliza debe permitir la adecuación de la interfaz web según la marca institucional

## DSpace Repository

DSpace is a digital service that collects, preserves, and distributes digital material. Repositories are important tools for preserving an organization's legacy; they facilitate digital preservation and scholarly communication.

## Comunidades en DSpace

Elija una comunidad para listar sus colecciones

- [Atoll Research Bulletin](#) [672]
- [Center for Folklife and Cultural Heritage](#) [52]
- [CONTENT STAGING](#) [1443]
- [Cooper-Hewitt National Design Museum](#) [1]

Bem-vindo ao Repositório Aberto

O Repositório Aberto colecciona, preserva e disponibiliza

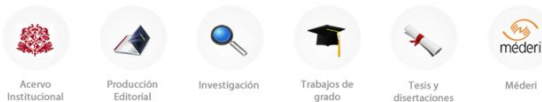
# Interfaz web

No es sólo una cuestión estética, es también fundamental desde el punto de vista **funcional**

- Define la navegación por el repositorio
- Permite el acceso a las funciones (búsqueda, filtrado, autoarchivo, novedades)
- Priorización de contenidos o colecciones

Buscar en EdocUR

Introduzca el texto a buscar en EdocUR

Intellectum es el Repositorio Institucional de la Universidad de La Sabana, se ha creado para administrar, preservar y difundir la producción intelectual, científica, cultural e histórica de la comunidad universitaria.



## Libros

Libros digitalizados y e-books



## Red UNCI

Artículos y ponencias de la Red UNCI



## Radio Universidad

Entrevistas y producciones artísticas, entre otros audios



## Recursos educativos abiertos

Material educativo de cursos dictados en la UNLP



## Red de museos

Material perteneciente a la red de museos de la UNLP



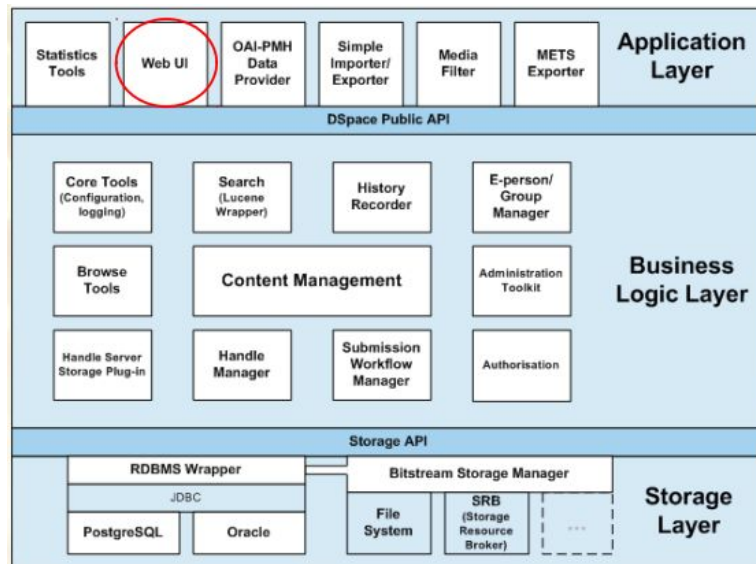
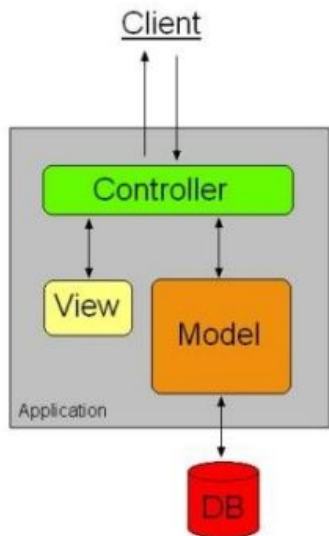
## Emergencia hídrica

Trabajos dentro del Plan de Gestión Integrada de Riesgos de Desastres



# Interfaz web

El software debe proveer, desde su diseño, una separación entre los datos, la lógica de negocios y la presentación



Arquitectura, capas y componentes de DSpace. Fuente <https://wiki.duraspace.org/>

# Web UI en DSpace

DSpace soporta dos tipos de interfaces: **JSPUI** y **XMLUI**

La interfaz JSPUI está implementada usando Java Servlets para gestionar la lógica de negocios, y JavaServer Pages (JSP) para producir el código HTML.

Los documentos HTML se construyen ensamblando "porciones" generadas por archivos JSP, organizados en un layout común a todas las páginas.



# Web UI en DSpace - JSP layout


(Header)

**Search DSpace**

Advanced Search  
Subject Search

Home

**Browse**

- Communities & Collections
- Titles
- Authors
- Subjects
- By Date

**Sign on to:**

- Receive email updates
- My DSpace authorized users
- Edit Profile
- Help
- About DSpace

DSpace at My University > (Location Bar)

**DSpace is Live**

Welcome to our digital repository of My University research!  
More exciting news to appear here.

**Search**

Enter some text in the box below to search DSpace.

This is a default installation of DSpace

**Search DSpace**

Advanced Search  
Subject Search

Home

**Browse**

- Communities & Collections
- Titles
- Authors
- Subjects
- By Date

**Sign on to:**

- Receive email updates
- My DSpace authorized users
- Edit Profile
- Help
- About DSpace

DSpace at My University >

**Browse by Title**

Jump to:  or enter first few letters:

Showing items 1-21 of 224.

Issue Date	Title	Author(s)
1996	<a href="#">The 1995 review and extension of the Nuclear Non-Proliferation Treaty</a>	edited by Thomas Morgan, Jason Zych and Clifford E. Singer
1998	<a href="#">Aadhi raat ke baad : "After midnight"</a>	S. Rashid Naim
2005	<a href="#">Anatomy of interlinking rivers in India : a decision in doubt</a>	A.C. Shukla and Vandana Asthana; Asthana, Vandana
Feb-2004	<a href="#">Asymptotic theory of ignition and failure of self-sustained detonations</a>	-
Jun-2001	<a href="#">Atomistic scale experimental observations and micromechanical/continuum models for the effect of hydrogen on the mechanical behavior of metals</a>	-

(Browse Layout)

W3C XHTML 1.0

DSpace Software Copyright © 2002-2007 MIT and Hewlett-Packard - Feedback



# Web UI en DSpace - XMLUI

Utiliza el framework Apache Cocoon

- múltiples componentes "simples"
- procesamiento en *pipeline*

Modularidad, extensibilidad y una organización multi-tier

- Capa de estilos (style tier)
- Capa de temas (theme tier)
- Capa de aspectos (aspect tier)



# Web UI en DSpace - XMLUI

Los distintos *tiers* permiten distintos grados de personalización

style	CSS y XHTML. Cambios menores sobre temas preexistentes
theme	XSLT, CSS y XHTML. Ajustes avanzados y nuevos temas
aspect	Cocoon, Java, XSLT. Definición de nuevas características, incorporación de contenido a los DRI (Digital Repository Interface)

Basado en archivos de configuración

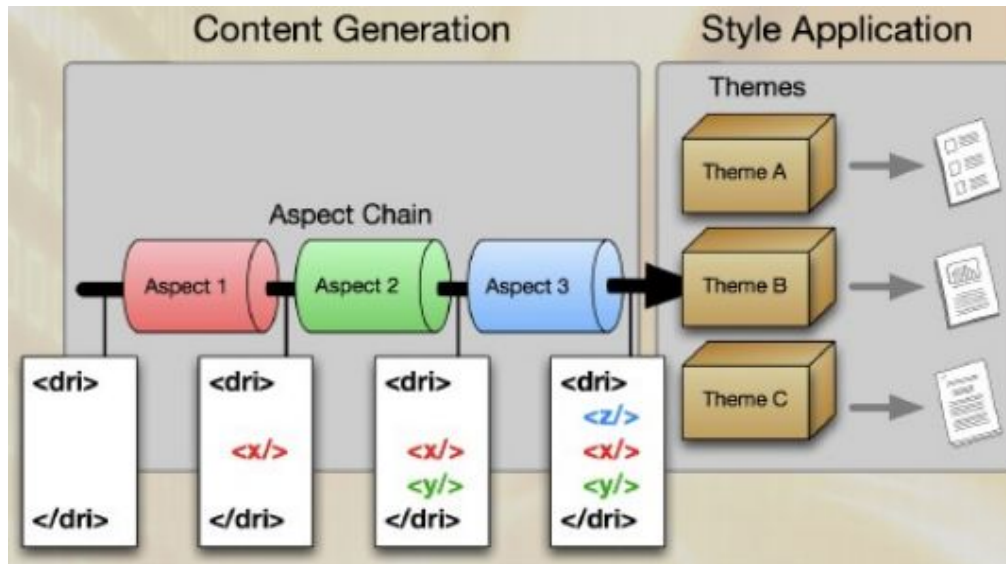
Permite personalizar ítems, colecciones y comunidades



# Web UI en DSpace - XMLUI - Aspect chain

## Aspectos del core de Manakin

- ViewArtifacts
- BrowseArtifacts
- SearchArtifacts
- E-Person
- Administrative
- Submission
- Statistics
- Workflow



Manakin request processing. Fuente: Phillips, S.; Green, C.; Maslov, A.; Mikeal, A. Leggett, J. "Manakin. A New Face for DSpace" D-Lib Magazine, November/December 2007. Volume 13 Number 11/12



# Gestión de autoridades

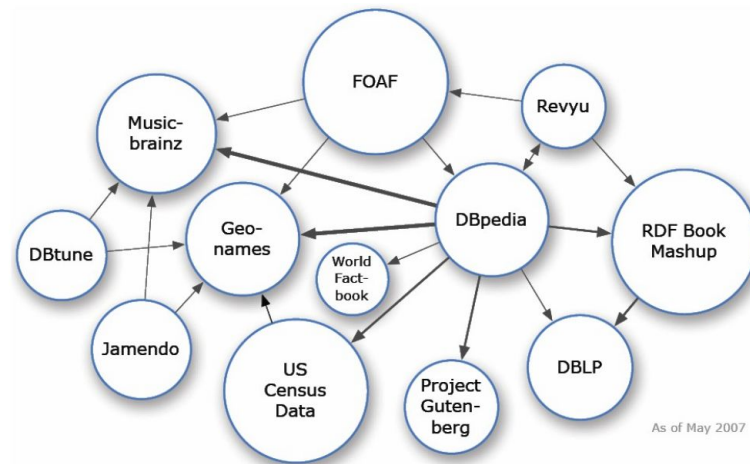
Vocabularios controlados e integraciones con  
DSpace



# Catalogación de recursos - gestión de autoridades

Uso de Vocabularios controlados gestionados en otros sistemas

- Tesoros
- Sistemas de clasificación / Taxonomías
- Lista de encabezamientos de materias
- Bases de datos de investigadores
- Jerarquía de instituciones
- Grados alcanzados



El software del repositorio debe integrarse con estos sistemas.

En DSpace estos vocabularios controlados son llamados **autoridades**.



# Catalogación de recursos - gestión de autoridades

## Ventajas:

1. Permite probar que dos valores son idénticos comparando por la clave de las autoridades.
2. Ayuda a completar metadatos con valores correctos.
3. Permite mejorar la calidad de los metadatos.
4. Mejora la interoperabilidad compartiendo un nombre de autoridades con otra aplicación.
5. Reduce el tiempo de carga



# Catalogación de recursos - gestión de autoridades

La elección del valor de un metadato pueden ser abiertas o cerradas.

- **Abiertas :**
  - Se pueden usar valores no incluidos en las elecciones presentadas.
- **Cerradas:**
  - Elección restringida solo para un conjunto de valores ofrecidos.

La elección del valor de un metadato pueden ser obligatorio u opcional.



# Catalogación de recursos - gestión de autoridades

DSpace nos da la posibilidad de utilizar autoridades.

- **Autoridad:** conjunto de valores fijos identificado por una clave.
  - **Registro de autoridad:** información asociada con uno de los valores de la autoridad
  - **Clave de autoridad:** un identificador persistente que se corresponde con el registro de autoridad

La ubicación de las autoridades pueden ser externas a DSpace.

Facilitar la integración de nuevas autoridades sin modificar el código en DSpace.



# Catalogación de recursos - gestión de autoridades

Las autoridades controladas utilizadas en **CIC-DIGITAL** se encuentra en una instalación de Drupal.

Drupal:

- Cuenta con un conjunto de **módulos** que facilitan el modelado, la importación de los contenidos con sus relaciones y su indexación, gracias a su soporte de nodos y taxonomías.
- La comunicación entre esta plataforma y CIC-DIGITAL es mediante consultas **SPARQL**, a través de un endpoint configurado en la instalación de Drupal el cual permite realizar las búsquedas.

SEDICI utiliza un software propio de gestión de autoridades



# Catalogación de recursos - gestión de autoridades

## Choice management:

Mecanismo para selección o elección de los posibles valores de un metadato, a partir de valores propuestos o consultas.

Ejemplo: Submission

### Autor Institucional:

Institución creadora del contenido intelectual de la obra

uni

Add

Universidad de Salamanca (USAL)

AGH University of Science and Technology (AGH )

Benemérita Universidad Autónoma de Puebla (BUAP)

Cadi Ayyad University (UCA)

Centro de Investigación en Economía Política y Comunicación (CIEPYC)

Editorial de la Universidad Nacional de La Plata (EduLP)

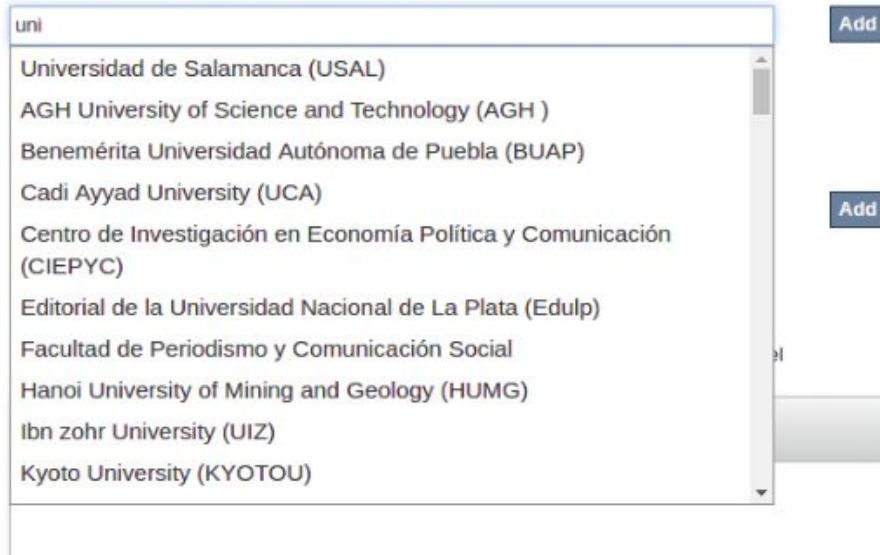
Facultad de Periodismo y Comunicación Social

Hanoi University of Mining and Geology (HUMG)

Ibn zohr University (UIZ)

Kyoto University (KYOTOU)

Add



# Catalogación de recursos - gestión de autoridades

## Indexación de autoridades

DSpace permite indexar las autoridades para crear facets de búsqueda que permiten al usuario filtrar resultados

Materia
Ciencias Físicas (84)
Óptica, Acústica (40)
Física (23)
Ingeniería de los Materiales (12)
Ciencias Químicas (9)
Diseño Arquitectónico (8)
Ingeniería de los Materiales (7)
Física Atómica, Molecular y Química (3)
Ingeniería Eléctrica y Electrónica (3)
Físico-Química, Ciencia de los Polímeros, Electroquímica (2)
... ver más



# Catalogación de recursos - gestión de autoridades

## Control de Autoridades en DSpace - API

Para implementar el control de autoridades DSpace ofrece un **API** con las siguientes clases e interfaces:

- **Choice:**
  - Clase que contiene los atributos authority, label, confidence y value.
- **Choices:**
  - Clase que contiene un conjunto de Choice.
- **ChoiceAuthority:**
  - Interfaz para suplir el mecanismo del control de autoridades.

# Catalogación de recursos - gestión de autoridades

## Control de Autoridades en DSpace - Niveles de confianza

Representa la “*calidad*” o confianza de un valor de autoridad.

1. **ACCEPTED** - Código 600.
2. **UNCERTAIN** - Código 500.
3. **AMBIGUOUS** - Código 300.
4. **FAIL** - Código 200.
5. **REJECTED** - Código 100.
6. **NOVALUE** - Código 0.
7. **UNSET** - Código -1.

# Catalogación de recursos - gestión de autoridades

## Control de Autoridades en DSpace - Base de datos

1. No es un reemplazo del valor de los metadatos.
2. Es configurado mediante un campo en la base de datos.

text_value	authority	Confidence
Universidad Nacional de La Plata (UNLP)	<a href="http://digital.cic.gba.gob.ar/auth/node/86555">http://digital.cic.gba.gob.ar/auth/node/86555</a>	600
Cañueto, Matías F.	<a href="http://digital.cic.gba.gob.ar/auth/node/204702">http://digital.cic.gba.gob.ar/auth/node/204702</a>	600
Attribution 4.0 International (BY 4.0)	<a href="http://creativecommons.org/licenses/by/4.0/">http://creativecommons.org/licenses/by/4.0/</a>	600
UNLP	<a href="http://digital.cic.gba.gob.ar/auth/node/86555">http://digital.cic.gba.gob.ar/auth/node/86555</a>	500
Cañueto, Matías G.		-1

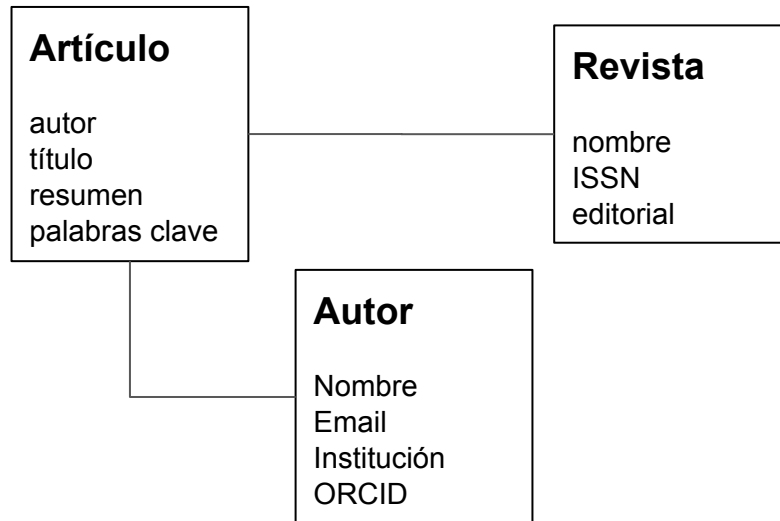


# Repositorio Semántico

Existen relaciones entre las entidades

- un artículo **pertenece** a una revista
- un artículo **tiene** autores

Algunas relaciones se manejan desde los metadatos, otras por medio de entidades externas



# Repositorio Semántico

Las relaciones establecen un modelo semántico

- Podemos especificar este modelo de manera formal, mediante una ontología
  - Método formal para describir taxonomías y redes de clasificación (RDF: Resource Description Framework)
  - Define la estructura del conocimiento para diferentes dominios
- Podemos representar las relaciones semánticas en una base de datos semántica
- Consultas en un lenguaje de consulta RDF como por ejemplo SPARQL

Mayor complejidad, pero ganamos flexibilidad para representar entidades y relaciones

# Vías de ingesta

Circuitos de carga y gestión de formularios



# Vías de ingesta

Cómo ingresan los recursos al repositorio?

Es posible personalizar las vías de ingesta e incorporar nuevas vías?

- Autoarchivo ⇒ asistente sencillo para los usuarios
- Importación ⇒ incorporación masiva de ítems en la cola de procesamiento
- Carga desde la administración
- SWORD ⇒ incorporación de ítems desde otros sistemas
- Otros (ej OAI client)



# Circuitos de carga

Cada vía de ingesta puede tener su propio circuito de trabajo

- distintos formularios de catalogación
  - por colección
  - por tipo de documento
- workflows de revisión y publicación particulares
- plantillas con datos pre cargados
- widgets HTML personalizables

Qué facilidades brinda el software del repositorio para permitir (y promover) esta variedad?



# Web UI en DSpace - Formularios

Distintos conjuntos de metadatos para distintos tipos de contenidos

- ítems de congresos: nombre de la conferencia, lugar, fecha de exposición
- tesis: grado alcanzado, director, subdirector
- libro: capítulos, ISBN, editorial
- artículo de revista: datos de la revista (nombre, ISSN), número, volumen, año

No es viable (ni aceptable) un gran formulario con todos los posibles campos para todos los posibles tipos de contenidos)

La interfaz de carga debe adaptarse al contenido

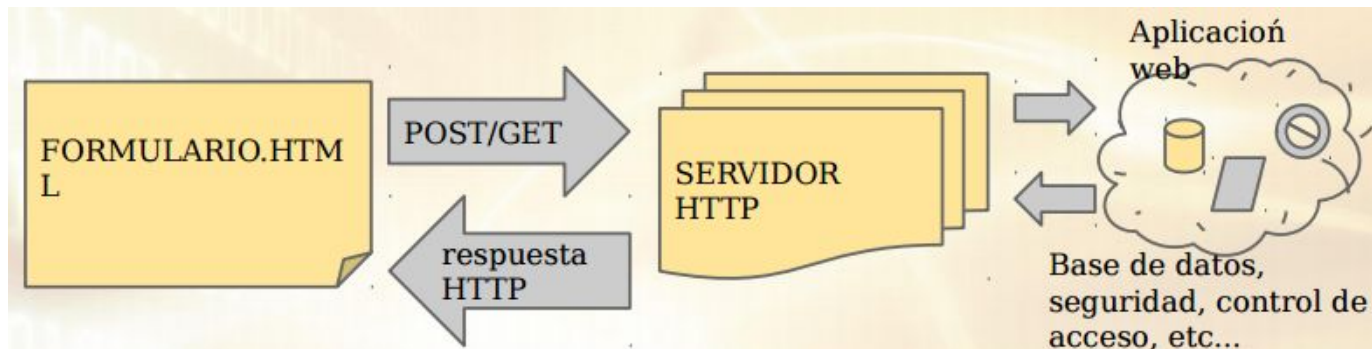


# Web UI en DSpace - Formularios

Distintos widgets (input fields) -> texto, selección, párrafo

Distintos dominios: números, fechas, si/no (boolean)

Distintas validaciones: campos obligatorios, selección de al menos x ítems



# Web UI en DSpace - Formularios

Cada formulario tiene varias etapas

Para cada etapa se presentan varios campos

The image shows a screenshot of the 'Envío de ítems' (Item Submission) form in DSpace. The form is divided into several sections, each with a corresponding UI component identified by a red box and an arrow:

- Submission progress bar:** Located at the top of the form, showing the current step in the submission process: **Describir** (highlighted) → Describir → Adjuntar → Revisar → Licencia CC → Licencia → Finalizar.
- Selector simple (select):** A dropdown menu for 'Tipo de documento (\*)' with the text 'Seleccione el Tipo de Documento que desea cargar' and the option '- Seleccione un Tipo -'.
- Texto de una línea (text):** A single-line text input field for 'Autor (\*)' with the label 'Autores de la obra' and an 'Agregar Otro' button.
- Texto multi línea (textarea):** A multi-line text area for 'Título (\*)' with the label 'El título principal de la obra'.
- Calendario:** A date selection field for 'Fecha de Publicación:' with the label '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/año.'

# Web UI en DSpace - Formularios

DSpace permite mantener distintos flujos de envíos (submission-processes)

Cada submission-process se compone de pasos o steps

El step Describe define los metadatos del ítem

```
<submission-process name="autoarchiveSubmission">
  <step>
    <processing-
class>org.dspace.submit.step.SkipInitialQuestionsStep</process
ing-class>
    </step>
    <step id="sediciDescribeItem"/>
    <step id="sediciUploadItem"/>
    <step id="sediciVerifyItem"/>
    <step id="sediciCCLicense"/>
    <step id="sediciSignOffLicense"/>
    <step id="sediciWorkflowEditedControl"/> <!-- setea el
flag indicando que se completó el proceso de edición desde el
workflow. Sin interface. -->
  </submission-process>
```



# Web UI en DSpace - Formularios

Los *steps* se definen de manera independiente de los *submission processes*

Esto permite la reutilización de *submission steps*

```
<step id="sediciCCLicense">
  <heading>submit.progressbar.CClicense</heading>
  <processing-
class>org.dspace.submit.step.SediciCCLicenseStep</pro
cessing-class>
  <jspui-
binding>org.dspace.app.webui.submit.step.JSPCCLicense
Step</jspui-binding>   <xmlui-
binding>org.dspace.app.xmlui.aspect.submission.SediciC
CLicenseStep</xmlui-binding>
    <workflow-editable>true</workflow-editable>
  </step>
```



# Web UI en DSpace - Formularios

DSpace trae una configuración básica (config/input-forms.xml) que implementa los formularios tradicionales para ingreso de metadatos

Es posible definir nuestros propios conjuntos de formularios personalizados

En DSpace, un **form** es un **conjunto de páginas** a través de las cuales los usuarios ingresan sus metadatos

- Un **form de DSpace** se compone de **múltiples forms HTML**
- Dentro de la estructura del XML, cada Dspace Form se descompone en una serie de páginas, y cada una de ellas representa una página web separada para recolectar metadatos



# Web UI en DSpace - Formularios



Estructura simplificada del archivo input-forms



Es posible especificar las propiedades de cada campo individual de cada form

Algunos campos sólo se muestran en ciertos tipos (dc.type) de contenidos (**type-bind**)

```
<form name="traditional">
  <page number="1">
    <!-- Tipo de documento: determina la carga de otros metadatos -->
    <field>
      <dc-schema>dc</dc-schema>
      <dc-element>type</dc-element>
      <label>Tipo de documento (*)</label>
      <input-type value-pairs-name="common_types">dropdown</input-type>
      <hint>Seleccione el Tipo de Documento que desea cargar</hint>
      <required>Debe seleccionar un tipo de documento, ya que el proceso de
    </field>
    <field>
      <dc-schema>sedici</dc-schema>
      <dc-element>subtype</dc-element>
      <label>Especifique el tipo de Artículo (*)</label>
      <input-type value-pairs-name="common_subtypes_articulo">dropdown</input-type>
      <hint>Seleccione el tipo de artículo que desea cargar</hint>
      <required>Debe seleccionar un subtipo de documento</required>
      <visibility>workflow</visibility>
      <type-bind>Articulo</type-bind>
    </field>
    <field>
      <dc-schema>sedici</dc-schema>
      <dc-element>subtype</dc-element>
      <label>Especifique el tipo de Tesis (*)</label>
      <input-type value-pairs-name="common_subtypes_tesis">dropdown</input-type>
      <hint>Seleccione el tipo de tesis que desea cargar</hint>
      <required>Debe seleccionar un subtipo de documento</required>
      <visibility>workflow</visibility>
      <type-bind>Tesis</type-bind>
    </field>
  </page>
</form>
```

## Localización de recursos

Identificadores persistentes. Exploración y búsqueda.



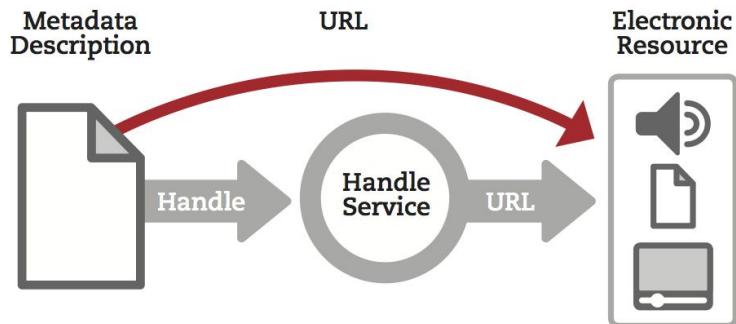
# Identificadores persistentes

Un IP asigna URLs **fijas** a recursos en línea, con el objetivo de garantizar su acceso incluso si estos cambian su URL de acceso

[sedici.unlp.edu.ar/handle/10915/54681](http://sedici.unlp.edu.ar/handle/10915/54681) ⇒ [hdl.handle.net/10915/54681](http://hdl.handle.net/10915/54681)

[revistas.ucr.ac.cr/index.php/eciencias/article/view/23690](http://revistas.ucr.ac.cr/index.php/eciencias/article/view/23690) ⇒ [dx.doi.org/10.15517/eci.v6i2.23690](http://dx.doi.org/10.15517/eci.v6i2.23690)

<http://revistas.unlp.edu.ar/raab/article/view/2177> ⇒ <http://dx.doi.org/10.17139/raab.2016.0018.02.09>



Fuente de la imagen <https://www.clarin.eu/content/persistent-identifiers>



# Identificadores persistentes

Por qué son necesarios?

- Las URLs pueden cambiar
  - en el dominio
  - en la ruta
- Cambios en el software
- Cambios en políticas institucionales

**10045/13546**  
↑ prefijo ↑ sufijo ↑

**Handle**

**10.4100/jhse.2010.52.15**  
↑ prefijo ↑ sufijo ↑  
**DOI**

http://purl.org/net/example/myFirstPURL  
Scheme host domain PURL name  
PURL id

**PURL**

Los cambios de URLs deben notificarse al **manejador de redirecciones**, para que éste actualice sus reglas de redireccionamiento



# Identificadores persistentes

Los recursos pueden tener más de un IP:

- al momento de la publicación (ej. artículo de revista)
- al momento del depósito en el repositorio institucional

## Información general

Fecha de publicación: diciembre 2016

Idioma del documento: Español

Revista: Revista Argentina de Antropología Biológica; vol. 18, no. 2

Dossier: Los estudios de histología ósea en Antropología Biológica

Institución de origen: Asociación de Antropología Biológica de la República Argentina (AABRA)

DOI: 10.17139/raab.2016.0018.02.01

ISSN: 1853-6387

Referencias geográficas: Argentina

Extensión: 3 p.

Materias: Antropología

Descriptores: Técnicas Histológicas; paleohistología; tejido óseo

## Descargar archivos



Documento completo  
Descargar archivo (428.7Kb) - PDF



Enlace externo  
[www.revistas.unlp.edu.ar/...](http://www.revistas.unlp.edu.ar/)

Creado el: 3 de junio de 2016

Disponible en SEDICI desde: 3 de junio de 2016

Por favor, utilice este identificador (URI) para citar o enlazar este ítem: <http://hdl.handle.net/10915/53190>

# Identificadores persistentes

Los identificadores persistentes pueden aplicarse:

- documentos de texto: artículos, tesis, libros
- datos de investigación: bases de datos, imágenes, audios
- personas

Dónde se usan?

- Catalogación y publicación de recursos
- Citas bibliográficas
- Menciones en línea (blogs, redes sociales, etc.)



**Gonzalo L. Villarreal**

**ORCID ID**  
id.orcid.org/0000-0002-3602-8211  
Ver versión pública

**Scopus**

Search Alerts

Back to results | 1 of 1

Villarreal, Gonzalo Luján  
Universidad Nacional de La Plata, Servicio de Difusión de la Creación Intelectual, La Plata, Argentina  
Author ID: 25628901000  
id http://orcid.org/0000-0002-3602-8211

# Servicios para localizar contenidos

El repositorio debe brindar mecanismos para ayudar a los usuarios a acceder a los contenidos que aloja

- exploración
- búsqueda
  - ◆ por metadatos
  - ◆ por texto completo
- faceting



# Exploración

## Acceso a los contenidos

- a partir de su organización lógica (comunidades y colecciones)
- a partir de su presentación en la web

Navegue por nuestras colecciones



**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



### Red de Universidades con Carreras en Informática (RedUNCI)

La Red de Universidades con Carreras en Informática (RedUNCI) coordina actividades académicas relacionadas con el perfeccionamiento docente, la actualización curricular, la utilización de recursos compartidos, y establece un marco de colaboración para el desarrollo de las actividades de postgrado en las carreras Ciencias de la Computación y/o Informática.

[Acceder al sitio web](#)

**Explorar:**  
Listar todo   Autores   Temas

---

**Subcomunidades en esta comunidad**  
Eventos

---

**Colecciones en esta comunidad**  
Libros, actas de congresos y otros documentos



### CACIC 2015

**Nombre del evento:** XXI Congreso Argentino de Ciencias de la Computación  
**Fecha:** 5 al 9 de octubre de 2015  
**Lugar:** Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA); Sede Junín  
**Comité organizador:** Claudia Russo; Tamara Ahmad; Natalia Anolles; Paula Lencina; Leonardo Mangold; Daniela Pérez; María Linda Trinidad Picco; Sabina Rodríguez; Eliana Serrano; Oscar Spada; Carlos Di Cicco; Mónica Sarobe  
**ISBN:** 978-987-3724-37-4  
**Materias:** Ciencias Informáticas

[Acceder al sitio web](#)

**Explorar:** Listar todo   Autores   Temas

# Búsqueda

La búsqueda puede realizarse a partir de los metadatos de los objetos digitales

- por autor
- por título
- por área temática

Una buena catalogación permite ofrecer un servicio de búsqueda de buena calidad



# Búsqueda por faceting

Un facet es un **filtro** generado dinámicamente a partir de una búsqueda o exploración previa

Permite refinar los resultados a partir de múltiples criterios

Muy útiles para localizar contenidos en grandes conjuntos de datos

**Refine su búsqueda**

Tipo de documento

- Artículo (9)
- Clase (2)
- Documento de trabajo (6)
- Objeto de conferencia (67)
- Preprint (6)
- Reporte (5)
- Tesis de doctorado (3)
- Tesis de grado (2)

Fecha de publicación

- 2010 - 2016 (100)**
- 2016 (5)
- 2015 (26)
- 2014 (22)
- 2013 (22)
- 2012 (14)
- 2011 (7)
- 2010 (4)

Materia

- Ciencias Informaticas (89)
- Bibliotecologia (77)
- Educacion (10)
- Informatica (4)
- Ciencias de la Educacion (1)

Unidades académicas → Presidencia →

## PREBI-SEDICI

Resultados de su búsqueda...

**Objeto de conferencia** Curso de Posgrado dictado en la Facultad de Informática  
 Curso de posgrado: Bibliotecas y repositorios digitales. Tecnología  
 De Giusti, Marisa Raquel; Servicio de Difusión de la Creación Intelectual (SE

**Objeto de conferencia** Semana del Acceso Abierto (Montevideo, Uruguay, 2013)  
 Jornada de sensibilización: Hacia la construcción de un repositori  
 De Giusti, Marisa Raquel

**Objeto de conferencia** III Conferencia de Bibliotecas y Repositorios Digitales de A  
 Infraestructura interoperable alrededor del repositorio institucional  
 De Giusti, Marisa Raquel; Lira, Ariel Jorge; Villarreal, Gonzalo Luján; Terruzz

**Objeto de conferencia** III Conferencia de Bibliotecas y Repositorios Digitales de  
 Discurso de inauguración del BIREDIAL 2013

# Búsqueda por texto completo

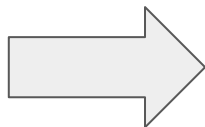
Permite localizar objetos digitales más allá de sus metadatos

Las búsquedas no se realizan directamente sobre los documentos (la carga de procesamiento sería excesiva!)

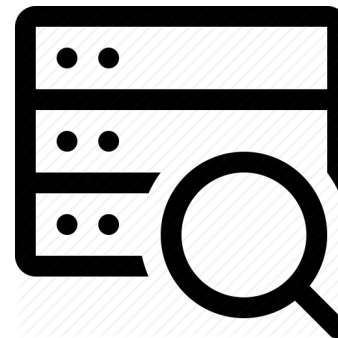
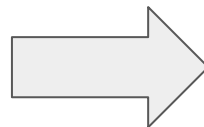
Requiere extraer el texto de los OD e integrarlo en un sistema de búsquedas



Documento



Texto extraído



Base de datos



# Base de datos de búsqueda

Se utiliza un indexador de texto

- optimizado para búsquedas
  - eficiencia
  - relevancia
  - filtrado
  - permite definir criterios de filtrado
  - permite especificar **pesos** en los campos
- permite gestionar grandes volúmenes de datos
- permite la **distribución** de conjuntos de datos (muy útil cuando estos crecen)



Más información: Solr in DSpace <https://wiki.duraspace.org/display/DSPACE/Solr>



# Interoperabilidad

Servicios, protocolos y directrices



# Servicios de Interoperabilidad

Los repositorios digitales deben pensarse como sistemas interoperable desde el principio

## Interoperabilidad **desde** el repositorio

- integrarse con otros sistemas de la institución
- ampliar el alcance y difusión de los contenidos
- incorporarse a sistemas o redes regionales e internacionales

## Interoperabilidad **hacia** el repositorio

- facilitar y/o agilizar la ingesta de contenidos
- modificar remotamente los contenidos



# Servicios de Interoperabilidad

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



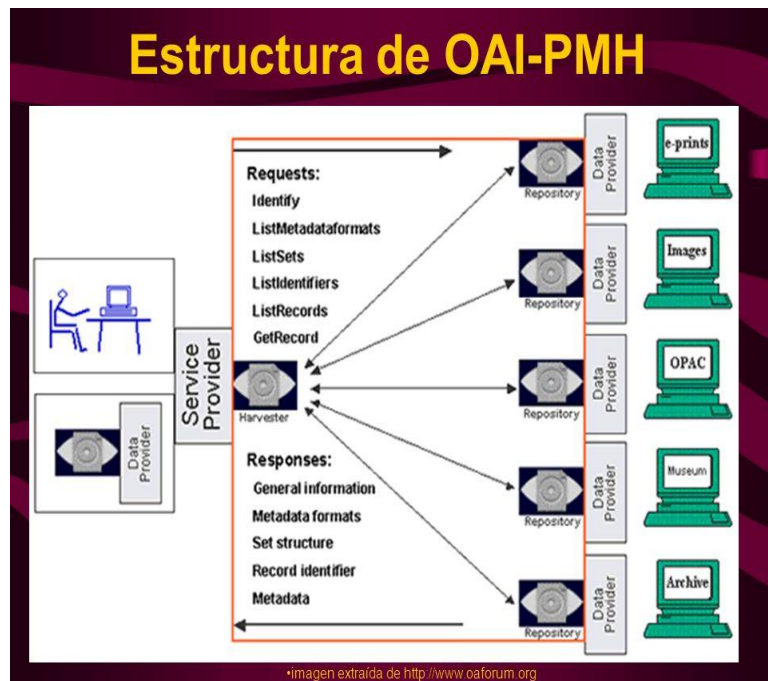
# Interoperabilidad - OAI PMH

## ¿Qué es?

*Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) es un mecanismo para permitir la interoperabilidad entre repositorios

## ¿Qué roles posee?

- Data Provider
- Service Provider (Harvester)



# Interoperabilidad - OAI PMH

¿Cómo se entienden los service y data provider?

- Modelo de metadatos Dublin Core
- Directrices
  - SNRD
  - Open Aire
  - Driver



# Interoperabilidad - OAI PMH - Directrices

## ¿Cuál es su objetivo?

El objetivo común de las distintas directrices es la normalización de la representación de los metadatos y el cumplimiento de ciertos metadatos de forma obligatoria, recomendada u opcional

## Ejemplo

	Driver 2.0	SNRD 2012	Open Aire 1.1
Uso de dc:rights	Uso: Recomendado Esquema: no especificado	Uso: Obligatorio Esquema: vocabulario OpenAIRE 1.1. Se debe exponer como primera instancia	Uso: Recomendado Esquema: vocabulario propio de 4 niveles de acceso: closed, embargoed, restricted y open Access



# Interoperabilidad - OAI PMH - Usos

¿ Qué utilidad tiene cumplir con las directrices ?

- SNRD ⇒ LA Referencia (<http://lareferencia.redclara.net/rfr/>)
- EUROPEANA (<http://www.europeana.eu/portal/en>)
- Tesis Latinoamericanas



# Interoperabilidad - OAI PMH

¿Cómo se comunican el service y data provider?

Utilizan verbs:

- Identify
- Get Record
- List Identifiers
- List Metadata Formats
- List Records
- List Sets



# Interoperabilidad - OAI PMH - Verbs

<i>Identify</i>	Se utiliza para obtener información del repositorio
<i>Get Record</i>	Se utiliza para obtener un registro específico
<i>List Identifiers</i>	Se utiliza para obtener los identificadores de los registros
<i>List Metadata Format</i>	Se utiliza para obtener los metadata format disponibles en el repositorio
<i>List Records</i>	Se utiliza para obtener todos los registros
<i>List Sets</i>	Se utiliza para obtener los sets

# Interoperabilidad - OAI PMH - DSpace

## DSpace como data provider

DSpace provee un módulo OAI 2.0 el cual permite al repositorio funcionar como data provider

## ¿Qué ventajas brinda?

Configuración flexible y adaptable basada en 4 conceptos:

- Contexto
- Filtro
- Transformador
- Mapeador



# Interoperabilidad - RSS y OpenSearch



RSS se utiliza para la sindicación de noticias y contenidos en línea

OpenSearch es un conjunto de tecnologías que permiten publicar los **resultados de una búsqueda** en un formato adecuado para la sindicación y agregación (fuente: Wikipedia)

- Permite que otras aplicaciones y sitios web expongan contenidos del repositorio
- Se integra fácilmente mediante RSS/Atom

## Últimos documentos agregados

Pamidronato endovenoso vs. Alendronato oral en el tratamiento de la osteoporosis establecida  
*Ferrari, Eliseo;*

Efecto de la fertilización con nitrógeno y fósforo sobre la acumulación de biomasa de pasturas de *Panicum coloratum* implantadas sobre suelos hidro-halomórficos de la Pampa Deprimida  
*Insausti, Mariano;*

Aislamiento, caracterización genética y análisis filogenético de *Rickettsia massiliae*, Ciudad Autónoma de Buenos Aires  
*Cicuttin, Gabriel L.;*

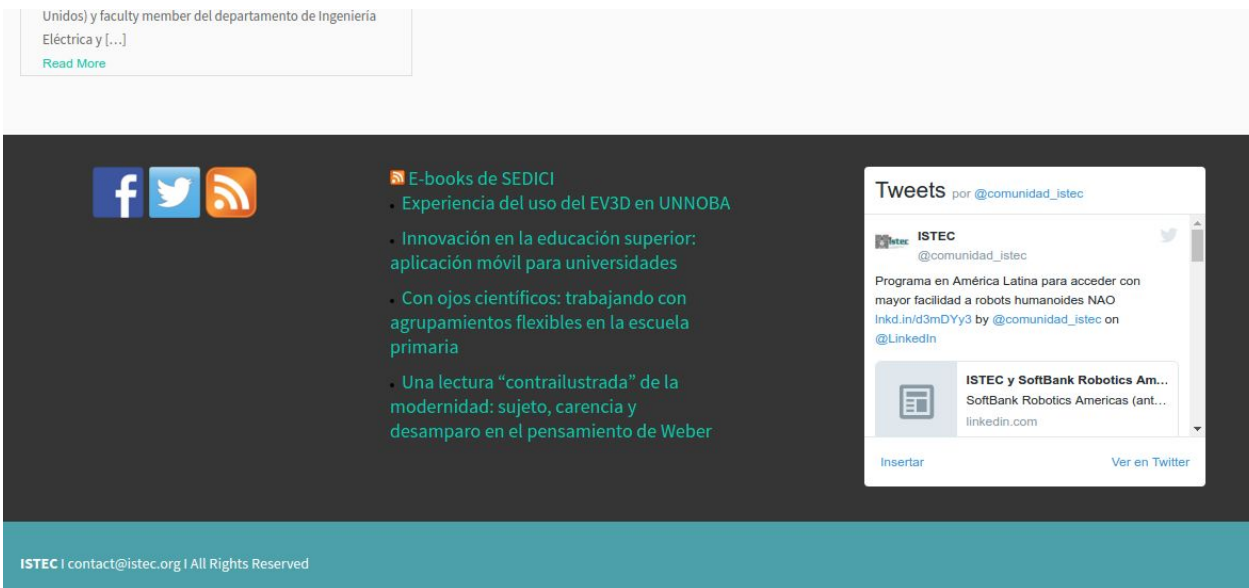
Propuesta pedagógica para la Cátedra de Climatología y Fenología Agrícola  
*Pinciroli, María;*

Agujeros negros astrofísicos  
*Pérez, Daniela;*







# Interoperabilidad - RSS y OpenSearch

Caso de uso: ISTECE ofrece a sus usuarios los últimos libros cargados en SEDICI




Unidos) y faculty member del departamento de Ingeniería Eléctrica y [...]  
[Read More](#)


 E-books de SEDICI

- Experiencia del uso del EV3D en UNNOBA
- Innovación en la educación superior: aplicación móvil para universidades
- Con ojos científicos: trabajando con agrupamientos flexibles en la escuela primaria
- Una lectura “contrailustrada” de la modernidad: sujeto, carencia y desamparo en el pensamiento de Weber

**Tweets** por @comunidad\_istec

 **ISTEC**  
@comunidad\_istec

Programa en América Latina para acceder con mayor facilidad a robots humanoides NAO  
lnkd.in/d3mDYy3 by @comunidad\_istec on @LinkedIn

 **ISTEC y SoftBank Robotics Am...**  
SoftBank Robotics Americas (ant...  
linkedin.com

Insertar [Ver en Twitter](#)

ISTEC | contact@istec.org | All Rights Reserved

Consorcio Iberoamericano para la Educación en Ciencia y Tecnología (ISTEC) [www.istec.org](http://www.istec.org)

sedici.unlp.edu.ar  
prebi.unlp.edu.ar



# Interoperabilidad - REST

- Basado sobre HTTP
- Muy simple y utilizado en aplicaciones web modernas
- Permite obtener datos o ejecutar operaciones sobre los datos
- Intercambio de información en cualquier formato: XML, JSON, etc
- Evita las abstracciones adicionales de otros protocolos de intercambio de mensajes (ej. SOAP)

AngularJS &  
RESTful API  
GET PUT POST DELETE



# Interoperabilidad - REST

→ El módulo REST API de Dspace provee una interfaz de acceso a Comunidades, Colecciones, Ítems y Bitstreams

→ Desde la versión 4 de DSpace

→ Provee varios *endpoints*

→ Soporta XML y JSON

- ◆ `curl -s -H "Accept: application/xml" http://localhost:8080/rest/communities | xmllint --format -`
- ◆ `curl -s -H "Accept: application/json" http://localhost:8080/rest/communities | python -m json.tool`

→ 4 métodos HTTP : GET (retornar), POST (crear), PUT (actualizar) y DELETE (eliminar)

Specific Bitstream	/bitstreams/:bitstreamID
Download a Bitstream	/bitstreams/:bitstreamID/retrieve
Bitstream Expands	parent, all

List Communities	/communities/
Specific Community	/communities/:communityID
Community Expands	parentCommunity, collections, subCommunities, logo, all

DSpace REST API <https://wiki.duraspace.org/display/DSDOC5x/REST+API>



# Interoperabilidad - SRU/SRW

SRU: Search / Retrieve Via URL.

- Protocolo estándar de búsqueda sobre Internet
- Las consultas se expresan en CQL (Contextual Query Language, Z39.50)
- Principal ventaja: fácil y sencillo
- Respuesta en formato XML

Ej. de request SRU:

<http://z3950.loc.gov:7090/voyager?version=1.1&operation=searchRetrieve&query=dinosaur>

The Library of Congress - SRU 2.0 <https://www.loc.gov/standards/sru/sru-2-0.html>



# Interoperabilidad - SRU/SRW

Search/Retrieve Web Service (SRU via HTTP SOAP)

Similar a SRU, pero

- utiliza XML para formatear los mensajes (human friendly)
- transferencia sobre SOAP y HTTP
- la interfaz SOAP simplifica la realización de consultas



# Interoperabilidad - SWORD

- SWORD (Simple Web-service Offering Repository Deposit) permite el **depósito remoto** de contenidos en el repositorio
- Se integra con otras aplicaciones
  - ◆ Existen librerías para desarrolladores en PHP, Java, Ruby, Python ( <http://swordapp.org/> )
- El servidor SWORD expone una interfaz (ServiceDocument), a la que se le envía un documento XML, que es procesado y transformado a la representación interna del repositorio
- Permite la transferencia de metadatos y binarios (archivos)
- Resulta particularmente útil para realizar el depósito de varios documentos a la vez



# Interoperabilidad - SWORD

## Caso de uso

### Portal de Revistas de la UNLP

- Una revista de la UNLP publica un nuevo número en su sitio web (OJS)
- Un administrador del sitio selecciona los artículos del nuevo número y los envía a una colección del repositorio SEDICI (< 1 min)
- Un administrador de SEDICI verifica y completa los registros, y publica el nuevo número en el repositorio (~30 min)

DSpace 5 SWORD v2 server

<https://wiki.duraspace.org/display/DSDOC5x/SWORDv2+Server>

Conector de depósito de importación/exportación SWORD

Punto de depósito: **sedici** **ANEXO: UNLP**

Nombre usuario/a:

Contraseña:

Punto de depósito: **Autoarchivo** **Actualizar**

Opciones:  Depositar galerías  Depositar los archivos de editorial más recientes

NÚMERO	TÍTULO	AUTORES/AS
<input checked="" type="checkbox"/> VOL_1_1006	Aplicação de processo oxidativo avançado baseado em fotocatalise heterogênea (TiO <sub>2</sub> /UV-vis) para o pré-tratamento de efluente têxtil	
<input checked="" type="checkbox"/> VOL_1_1005	Avaliação comparativa de íscas atrativas a partir da rigidez de espécies de formigas (Hymenoptera: Formicidae) numa floresta de Eucalyptus grandis, em Santa Maria, Rio Grande do Sul, Brasil	
<input checked="" type="checkbox"/> VOL_1_1004	Avaliação da Redução da Poluição do Chorume Tratado por Processo Fotoquímico	



**SEDICI**  
REPOSITÓRIO INSTITUCIONAL DE LA UNLP

Inicio Buscar material Subir material

Administración

**Tareas del flujo de trabajo**

Estas tareas son ítems que están esperando aprobación antes de ser añadidas al repositorio. Hay dos colas de tareas, una para tareas que ha aceptado y otra para tareas que todavía no han sido asumidas por nadie.

**Sus tareas**

Tarea	Ítem	Colección	Remitente
No tiene asignada ninguna tarea			

**Tareas en cola**

Tarea	Ítem	Colección	Remitente
<input type="checkbox"/> Revisión SEDICI	Aplicação de processo oxidativo avançado baseado e ...	Autoarchivo	Portal de Revistas UNLP
<input type="checkbox"/> Revisión SEDICI	Avaliação comparativa de íscas atrativas a partir ...	Autoarchivo	Portal de Revistas UNLP
<input type="checkbox"/> Revisión SEDICI	Avaliação da Redução da Poluição do Chorume Tratad ...	Autoarchivo	Portal de Revistas UNLP

# Servicios para preservar los contenidos

## *Preservación digital no es hacer backups*

¿Qué servicios trae el software del repositorio para "ayudar" a preservar los recursos?

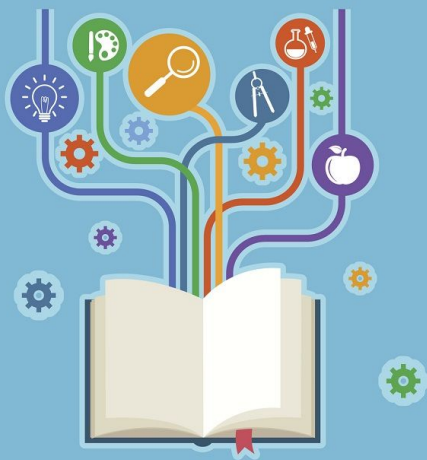
- control de integridad
- metadatos de preservación
- permite integración con aplicaciones o servicios de terceros



# Estadísticas

- ¿Qué tipo de información sobre el uso registra? (accesos, descargas, navegación)
- ¿Cómo se recolectan estos datos? logs, base de datos, software dedicado, servicio tercerizado
- ¿Cuán fiable son estos datos? Múltiples accesos, bots
- ¿Cómo se interpretan los datos, qué estadísticas genera?
- ¿Cómo se visualizan las estadísticas? cada cuánto se actualizan?
- Diferencia entre estadísticas web estándares (google analytics, PIWIK, awstats) y estadísticas propias de un repositorio





## Clase 3

### Derechos, licencias, vías de publicación y visibilidad



## Clase 3

- Propiedad intelectual. ¿Cómo mantener los derechos? El derecho de autor en la Argentina
- El **acceso abierto** y sus vías alternativas a los circuitos de publicación tradicionales.
- Información sobre editores y sus políticas. Contratos y adendas.
- Licencias de uso **Creative Commons**.
- Políticas del SEDICI.
- ¿Por qué conviene depositar en abierto? Visibilidad e impacto.
- El ranking web de webometrics.



# Objetivos

Que conozcan sus derechos como autores.

Destacar la importancia de gestionar correctamente los derechos de autor de una obra.

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 interesados.

Conocer dónde conviene publicar para tener más impacto y visibilidad.



# ¿Cómo mantenerlos?



## ¿Qué es la propiedad intelectual?

Versiones en otros idiomas: No todas las páginas están disponibles en todos los idiomas. Si una página no está disponible en el idioma de su elección, se abrirá la página en inglés.

La propiedad intelectual (P.I.) se relaciona con las creaciones de la mente: invenciones, obras literarias y artísticas, así como símbolos, nombres e imágenes utilizados en el comercio.

La legislación protege la P.I., por ejemplo, mediante las [patentes](#), el [derecho de autor](#) y las [marcas](#), que permiten obtener reconocimiento o ganancias por las invenciones o creaciones. Al equilibrar el interés de los innovadores y el interés público, el sistema de P.I. procura fomentar un entorno propicio para que prosperen la creatividad y la innovación.

- ¿Qué es la P.I.? [PDF](#)
- [Entender el derecho de autor y los derechos conexos](#) | [Entender la propiedad industrial](#)
- [WIPO Intellectual Property Handbook](#) - Una guía detallada sobre políticas, normas y uso de la P.I.

### DESTACADO



(FOTO: ISTOCKPHOTO.COM/ANDRESR)

### El deporte y la P.I.

Conozca cómo se usa la propiedad intelectual en el mundo del deporte.



**OMPI**

ORGANIZACIÓN MUNDIAL DE LA PROPIEDAD INTELECTUAL

[Prensa](#)[Reuniones](#)[Contacto](#)[Mi cuenta](#)[Español](#)[Servicios de P.I.](#)[Política](#)[Cooperación](#)[Conocimiento](#)[Sobre la P.I.](#)[La OMPI](#)

### Bases de datos de P.I.

[PATENTSCOPE](#)[Base Mundial de Datos sobre Marcas](#)[ROMARIN](#)[Base Mundial de Datos sobre Dibujos y Modelos](#)[Hague Express](#)[Artículo 6ter](#)

### Recursos jurídicos

[Leyes y tratados sobre P.I. \(WIPO Lex\)](#)[Tratados administrados por la OMPI](#)

### Recursos técnicos

[Clasificaciones internacionales](#)[Normas \(Manual de la OMPI\)](#)[Terminología \(WIPO Pearl\)](#)

### Recursos de información

[Documentos](#)[Estadísticas](#)[Publicaciones](#)[Perfiles nacionales](#)[Estudios de caso](#)[Biblioteca](#)

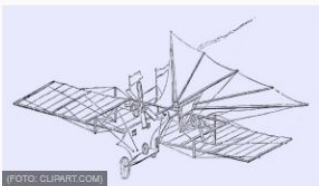
## Tipos de propiedad intelectual



(FOTO: ISTOCKPHOTO.COM/PROFESSOR25)

### Derecho de autor

En la terminología jurídica, la expresión derecho de autor se utiliza para describir los derechos de los creadores sobre sus obras literarias y artísticas. Las obras que abarca el derecho de autor van desde los libros, la música, la pintura, la escultura y las películas hasta los programas informáticos, las bases de datos, las publicidades, los mapas y los dibujos técnicos.



(FOTO: CLIPART.COM)

### Patentes

Una patente es un derecho exclusivo que se concede sobre una invención. En términos generales, una patente faculta a su titular a decidir si la invención puede ser utilizada por terceros y, en ese caso, de qué forma. Como contrapartida de ese derecho, en el documento de patente publicado, el titular de la patente pone a disposición del público la información técnica relativa a la invención.



(FOTO: IMPOISEN/A)

### Marcas

Una marca es un signo que permite diferenciar los productos o servicios de una empresa de los de las demás. Las marcas se remontan a los tiempos en que los artesanos reproducían sus firmas o "marcas" en sus productos.



(FOTO: MINALISTAMATI)

### Diseños industriales

Un diseño industrial (dibujo o modelo industrial) constituye el aspecto ornamental o estético de un artículo. El diseño puede consistir en rasgos tridimensionales, como la forma o la superficie de un artículo, o en rasgos bidimensionales, como motivos, líneas o colores.



(FOTO: ISTOCKPHOTO.COM/MATTJEACOCK)

### Indicaciones geográficas

Una indicación geográfica es un signo utilizado para productos que tienen un origen geográfico concreto y cuyas cualidades, reputación o características se deben esencialmente a su lugar de origen. Por lo general, la indicación geográfica consiste en el nombre del lugar de origen de los productos.



# Categorías de propiedad intelectual

Según la [Organización Mundial de la Propiedad Intelectual \(OMPI\)](#), el propietario o titular de una obra puede disponer de ésta como le plazca y ninguna otra persona física o jurídica podrá disponer legalmente de su propiedad sin su consentimiento. Naturalmente, el ejercicio de este derecho está sujeto a limitaciones.

Según del bien que trate la propiedad intelectual se divide en Propiedad Industrial y Derechos de Autor.



# ¿Qué es una patente?

Una patente es un derecho exclusivo que se concede sobre una invención. En términos generales, una patente faculta a su titular a decidir si la invención puede ser utilizada por terceros y, en ese caso, de qué forma. Como contrapartida de ese derecho, en el documento de patente publicado, el titular de la patente pone a disposición del público la información técnica relativa a la invención.



# ¿Qué es una patente?

## ¿Qué tipo de protección ofrece una patente?

En principio, el titular de la patente goza del derecho exclusivo a impedir que la invención patentada sea explotada comercialmente por terceros. La protección por patente significa que una invención no se puede producir, usar, distribuir con fines comerciales, ni tampoco vender, sin que medie el consentimiento del titular de la patente



# ¿Qué es una patente?

## ¿Una patente es válida en todo el mundo?

Las patentes son derechos territoriales. Por lo general, los derechos exclusivos correspondientes solos tienen validez en el país o la región en los que se ha presentado la solicitud y se ha concedido la patente, de conformidad con la normativa de ese país o esa región



# ¿Qué es una patente?

## ¿Cuánto dura el plazo de protección de una patente?

La protección se concede por un período limitado, que suele ser de 20 años a partir de la fecha de presentación de la solicitud.



# ¿Qué puede protegerse por derecho de autor?

La legislación no suele contener una lista exhaustiva de las obras que ampara el derecho de autor. No obstante, en términos generales, entre las obras habitualmente protegidas por el derecho de autor en todo el mundo están:

- las obras literarias como las novelas, los poemas, las representaciones escénicas, las obras de referencia, los artículos periodísticos;
- los programas informáticos y las bases de datos;
- las películas, las composiciones musicales y las coreografías;
- las obras artísticas como los cuadros, los dibujos, las fotografías y las esculturas;
- la arquitectura; y
- los anuncios, los mapas y los dibujos técnicos.



# ¿Qué puede protegerse por derecho de autor?

- La protección del derecho de autor abarca sólo las expresiones, pero no las ideas, procedimientos, métodos de operación o conceptos matemáticos en sí. El derecho de autor puede amparar o no elementos como los títulos, los lemas o logotipos, dependiendo de que la paternidad de la obra sea suficiente.



# ¿Qué derechos comprende el derecho de autor?

- Los derechos patrimoniales, que permiten que el titular de los derechos obtenga compensación financiera por el uso de sus obras por terceros; y

En la mayoría de los casos, en la legislación de derecho de autor se estipula que el titular de los derechos goza del derecho patrimonial a autorizar o impedir determinados usos de la obra o, en algunos casos, a recibir una remuneración por el uso de la obra.



# ¿Qué derechos comprende el derecho de autor?

El titular de los derechos patrimoniales de una obra puede prohibir o autorizar:

- la reproducción de su obra de varias formas, como la publicación impresa o la grabación sonora;
- la interpretación o ejecución públicas, por ejemplo en una obra dramática o musical;
- la grabación de la obra, por ejemplo en forma de discos compactos o DVD;
- la radiodifusión de la obra por radio, cable o satélite;
- la traducción de la obra a otros idiomas; y
- la adaptación de la obra, como en el caso de una novela adaptada para un guión.



# ¿Qué derechos comprende el derecho de autor?

- Los derechos morales, que protegen los intereses no patrimoniales del autor.
- Entre los ejemplos de derechos morales universalmente reconocidos están el derecho a reivindicar la paternidad de la obra y el derecho a oponerse a toda modificación de la obra que pueda perjudicar la reputación del creador.



## Derechos morales

- Otorgan al autor el derecho a ser reconocido como tal, y a la **integridad** de la obra.
  - NO SON TRANSFERIBLES.
  - NO CADUCAN
- Están reconocidos en la legislación argentina (Artículos 51 y 52 de la ley de PI 11.723), en la mayoría de jurisdicciones europeas y de latinoamérica, pero no en otras como en el caso de Estados Unidos.

# Derechos Patrimoniales o de explotación

- Derecho a publicar y distribuir, a reproducir la obra, realizar traducciones, mostrar en público y autorizar a otros a que ejerzan cualquiera de estos derechos,
- Son transferibles.
- En el derecho anglosajón se utiliza la noción de *copyright* (traducido literalmente como "derecho de copia") que —por lo general— comprende la parte patrimonial de los derechos de autor.

## ¿Se registra la obra?

En la mayoría de los países, y conforme a lo que se dispone en el [Convenio de Berna](#), la protección del derecho de autor se obtiene automáticamente sin necesidad de efectuar ningún registro ni otros trámites.

No obstante, en la mayoría de los países existe un sistema de registro y depósito facultativo de obras; estos sistemas facilitan, por ejemplo, las aclaraciones de las controversias relacionadas con la titularidad o la creación, las transacciones financieras, las ventas, las cesiones y transferencias de derechos.

La OMPI no cuenta con un sistema de registro del derecho de autor ni con una base de datos de derecho de autor que permita realizar búsquedas



# ¿Una traducción es una nueva obra?

La transformación de una obra comprende su traducción, adaptación y cualquier otra modificación en su forma (no de su contenido) de la que se derive una obra diferente. Para realizar el acto de transformación se precisará en todo caso, como para cualquier explotación de la obra, la correspondiente autorización del autor o del titular de los derechos.

Una vez realizada la transformación existen dos obras diferentes, la preexistente y la derivada, con lo que ambas generan derechos de explotación de titularidad independiente. No obstante, para la explotación de la obra resultante se precisará de la necesaria autorización del autor de la obra original, durante la duración del plazo legal de la protección.



# ¿Una traducción es una nueva obra?

Es decir, en el caso de una transformación se precisan dos autorizaciones: la primera para realizar la modificación de la obra preexistente y la segunda para la explotación, en cualquier forma de la obra derivada surgida. Dicho de otra forma, con la autorización de la transformación irán ligados los extremos de cómo se puede usar esa obra derivada.



# APRENDER DEL PASADO PARA CREAR EL FUTURO: LAS CREACIONES ARTÍSTICAS Y EL DERECHO DE AUTOR



OMPI  
ORGANIZACIÓN MUNDIAL  
DE LA PROPIEDAD  
INTELLECTUAL

Plagio es el acto de copiar una obra, entera o parcialmente, pretendiendo ser su autor original.

Como hemos visto en la sección anterior, las ideas en sí no están protegidas por el derecho de autor, por lo tanto no hay nada malo en escribir sobre una idea que encontramos en otra obra.

Sin embargo, para evitar el plagio, tenemos que expresar la idea de una manera personal y única.



sedici.unlp.edu.ar  
prebi.unlp.edu.ar



# Copyleft

El ***copyleft*** es una práctica legal que consiste en el ejercicio del [derecho de autor](#) con el objetivo de propiciar el libre uso y distribución de una obra, exigiendo que los concesionarios preserven las mismas libertades al distribuir sus copias y derivados.

El término surge en la comunidad del [software libre](#) como un juego de palabras en torno a [copyright](#): «derecho de autor» en inglés (literalmente «derecho de copia») con otro sentido, el de *left*: pretérito del verbo dejar o permitir (literalmente «dejada-copia»), así como izquierda, en contraste con *right*, que es derecho. (Véase la sección de [Etimología](#)).



# Copyleft

Los autores pueden aplicar una licencia con copyleft a cualquier tipo de trabajo creativo que sea regido por el [derecho de autor](#).

Es una alternativa contra las restricciones al público en las que incurren los editores y la industria del entretenimiento. Se pretende así ofrecer a quienes poseen los [derechos patrimoniales](#) la posibilidad de que ofrezcan una [licencia libre](#) para su obra; al mismo tiempo que una cláusula adicional (el copyleft) protege los derechos expuestos en la licencia de intentos subsecuentes de privatización (mientras dure el derecho de autor). Las licencias con copyleft son entonces una de las formas principales de licencia libre; en contraste con las llamadas [licencias permisivas](#) o sin copyleft, y en contraste con el [dominio público](#).

Referencia: Wikipedia: <https://es.wikipedia.org/wiki/Copyleft>



# Licencias libres

En el mundo del software libre existen diversos tipos de licencias bajo las cuales se amparan las producciones realizadas por los desarrolladores y/o usuarios, por ejemplo:

**GPL:** GNU General Public License. Es la más conocida, cubre la mayor parte del software de la Free Software Foundation, y otros muchos programas.

**Creative Commons:** está inspirada en la licencia GPL de la Free Software Foundation. La idea principal es posibilitar un modelo legal y ayudado de herramientas informáticas para así facilitar la distribución y el uso de contenidos para el dominio público.

# Derechos de autor

Desde la creación de una obra, la ley le reconoce al autor unos derechos como tal, los que básicamente se dividen en dos ramas

DERECHOS DE AUTOR	
<b>MORALES</b>	<b>DE EXPLOTACIÓN O PATRIMONIALES (COPYRIGHT)</b>
<ul style="list-style-type: none"><li>- Reconocimiento de autoría</li><li>- Integridad de la obra</li></ul>	<ul style="list-style-type: none"><li>- Distribución</li><li>- Reproducción</li><li>- Comunicación pública</li><li>- Transformación</li></ul>
<ul style="list-style-type: none"><li>- Intransferibles</li><li>- No caducan</li></ul>	<ul style="list-style-type: none"><li>- Transferibles</li><li>- Caducan</li></ul>

TRANSFERENCIA DE LOS DERECHOS DE EXPLOTACIÓN	
<b>A editores</b> Copyright transfer	<p><b>Exclusiva</b></p> <p>El autor pierde todos los derechos cedidos sobre su obra según la vigencia del contrato.</p> <hr/> <p><b>No exclusiva</b></p> <p>El autor mantiene todos los derechos sobre su obra. Se establecen unas condiciones de explotación entre el autor y la otra parte.</p>
<b>A lectores/usuarios</b> Licencias de uso	<p>El autor autoriza algunos usos y define restricciones. Posibles usos: Redistribuir/copiar, adaptar.</p> <p><b>Tipos de licencias:</b></p> <ul style="list-style-type: none"><li>- <b>Preexistentes:</b> Creative Commons, GPL, Open Data Commons, etc.</li><li>- <b>Ad-hoc:</b> el autor establece la licencia de uso de sus obras en términos propios.</li></ul>



# Duración de los derechos de autor

- **Una obra pasa al dominio público cuando los derechos patrimoniales han expirado.** Esto sucede habitualmente trascurrido un determinado plazo desde la muerte del autor: en tal momento puede ser utilizada en forma libre, respetando los derechos morales.
- En Argentina, el Derecho de autor está enmarcado, en principio, por el **artículo 17 de la Constitución** que expresa que *"Todo autor o inventor es propietario exclusivo de su obra, invento o descubrimiento, por el término que le acuerde la ley"*.
- El **artículo 5 de la Ley de la Ley 11723** que regula el Régimen Legal de la Propiedad Intelectual en Argentina dice que: *"La propiedad intelectual sobre sus obras corresponde a los autores durante su vida y a sus herederos o derechohabientes hasta setenta años contados a partir del 1 de Enero del año siguiente al de la muerte del autor"*.



# Legislación en Ecuador

[http://www.wipo.int/wipolex/es/text.jsp?file\\_id=195678](http://www.wipo.int/wipolex/es/text.jsp?file_id=195678)



# Derechos de autor. Focalización e intereses

- Esta parte de la charla se concentrará en los derechos patrimoniales, de explotación o copyright, a los que además interesa vincularlos y especialmente ver cómo explotarlos en relación a las vías de publicación en abierto: repositorios y revistas de acceso abierto.



# ¡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.

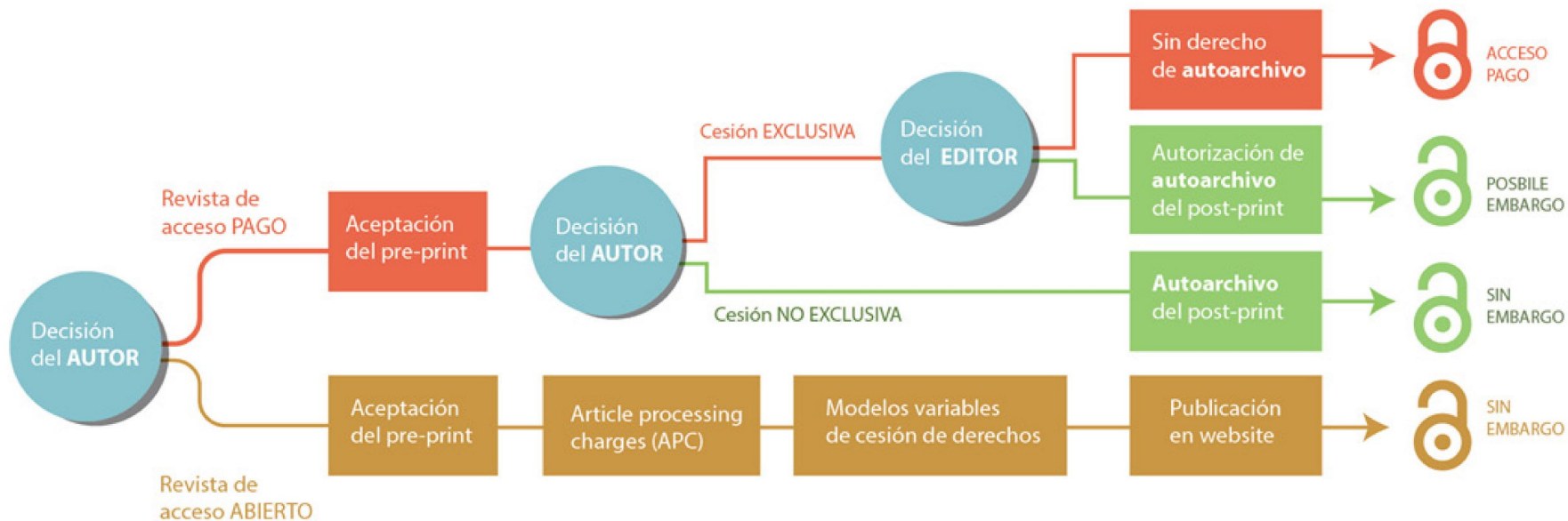




## Ruta roja, verde y dorada

Foto: Orange Gold Red Green Autumn Forest Trees. By Jansen Fenstermacher







# VÍA ROJA

Circuito tradicional de publicación

(son las revistas a las que sólo se accede por pago)



# Vía roja: Cesión Exclusiva



Si la revista retiene los derechos de explotación de la obra **de manera exclusiva** (el autor los cede por completo).

Según la cesión firmada puede ocurrir que:

- No se permita el autoarchivo en un repositorio
- El lector paga para ver los artículos.

Además, el autor deberá pedir permiso para cualquier uso que desee realizar con el trabajo.



# Vía roja: Cesión Exclusiva



Es necesario analizar el contrato, hacer agregados o acuerdos para tener alguna/s alternativa que le resulte interesante:

- publicar en la página web personal,
- compartir en la institución a la que pertenece con alumnos y colaboradores,
- usar su trabajo para la creación de otro trabajo,
- publicar en un repositorio institucional,
- publicar en un repositorio temático...

# Vía roja: Cesión Exclusiva



Si el contrato no dice explícitamente que la cesión es **NO EXCLUSIVA**, entonces **ES EXCLUSIVA**.

# Recomendaciones a los autores

- Evitar ceder exclusivamente todos los derechos patrimoniales o de explotación al celebrar un contrato con una editorial.
- En caso que sólo se pueda firmar una cesión exclusiva:
  - o Utilizar una adenda al contrato con la editorial:
    - consiste en añadir un apéndice especial en el contrato para mantener algunos derechos de uso. (ej: depositar en el repositorio institucional)
  - o Hay muchos modelos de adenda e incluso herramientas para obtener adendas en línea. Ver SPARC





# VÍA VERDE

Los repositorios



# La vía verde



Si el autor

- no cedió derechos sobre la obra, o
  - cedió sus derechos de forma **NO-EXCLUSIVA**, o
  - incorporó una adenda a la cesión Exclusiva firmada con el editor
- puede realizar el depósito de su obra en un repositorio institucional o temático. Es decir, habilita la vía verde



# Repositorios



- Los repositorios son archivos digitales provistos de 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 organización.
- Los repositorios deben tener una política definida y asegurar:
  - El autoarchivo
  - La interoperabilidad
  - El libre acceso
  - La preservación a largo plazo



# Repositorios: autoarchivo



La acción por parte del autor de depositar un artículo o cualquier otra obra en estos repositorios es lo que se conoce con el nombre de AUTOARCHIVO.

La mayoría de los editores científicos de suscripción permiten que los autores **autoarchiven** sus trabajos en repositorios abiertos. Aunque, en función de las condiciones establecidas por los editores, el autor podrá o no autoarchivar versiones diferentes de sus artículos.



# Las ventajas del repositorio



- Se ofrece un punto de acceso uniforme a la información de la institución y del autor.
- El material es catalogado y descrito en profundidad, maximizando su accesibilidad.
- Se preserva la integridad de las obras
- Trabaja bajo estándares que permiten la exposición automática de los registros a otros portales académicos como OAISTER/OCLC, RECOLECTA, BASE-SEARCH, entre otros.
- El material se distribuye junto al detalle de sus derechos de uso, notificando a los lectores los usos permitidos.



# Los recolectores académicos. Otros



- BASE - <http://www.base-search.net/>
- Cybertesis - <http://www.cybertesis.net/>
- DART-Europe - <http://www.dart-europe.eu/>
- DIALNET - <http://dialnet.unirioja.es/>
- DRIVER - <http://search.driver.research-infrastructures.eu/>
- EUROPEANA - <http://www.europeana.eu/>
- HISPANA - <http://hispana.mcu.es>
- OAISTER - <http://www.oclc.org/oaister/>
- RECOLECTA - <http://www.recolecta.net/>
- TESISENRED - <http://www.tesisenred.net/>



# VÍA DORADA

## Las Revistas Abiertas



# Tipos de Revistas de Acceso Abierto



**Las revistas de acceso abierto** llamadas **genuinas** son las que exponen de manera gratuita por vía electrónica TODOS sus artículos sin cargos de suscripción **para el lector**.

- Ej: PLOS, SCIELO

**Las revistas híbridas** son las que exponen de manera gratuita por vía electrónica ALGUNOS de sus artículos sin cargos de suscripción **para el lector**. El acceso al resto es a través del pago por parte del lector.

- Ejemplo: OpenChoice de Springer

# Revistas genuinas de Acceso Abierto

## Modelo de financiación

Las revistas en acceso abierto se sostienen a través de diferentes modelos de financiación y son publicadas por editores científicos bajo modelo de suscripción o por editores e iniciativas completamente dedicadas al acceso abierto.

- Ejemplos de este tipo de revistas se recogen en el Directorio de Revistas de Acceso Abierto (DOAJ).

**DOAJ** DIRECTORY OF  
OPEN ACCESS  
JOURNALS



# Revistas genuinas de Acceso Abierto.

## Modelo de financiación



- El caso más puro según la definición de open access sería aquella revista que ni lector ni autor paga por publicar y son los autores los que retienen el copyright sobre sus trabajos cediendo los derechos no exclusivos de publicación a la revista.
- Obviamente alguien *“tiene que pagar”*, en este caso, las revistas generalmente pertenecen a instituciones académicas o sociedades profesionales, cuyos recursos incluyen también el mantenimiento de estas publicaciones.



# Modelo Alternativo: el autor paga



## Cargos de publicación bajo Acceso Abierto

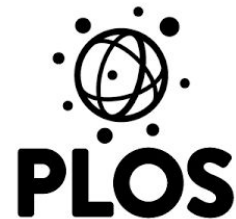
Algunas revistas de acceso abierto cobran ciertos gastos de publicación para dar acceso electrónico directo a los artículos.

En el caso de algunas revistas, estos honorarios son **pagados por los miembros de una Sociedad o institución o por el propio autor.**

# Modelo Alternativo: el autor paga. Caso PLoS

- **Es una Editorial Open Access**
- mantiene 9 publicaciones periódicas en las áreas de medicina, biología y temáticas relacionadas
- Los lectores ven abiertamente los artículos.
- **las tasas de la publicación son pagadas por los autores** (o sus instituciones ).
- Los números de los que se habla resultan muy altos.
  - PLoS Biology *US\$2900*
  - PLoS Medicine *US\$2900*
  - PLoS Computational Biology *US\$2250*
  - PLoS Genetics *US\$2250*
  - PLoS Pathogens *US\$2250*
  - PLoS Neglected Tropical Diseases *US\$2250*
  - PLoS ONE *US\$1350*

<http://www.plos.org/publish/pricing-policy/publication-fees/>



# Revistas híbridas de Acceso Abierto



## Un Panorama complejo

Entre las revistas de modelo de acceso pago (tradicional) y las de modelo totalmente abierto, se dan casi todas las posibles combinaciones entre pago y derechos de copyright.



# Revistas híbridas de Acceso Abierto



Las grandes editoriales (Elsevier, Springer, Blackwell, etc. ) ofrecen modelos híbridos en algunas de sus revistas:

- se **paga** a la editorial para que el artículo esté en acceso abierto, junto con otros que no lo están.
- Los precios varían de editorial a editorial y de revista a revista.
- Elsevier lo llama “*sponsored access*”:  
<http://www.elsevier.com/physical-sciences/mathematics/sponsored-access> y lo ofrece para todas sus revistas.
- Springer lo llama "Open Choice".

Opiniones:

<https://blogs.ch.cam.ac.uk/pmr/2012/08/05/elsevier-replies-about-hybrid-openaccess-i-am-appalled-about-their-practices-breaking-licences-and-having-to-pay-to-read-open-access/>

# Recursos para aprender



- El proyecto [Open Access Publishing in European Networks](#) (OAPEN) brinda la información básica y una lista de editores que se especializan en la publicación de libros de acceso abierto.
- El Directorio de Acceso Abierto (OAD) es un compendio sobre hechos en acceso abierto (OA) a la ciencia mantenido por la comunidad de la OA en general: [http://oad.simmons.edu/oadwiki/Main\\_Page](http://oad.simmons.edu/oadwiki/Main_Page)

# Recursos para aprender

- Sherpa-Romeo
  - permite buscar información sobre la política de derechos de autor de 1823 editores en relación con el acceso abierto. (Al 4/4/2015): <http://www.sherpa.ac.uk/romeo/statistics.php>

Siempre es necesario contrastar esta información con el contenido de la transferencia de derechos que le fueron presentados a un autor al firmar el contrato. Sólo el acuerdo firmado en sí mismo es jurídicamente vinculante.



# Recomendaciones a los autores

## El camino verde, ¡la mejor vía!

Al depositar una copia de la versión de presentada al editor en un repositorio institucional/temático tan pronto como sale a la luz se garantiza que la obra gane rápidamente **visibilidad** y se mantenga permanentemente **accesible**.



# 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 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?.



# Derechos de autor. Entre editor y autor

Los acuerdos con las editoriales no están estandarizados, no son fáciles de entender y mucho menos de encontrar en las páginas de los editores.

El editor de una revista, en general, pide como requisito para publicar el artículo de un autor la cesión, mediante un contrato por escrito, de algunos o de todos los derechos de explotación. Lo que el autor pueda hacer con su artículo a posteriori, dependerá de las condiciones que haya firmado con el editor.

Si ha firmado una cesión EXCLUSIVA, el editor obtiene el control de la explotación de la obra, y el autor la pierde completamente. En este caso, es probable que el autor deba pedir permiso al editor para subirlo a su propia página web, al campus virtual de su universidad, para traducirlo, compartirlo en redes sociales con colegas de su especialidad, para reusar partes de la obra para generar una nueva, o bien depositarlo en un repositorio abierto.

# Clave: conocer el contrato y los permisos

Si un autor, leyendo el contrato que le propone el editor, no está de acuerdo con las condiciones establecidas, puede y es aconsejable que negocie con el mismo un ADDENDUM o adenda en el cual se describan los derechos que desea retener. Si bien muchos editores aceptan este tipo de enmiendas, algunos no lo hacen. En todos los casos se aconseja revisar detenidamente los términos de contrato que propone la revista en cuestión.

Como un ejercicio posible en relación a las dificultades para comprender los términos de la licencia se propone la lectura de esta propuesta:

[http://www.sciencemag.org/site/feature/contribinfo/prep/lic\\_info.pdf](http://www.sciencemag.org/site/feature/contribinfo/prep/lic_info.pdf)



# Relación entre el autor y el editor

Hasta aquí resulta claro que los autores deben conocer cuáles son sus derechos referidos a:

- compartir su obra con otros investigadores;
- reutilizarla con fines educativos;
- difundirla en acceso abierto en repositorios (institucionales o temáticos);
- reutilizarla en un trabajo posterior;
- traducirla, etc.

En la relación autor-editor hay un vínculo contractual, es decir un contrato que el autor establece (y firma) con la editorial **y es lo que debe conocer de antemano.**



# Conocer los permisos de las editoras

## La Editorial Elsevier

[www.elsevier.com/wps/find/authorsview.authors/rights](http://www.elsevier.com/wps/find/authorsview.authors/rights) permite el autoarchivo de preprints sin requerir autorización en cualquier tipo de repositorios. Sin embargo, prohíbe el autoarchivo de postprints (del editor) en cualquier repositorio temático o institucional, aun en los casos que exista un mandato institucional. En ciertos casos, que explícitamente detallan, se han firmado acuerdos para permitir que se cumplan estos mandatos institucionales. Esos acuerdos, en general, obligan a los autores o entidades financiadores a pagar una suma que “cubra” los gastos del proceso de revisión por pares llevado adelante por el cuerpo editorial

# Conocer los permisos de las editoras

La **Editorial Wiley-Blackwell** en su apartado “Author Rights”: [authorservices.wiley.com/bauthor/benefits.asp](http://authorservices.wiley.com/bauthor/benefits.asp) menciona en un párrafo breve que los autores pueden usar sus artículos en un número dado de modalidades que incluyen el reuso para nuevos trabajos y como material en los cursos de su institución pero que tales prácticas varían según la revista y aconseja a los autores a consultar la forma de copyright en cada caso. También incluye un set de preguntas frecuentes: [http://authorservices.wiley.com/bauthor/faqs\\_copyright.asp](http://authorservices.wiley.com/bauthor/faqs_copyright.asp)



## SHERPA/ROMEO <http://www.sherpa.ac.uk/romeo/>

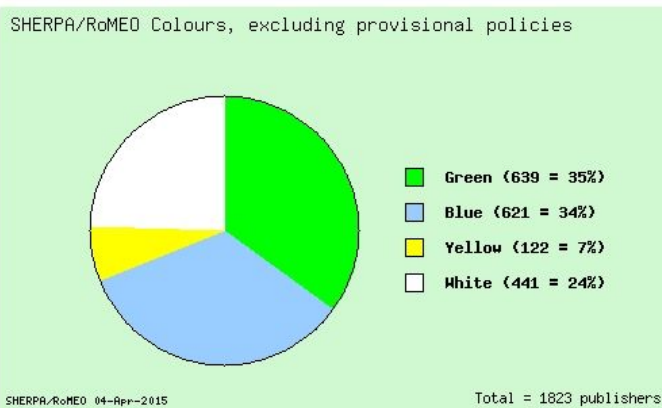
Color ROMEO	Política de archivo
Verde	permite archivar el preprint y el postprint o version final.
Azul	permite archivar el postprint o version final .
Amarillo	permite archivar el preprint, la primera revisión del trabajo antes de la revisión por pares.
Blanco	no permite archivar o no esta definido.

Un listado de 1823 editores y las 2716 revistas con sus colores es expuesto en:  
<http://www.sherpa.ac.uk/romeo/journalbrowse.php?la=en&fIDnum=|&mode=simple>

# Estadísticas de 1823 editoriales en la base de datos RoMEO

RoMEO colour	Archiving policy	Publishers	%
<a href="#">green</a>	Can archive pre-print and post-print	639	35
<a href="#">blue</a>	Can archive post-print (ie final draft post-refereeing)	621	34
<a href="#">yellow</a>	Can archive pre-print (ie pre-refereeing)	122	7
<a href="#">white</a>	Archiving not formally supported	441	24

Summary: **76%** of publishers on this list formally **allow** some form of self-archiving.



# Adicionalmente

Romeo realiza consultas sobre:

- El servicio Zetoc de la British Library
- DOAJ:
  - 10,031 Journals
  - 7,280 searchable at Article level
  - 122 Countries
  - 2,594,511 Articles
- Entrez: portal de acceso a la biblioteca nacional de medicina y sus journals.



# Ejemplo de Sherpa Romeo con IEEE Transactions on Signal Processing

## Search - Publisher copyright policies & self-archiving

English | [Español](#) | [Magyar](#) | [P](#)

One journal found when searched for: **ieee transactions on signal processing**

<b>Journal:</b>	<a href="#">IEEE Transactions on Signal Processing</a> (ISSN: 1053-587X)
<b>RoMEO:</b>	This is a <b>RoMEO green</b> journal
<b>Paid OA:</b>	A paid open access option is <b>available</b> for this journal.
<b>Author's Pre-print:</b>	✓ author <b>can</b> archive pre-print (ie pre-refereeing)
<b>Author's Post-print:</b>	✓ author <b>can</b> archive post-print (ie final draft post-refereeing)
<b>Publisher's Version/PDF:</b>	✗ author <b>cannot</b> archive publisher's version/PDF
<b>General Conditions:</b>	<ul style="list-style-type: none"> <li>• Author's pre-print on Author's personal website, employers website or publicly accessible server</li> <li>• Author's post-print on Author's server or Institutional server</li> <li>• Author's pre-print must be removed upon publication of final version and replaced with either full citation to IEEE work with a Digital Object Identifier or link to article abstract in IEEE Xplore or replaced with Authors post-print</li> <li>• Author's pre-print must be accompanied with set-phrase, once submitted to IEEE for publication ("This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible")</li> <li>• Author's pre-print must be accompanied with set-phrase, when accepted by IEEE for publication ("(c) 20xx IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/ republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.")</li> <li>• IEEE must be informed as to the electronic address of the pre-print</li> <li>• If funding rules apply authors may post Author's post-print version in funder's designated repository</li> <li>• Author's Post-print - Publisher copyright and source must be acknowledged with citation (see above set statement)</li> <li>• Author's Post-print - Must link to publisher version with DOI</li> <li>• Publisher's version/PDF cannot be used</li> <li>• Publisher copyright and source must be acknowledged</li> </ul>
<b>Mandated OA:</b>	(Awaiting information)



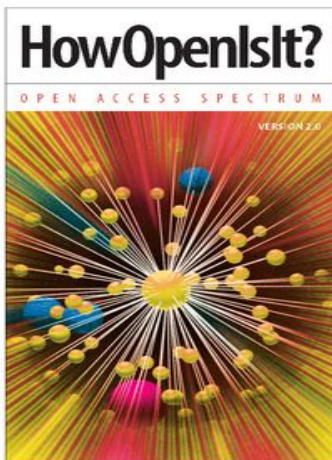
## Licencia general de IEEE relevada de su página

La licencia de IEEE, es bastante restrictiva porque es una cesión exclusiva y obviamente no permite la redistribución libre (al estilo CC). Sin embargo, permite el uso y distribución por parte del autor siempre que no se venda y además, en el punto 5 (segunda hoja) afirma que el autor es libre de cumplir cualquier imposición que le hagan sus organismos de financiamiento, lo que en el caso de Argentina implicaría el depósito en un repositorio con un embargo de 6 meses.

[http://www.ieee.org/publications\\_standards/publications/rights/index.html](http://www.ieee.org/publications_standards/publications/rights/index.html)



# “HowOpenIsIt?®” Open Access Spectrum (OAS)



The “[HowOpenIsIt?®](#)” Open Access Spectrum (OAS) guide standardizes Open Access terminology in an easily understandable, comprehensive resource created by PLOS, the Scholarly Publishing and Academic Resources Coalition ([SPARC](#)) and the Open Access Scholarly Publishers Association ([OASPA](#)). The guide defines core components of Open Access derived from the articulation of basic tenets in the 2002 Budapest Open Access Initiative ([BOAI](#)).

“[HowOpenIsIt?®](#)” OAS moves the conversation from “Is It Open Access?” to “[HowOpenIsIt?®](#)” and illustrates a nuanced continuum of more versus less open to enable users to compare and contrast publications and policies across a grid of clearly defined components related to readership, reuse, copyright, author and automatic posting, and machine readability.

The guide has been vetted and refined in a practical use pilot of 100 journals by PLOS, SPARC and [Copernicus Publications](#) in consultation with OASPA, Securing a Hybrid Environment for Research Preservation and Access ([SHERPA](#)) and Infrastructures for Open Access ([IS4OA](#)).

# Ayudas para conocer los derechos

Enlaces extraídos del portal español Acceso abierto.net, <http://www.accesoabierto.net/> pueden ser de interés para determinar cuáles son las condiciones en que algunas revistas permiten la reutilización de sus artículos y cómo afectan al posterior autoarchivo.

**Journal Info** <http://jinfo.lub.lu.se/jinfo?func=findJournals>

**OAK list** <http://www.oaklist.qut.edu.au/database/Basic.action>

Creative Commons License <http://creativecommons.org/license/>

Scholar's copyright project <http://www.sciencecommons.org/projects/publishing/>

Copyright toolbox <http://copyrighttoolbox.surf.nl/copyrighttoolbox/authors/licence>

SPARC Author Rights [www.arl.org/sparc/author/index.html](http://www.arl.org/sparc/author/index.html)

<http://scholars.sciencecommons.org/>

## Cómo insertar licencias CC en documentos de Office

**MS:** [http://wiki.creativecommons.org/Microsoft\\_Office\\_Addin](http://wiki.creativecommons.org/Microsoft_Office_Addin)

[http://roderic.uv.es/oa\\_es.html#nodo10](http://roderic.uv.es/oa_es.html#nodo10)



# Información sobre editores y políticas de permisos

[SHERPA-ROMEIO](#): Resumen de las políticas de más de 700 editores.

[DULCINEA](#): Permisos de revistas científicas españolas

[Blimunda](#): Permisos de revistas y editores científicos portugueses

[Lista de editores que permiten el depósito de sus PDF inmediatamente](#)



## Para mantener los derechos

Leer atentamente la información de la revista

Aunque la información de las páginas previas es de ayuda para una primera aproximación, **hay que leer la web del editor y de la revista en particular.**

Analizar las restricciones sobre el autoarchivo y el acuerdo real así como cualquier restricción que pone el editor sobre el lugar donde autoarchivar el trabajo: web personal, repositorio institucional, temático, etc.

Lo importante es que el autor analice la mejor opción para dar visibilidad a su obra y en tal sentido, proceder.

# ¿Dónde publicar?

Las revistas científicas especializadas más prestigiosas del mercado, con mayor índice de citas, son las elegidas o a las que apuntan los autores más prestigiosos (en general). En algunos casos puede suceder que la revista de interés no permita la clase de autoarchivo que sí interesa al autor, por lo que el autor tiene dos caminos:

**A)** Reconsiderar la elección y publicar en una revista de menos prestigio pero con una política más adecuada para el autoarchivo. La RAZÓN es que cualquier autor puede POR SÍ MISMO comprobar el incremento en el número de citas de su artículo cuando es publicado en una revista abierta o en un repositorio de acceso abierto.

**B)** Pedir al editor que acepte una modificación en cuanto a la política de autoarchivo, esto es lo que se llama ADENDUM o ADENDA.



# Cambios en el tiempo: Adendas

La iniciativa **SPARC** ofrece el **Sparc Author Addendum**:

<http://www.arl.org/sparc/author/addendum.shtml>

Esta adenda permite que el autor mantenga los derechos de reproducción de su obra, le brinda acceso gratuito al artículo y permite que redistribuya el material con una licencia de aceptación de trabajos derivados aunque no comercial. Consecuentemente, se reservan derechos suficientes como para realizar autoarchivo o para usar los materiales en clase o con colaboradores.



# Adenda en línea

## Scholars Copyright Addendum Engine

<http://scholars.sciencecommons.org/> de la propia **Science Commons** que, en base a los datos ingresados del autor, obra y del editor, genera una Adenda en PDF que se puede descargar. Hay cuatro tipos de adendas:

- 1- Acceso Reuso: Sparc Author Addendum
- 2- Acceso Inmediato: acceso inmediato a la version final
- 3- Acceso Retardado/diferido: acceso en diferido de 6 meses a la versión final
- 4- MIT Copyright Amendment: acceso inmediato, y con permiso para publicar en MIT Libraries y en PubMed

**Science Commons:** <http://sciencecommons.org/about/>

Science Commons tiene como misión lograr la apertura y el intercambio que han hecho de las licencias Creative Commons un éxito en las artes y las ciencias.



## Otras posibilidades para el autor

- El autor también puede optar por retener todos sus derechos y otorgar al editor aquellos que necesite para publicar su artículo. Si el editor no está de acuerdo, puede elegir otro que sí esté de acuerdo.
- Si el autor decide reservarse todos los derechos, es responsable de proteger la obra de las infracciones que puedan producirse, esta opción sólo es aconsejable para autores de reconocido prestigio o para quienes estén comprometidos con el acceso abierto.
- **Si no existe un contrato por escrito de cesión de derechos con el editor y el artículo se publica el autor conserva todos los derechos.**

# Entre el autor y otros: Lectores, usuarios, NO editores







## Una posibilidad: Licencia Creative Commons

El autor de la obra puede ceder a otros, no sólo editores, determinados derechos para hacer uso de la misma. Una forma posible es utilizar licencias de tipo CC, mediante las cuales se mantienen los derechos como autor, al mismo tiempo que se permite a otros ciertos usos siempre y cuando se reconozca la autoría correspondiente.



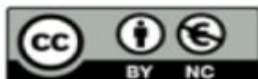
Las licencias **CC** se basan en 4 condiciones: una de ellas de carácter obligatorio denominada Reconocimiento, y 3 opcionales:

- 
-  **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.

Estas condiciones dan lugar a 6 licencias que parten de Reconocimiento y que luego se van cerrando con las otras alternativas. Las licencias son:



**Reconocimiento (by):** Se permite cualquier explotación de la obra, incluyendo una finalidad comercial, así como la creación de obras derivadas, la distribución de las cuales también está permitida sin ninguna restricción.



**Reconocimiento - NoComercial (by-nc):** Se permite la generación de obras derivadas siempre que no se haga un uso comercial. Tampoco se puede utilizar la obra original con finalidades comerciales.



**Reconocimiento - NoComercial - CompartirIgual (by-nc-sa):** No se permite un uso comercial de la obra original ni de las posibles obras derivadas, la distribución de las cuales se debe hacer con una licencia igual a la que regula



**Reconocimiento - NoComercial - SinObrasDerivadas (by-nc-nd):** No se permite un uso comercial de la obra original ni la generación de obras derivadas.

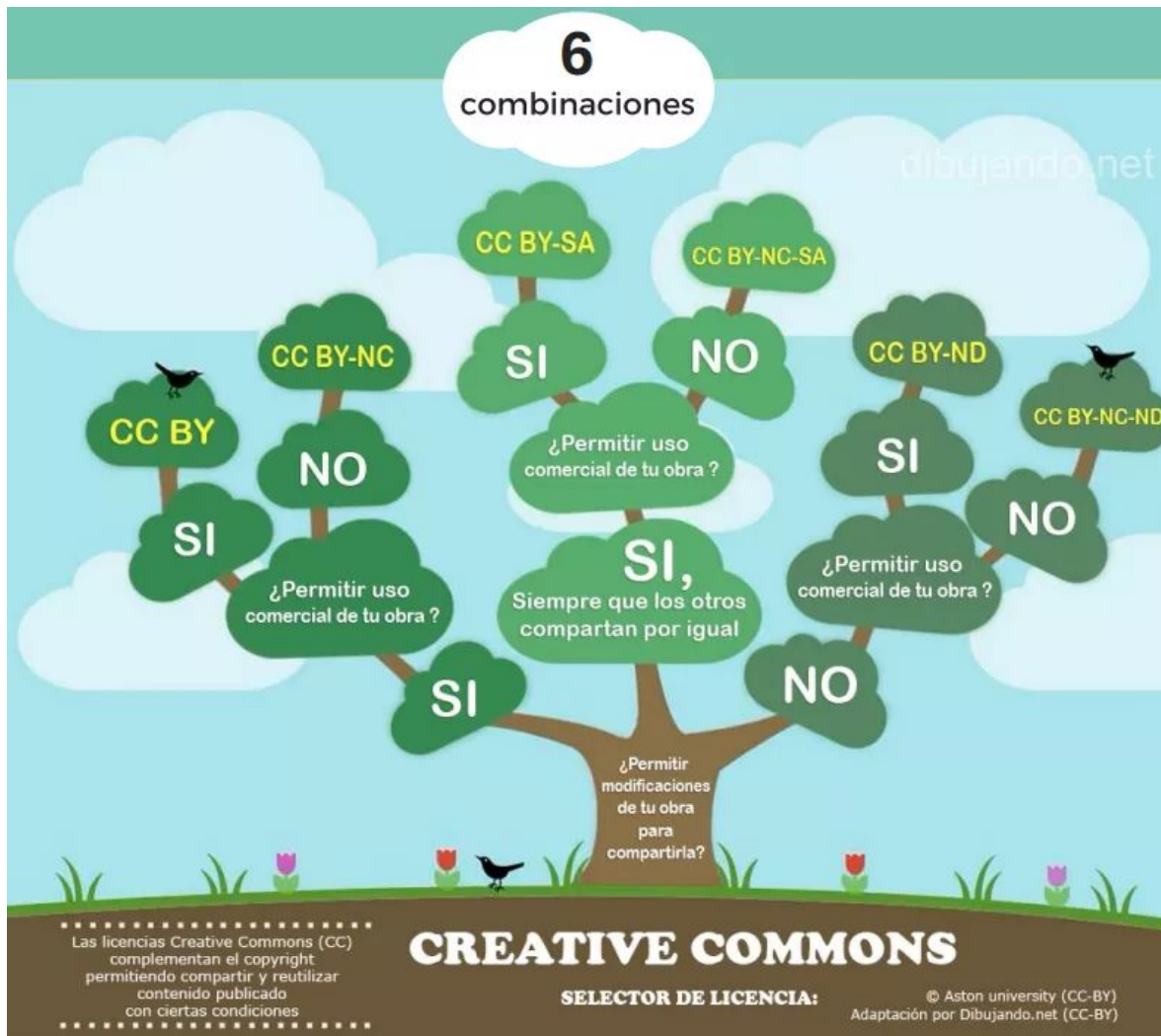


**Reconocimiento - CompartirIgual (by-sa):** Se permite el uso comercial de la obra y de las posibles obras derivadas, la distribución de las cuales se debe hacer con una licencia igual a la que regula la obra original.



**Reconocimiento - SinObrasDerivadas (by-nd):** Se permite el uso comercial de la obra pero no la generación de obras derivadas.





# El Servicio de Difusión de la Creación Intelectual - SEDICI

SEDICI

REPOSITORIO INSTITUCIONAL DE LA UNLP



Inicio
Buscar material
Subir material
Institucional
Preguntas frecuentes
Contacto





**Buscar material**

Busque entre los mas de **40000** recursos disponibles en el repositorio

**Recursos para su investigación**  
Acceda a más de 25.000 recursos científicos y académicos producidos dentro de la UNLP

**Preservación digital**  
Asegure el acceso a su obra a largo plazo, aún frente a los cambios tecnológicos

**Visibilidad de sus obras**  
Al publicar sus trabajos en SEDICI aumentará la visibilidad e impacto de sus obras

**Contacto con pares**  
Contactese con investigadores que se encuentran trabajando en áreas relacionadas a la suya

**Subir material**

Suba sus trabajos a SEDICI, para mejorar notoriamente su visibilidad e impacto



¿CÓMO CARGO MIS TRABAJOS?

**NOTICIAS**

- 31/03 | Novedades OA: el acceso abierto, hoy
- 26/03 | Cervantinos hasta los huesos
- 17/03 | Volver al futuro: la permanencia en el tiempo de los archivos digitales
- 13/03 | Linked Data, recomendaciones de la W3C
- 10/03 | Visita SEDICI al Vicepresidente Institucional Fernando Tauber
- 03/03 | Humanidades Digitales: Sócrates en la nube
- 27/02 | El impacto por venir
- 19/02 | La NASA pondrá en abierto los datos y productos de sus investigaciones
- 09/02 | Revisión por pares: debates para el porvenir
- 03/02 | SEDICI entró en el top 50 del ranking mundial de repositorios
- 13/01 | Novedades OA: el acceso abierto, hoy
- 09/01 | Santiago Sarandón: "Hay modelos de hacer agricultura que son insumodependientes y no todo el mundo tiene

**DOCUMENTOS DESTACADOS**

Implementación del dispositivo Punto de Encuentro familiar en Mar del Plata  
*Mimozzi, Mercedes;*

Aplicaciones aritméticas usando lógica programable  
*Jaquezod, Guillermo A.; De Gluzi, María Raquel;* ...

Los múltiples Sendero Luminoso en el actual Perú  
*Daz, Fernanda Daniela;*

Efectividad en colmenas de mezclas de aceites esenciales sobre loque americana  
*Abo, Graciela N.; Cedrele, Elia;* ...

El estudio de la Seguridad Internacional: ampliación y profundización del debate en torno a la nueva agenda de Seguridad  
*Tibera, Juan Cruz;*

**Iniciar sesión**

**Navegue por nuestras colecciones**

-  **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
-  **Radio Universidad**  
Entrevistas y producciones artísticas, entre otros audios
-  **Recursos educativos abiertos**  
Material educativo de cursos dictados en la UNLP
-  **Emergencia hídrica**  
Trabajos dentro del Plan de Gestión Integrada de Riesgos de Desastres

# Autoarchivo en SEDICI

SEDICI

REPOSITORIO INSTITUCIONAL DE LA UNLP


UNLP

Inicio
Buscar material
Subir material
Institucional
Preguntas frecuentes
Contacto





**Buscar material**

Busque entre los mas de **40000** recursos disponibles en el repositorio

Q

**Subir material**

Suba sus trabajos a SEDICI, para mejorar notoriamente su visibilidad e impacto



¿CÓMO CARGO MIS TRABAJOS?



**Analia Pinto** Salir

Editar perfil

Mis documentos

## Carga de documentos

### Cargas Pendientes

Envíos pendientes. Podría **Nuevo envío**.

Título	Colección	Remitente
No hay envíos incompletos.		

### Tareas del flujo de trabajo

Estas tareas son ítems que están esperando aprobación antes de ser añadidas al repositorio. Hay dos colas de tareas, una para tareas que ha aceptado y otra para tareas que todavía no han sido asumidas por nadie.

**Sus tareas**

Tarea	Ítem	Colección	Remitente
<input type="checkbox"/>	Revisión	Guía de buenas prácticas para la	Libros
			Analia



# ¿Qué hay que verificar cuando se quiere depositar en un Repositorio?

## 1) Quién posee los derechos de explotación de la obra:

Si el autor es el titular, entonces no debe buscar ningún permiso para depositar el trabajo en Repositorio.

Si se va a subir un trabajo previamente publicado (por ejemplo, un artículo, un libro, una comunicación de un congreso) y el autor ha conservado los derechos sobre todo el contenido y ha firmado simplemente una licencia de publicación con el editor, es necesario comprobar los términos del acuerdo, aunque probablemente será posible realizar el depósito en el Repositorio, a menos que se haya firmado una licencia “exclusiva”.



## ¿Qué hay que verificar cuando se quiere depositar en un Repositorio?

Si, por el contrario, se ha realizado una cesión exclusiva de los derechos a un editor hay que comprobar los permisos de depósito que dan las editoriales a los repositorios institucionales. Para los artículos, estos permisos se refieren a si se puede subir el texto completo de la publicación en un repositorio abierto y/o en la web personal del autor y si es así qué versión de la obra (preprint, post-print de autor o post-print de editor) y en qué condiciones (embargos, modo de citarla, etc).

# ¿Qué hay que verificar cuando se quiere depositar en un Repositorio?

- En el caso de que no esté autorizada la subida del **PDF o post-print del editor**,
  - intentar localizar la versión del texto que sí puede subirse en el repositorio: el **pre-print** del texto (la primera versión del trabajo que un autor envía al editor)
  - y/o el **post-print de autor** (la versión del texto que ha incorporado las sugerencias del comité de pares y ha sido aprobado para su publicación).
- Ver si el editor ha aplicado un **periodos de embargo** antes de poder subir una versión del trabajo en un repositorio.



## ¿Qué hay que verificar cuando se quiere depositar en un Repositorio?

- Cuando los permisos de una editorial no aparezcan en ninguna de las bases de datos presentes en “Enlaces de interés” o la información no sea clara, es aconsejable visitar la web del editor y buscar la información bajo los epígrafes “Authors”, “FAQ”, “Permissions”, “Policies” o similar.
- Como último recurso contactar al editor: el autor, la biblioteca y/o el repositorio.

## 2) **¿Está sujeto a copyright el material científico y divulgativo depositado en el repositorio?**

Todo el material depositado en SEDICI está sujeto a copyright, tal y como indica la licencia de Digital.CSIC. Cuando un autor archiva un trabajo suyo en SEDICI confirma que posee los derechos o que ha sido autorizado por el titular de los derechos para depositarlo en el repositorio.

Mediante la Licencia de SEDICI está simplemente autorizando a que el repositorio difunda en modo no exclusivo, almacene y preserve el objeto digital depositado.



### Buscar material

Busque entre los mas de **43000** recursos disponibles en el repositorio

### Subir material

Suba sus trabajos a SEDICI, para mejorar notoriamente su visibilidad e impacto



## Políticas del repositorio

SEDICI es el Repositorio Institucional de la Universidad Nacional de La Plata creado en el año 2003 para constituirse como portal de acceso central que reúna toda la producción digital generada en el ámbito de la Universidad y para dar a sus creaciones una mayor visibilidad, aumentar su uso e impacto y asegurar su acceso y preservación a largo plazo. El repositorio SEDICI adhiere a la Iniciativa de Acceso Abierto que sustenta el principio de la libre disposición de la información. A continuación se exponen las políticas del repositorio, las cuales son sometidas regularmente a revisión.

- Política de contenidos
- Política de acceso a datos
- Política de depósito
- Política de preservación
- Política de Metadatos
- Política de Servicios

### Política de contenidos

SEDICI recibe la producción de toda la UNLP en sus actividades de enseñanza, investigación y extensión. Los materiales a incorporar en el repositorio deben cumplir con al menos uno de los siguientes requisitos:





# Versiones de los artículos científicos y acceso abierto



**Draft / Discussion Paper**  
(Borrador de trabajo)



**Preprint / Submitted Version**  
(Versión enviada al editor)



**Revisión por pares**



**Postprint / Accepted Manuscript**  
(Versión final de los autores que incluye los cambios propuestos por los revisores)  
También llamada Author's final version



**Published Version**  
(Versión final publicada por el editor)



**Consultar** la política de la revista/editorial para comprobar qué versión se puede depositar en un repositorio



**Investigador:** conserva todas las versiones del ciclo de vida de tus documentos



Las grandes editoriales científicas suelen permitir depositar esta **versión en repositorios**



**Versiones aptas para cumplir** los requerimientos de la Ley de la Ciencia y Horizonte 2020





## Beneficios de publicar en los repositorios institucionales

### ¿Qué son los repositorios institucionales?

Son sistemas en línea de acceso abierto que tienen el objetivo de almacenar, preservar y difundir la producción científica y académica de la comunidad universitaria.



#### → Objetivos

- Difusión
- Visibilidad
- Impacto
- Preservación

### ¿Qué beneficios proporciona un repositorio?



#### al investigador



- Permite publicar en abierto los documentos resultado de investigación y así cumplir los mandatos (de organismos financiadores).
- **Más visibilidad e impacto**, más citas.
- Garantiza una **correcta gestión de los derechos de autor**.
- **Acceso perpetuo a los trabajos** mediante enlaces permanentes.
- **Datos de uso** que facilitan la participación en evaluaciones.
- Permite el depósito de **todo tipo de documentos**, incluso inéditos (artículos, monografías, capítulos de monografías, comunicaciones en eventos, tesis, trabajos académicos, datasets, videos, etc.) **y en todo tipo de formatos**.



#### a la universidad



- **Reunir y difundir** al mundo la producción científica y académica de la institución.
- **Aumento de la visibilidad** de la institución a través de las obras de sus autores y mejora del posicionamiento en Google.
- **Preservación para el futuro** de la obra de los autores y de la actividad intelectual de la universidad.



#### a la sociedad



- **Acceso al conocimiento** para toda la sociedad y reutilización en beneficio de todos.
- Permite **visibilizar** y rendir cuentas de la inversión pública realizada en investigación.
- **Disminuye la brecha de acceso a la información** entre las instituciones y países.



# El acceso abierto: consecuencias

- 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, cuestionando 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...).
  - Abre las puertas a estándares internacionales abiertos y nuevos modelos de medición que superen el modelo dominante.



# Visibilidad e Impacto

Métricas de visibilidad. Herramientas y servicios



Discovery Space Shuttle  
Fuente: NASA.gov



# Visibilidad Web

Un repositorio institucional preserva la producción de una institución

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

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

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



# Visibilidad Web

Abarca gran variedad de contenidos:

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



# Visibilidad Web

Involucra múltiples actores:

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



# Visibilidad Web

A considerar:

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

# Visibilidad Web

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

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

\*\* visibilidad, accesibilidad, facilidad de encontrar



# Visibilidad Web

En su investigación, A. Swan realiza ciertas consideraciones:

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



# Visibilidad Web

De lo anterior se desprende que:

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

# Visibilidad Web

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

Detectó una importante variabilidad por disciplina:

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



# Visibilidad Web

Estrategia: entender qué y cómo se expone

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



# Visibilidad Web

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

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



# Visibilidad Web

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

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

Ej: crawler4j, Nutch, Ankhor, Webphinx



The image shows the Webphinx Crawler interface on the left and its log output on the right. The interface includes a 'Crawl' section with 'the Web' selected, 'all links' as the crawler, and 'http://www.unlp.edu.ar' as the starting URL. The depth is set to 5 hops using Breadth first. The 'Graph' section shows a network diagram of subsites and connections. The log on the right displays a series of actions performed by the crawler, such as downloading and visiting various pages on the UNLP website.

```
Wed Aug 13 08:36:36 ART 2014: downloaded Escuela Anexa [http://www.unlp.edu.ar/escuela_graduada]
Wed Aug 13 08:36:36 ART 2014: visited Escuela Anexa [http://www.unlp.edu.ar/escuela_graduada]
Wed Aug 13 08:36:36 ART 2014: retrieving Escuela de Agricultura y Ganaderia [http://www.unlp.edu.ar/escu
y_ganaderia]
Wed Aug 13 08:36:36 ART 2014: downloaded Colegio Nacional [http://www.unlp.edu.ar/colegio_nacional]
Wed Aug 13 08:36:36 ART 2014: visited Colegio Nacional [http://www.unlp.edu.ar/colegio_nacional]
Wed Aug 13 08:36:36 ART 2014: retrieving Dirección [http://www.unlp.edu.ar/direccion_ead]
Wed Aug 13 08:36:37 ART 2014: downloaded Liceo Victor Mercante [http://www.unlp.edu.ar/liceo]
Wed Aug 13 08:36:37 ART 2014: visited Liceo Victor Mercante [http://www.unlp.edu.ar/liceo]
Wed Aug 13 08:36:37 ART 2014: retrieving Áreas EAD [http://www.unlp.edu.ar/areas_ead]
Wed Aug 13 08:36:37 ART 2014: downloaded Bachillerato de Bellas Artes [http://www.unlp.edu.ar/bachillera
Wed Aug 13 08:36:37 ART 2014: visited Bachillerato de Bellas Artes [http://www.unlp.edu.ar/bachillerato
Wed Aug 13 08:36:37 ART 2014: retrieving Historia [http://www.unlp.edu.ar/historiaead]
Wed Aug 13 08:36:37 ART 2014: downloaded Escuela de Agricultura y Ganaderia [http://www.unlp.edu.ar/escu
y_ganaderia]
Wed Aug 13 08:36:37 ART 2014: visited Escuela de Agricultura y Ganaderia [http://www.unlp.edu.ar/escuela
anaderia]
Wed Aug 13 08:36:37 ART 2014: retrieving Comisión y Unidad de Gestión [http://www.unlp.edu.ar/comision_y
Wed Aug 13 08:36:38 ART 2014: downloaded Dirección [http://www.unlp.edu.ar/direccion_ead]
Wed Aug 13 08:36:38 ART 2014: visited Dirección [http://www.unlp.edu.ar/direccion_ead]
Wed Aug 13 08:36:38 ART 2014: retrieving Metodología [http://www.unlp.edu.ar/metodologiaead]
Wed Aug 13 08:36:39 ART 2014: downloaded Áreas EAD [http://www.unlp.edu.ar/areas_ead]
Wed Aug 13 08:36:39 ART 2014: visited Áreas EAD [http://www.unlp.edu.ar/areas_ead]
Wed Aug 13 08:36:39 ART 2014: retrieving Publicaciones [http://www.unlp.edu.ar/publicaciones]
Wed Aug 13 08:36:39 ART 2014: downloaded Historia [http://www.unlp.edu.ar/historiaead]
Wed Aug 13 08:36:39 ART 2014: visited Historia [http://www.unlp.edu.ar/historiaead]
Wed Aug 13 08:36:39 ART 2014: retrieving Entorno Virtual [http://www.unlp.edu.ar/entorno_virtual_ead]
Wed Aug 13 08:36:39 ART 2014: downloaded Comisión y Unidad de Gestión [http://www.unlp.edu.ar/comision_y
Wed Aug 13 08:36:39 ART 2014: visited Comisión y Unidad de Gestión [http://www.unlp.edu.ar/comision_y
Wed Aug 13 08:36:39 ART 2014: retrieving Capacitación [http://www.unlp.edu.ar/capacitacion_docente_ead]
Wed Aug 13 08:36:40 ART 2014: downloaded Metodología [http://www.unlp.edu.ar/metodologiaead]
Wed Aug 13 08:36:40 ART 2014: visited Metodología [http://www.unlp.edu.ar/metodologiaead]
Wed Aug 13 08:36:40 ART 2014: retrieving Proyectos EAD [http://www.unlp.edu.ar/proyectosead]
```

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

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

# Visibilidad Web

Dónde impacta la visibilidad web?

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



# Visibilidad Web - Webometrics

Ranking elaborado por el CISC

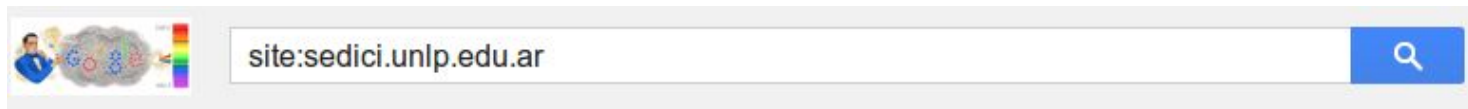
El Ranking de repositorios analiza diversos indicadores:

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



# Visibilidad Web - Webometrics

## Size: el tamaño sí importa



**Web** Images News More ▾ Search tools

About 745,000 results (0.50 seconds)

Google promotion

**Try Google Webmaster Tools**

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

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

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

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

Repositorio institucional de la Universidad Nacional de La Plata.

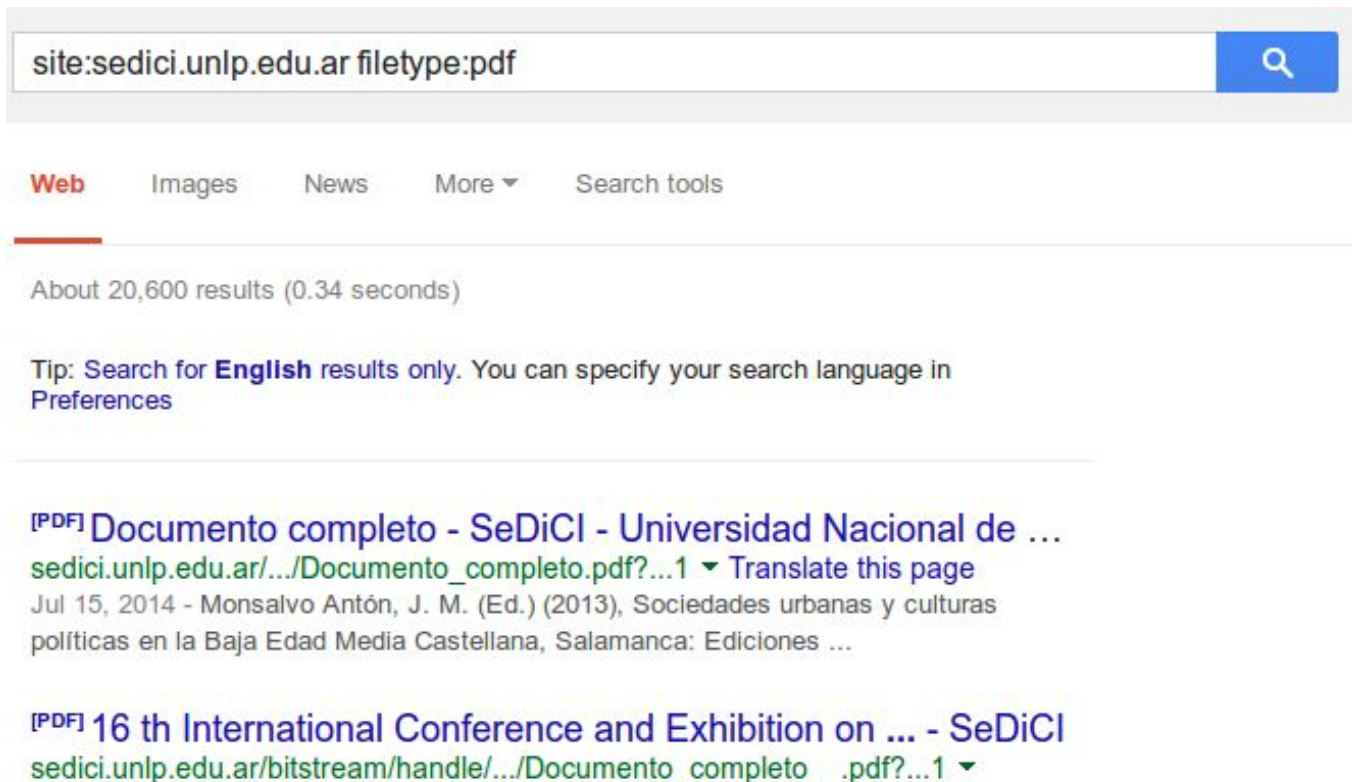
**SEDICI - Buscar**

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



# Visibilidad Web - Webometrics

## Rich files



The screenshot shows a Google search interface. The search bar contains the query 'site:sedici.unlp.edu.ar filetype:pdf'. Below the search bar, there are navigation tabs for 'Web', 'Images', 'News', 'More', and 'Search tools'. The 'Web' tab is selected and highlighted with a red underline. Below the tabs, it indicates 'About 20,600 results (0.34 seconds)'. A tip suggests searching for English results only. Two search results are visible, both marked as PDFs. The first result is 'Documento completo - SeDiCI - Universidad Nacional de ...' with a URL 'sedici.unlp.edu.ar/.../Documento\_completo.pdf?...1'. The second result is '16 th International Conference and Exhibition on ... - SeDiCI' with a URL 'sedici.unlp.edu.ar/bitstream/handle/.../Documento\_completo\_\_pdf?...1'.

site:sedici.unlp.edu.ar filetype:pdf

**Web** Images News More Search tools

About 20,600 results (0.34 seconds)

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

[PDF] Documento completo - SeDiCI - Universidad Nacional de ...  
sedici.unlp.edu.ar/.../Documento\_completo.pdf?...1 ▾ Translate this page  
Jul 15, 2014 - Monsalvo Antón, J. M. (Ed.) (2013), Sociedades urbanas y culturas políticas en la Baja Edad Media Castellana, Salamanca: Ediciones ...

[PDF] 16 th International Conference and Exhibition on ... - SeDiCI  
sedici.unlp.edu.ar/bitstream/handle/.../Documento\_completo\_\_pdf?...1 ▾



# Visibilidad Web - Webometrics

Visibility: enlaces entrantes

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

Priorizamos enlaces externos o tamaño del dominio?





SITE EXPLORER

SEO REPORT

BACKLINKS REPORT

APPS

LABS

PLANS & PRICING

OVERVIEW

Dashboard

Top Pages

INBOUND LINKS

Links

New / Lost

Anchors

Referring Domains

New / Lost

Referring IPs

OUTGOING LINKS

Linked Domains

Anchors

EXPORT

CSV

PDF

URL Rank

38

Ahrefs Domain Rank

65

Backlinks

41 K

Referring Domains

522



Referring Pages 27,984

Total Backlinks 41,298

Crawled Pages 133,810

Referring IPs 432

Referring Subnets 390

Referring Domains 522

Governmental 8

Educational 64

.com 149

.net 36

.org 82

### Referring Pages

JUN JUL AUG ALL ONE YEAR LAST 30 DAYS



### Backlink Types:

text 27,817

dofollow 26,096

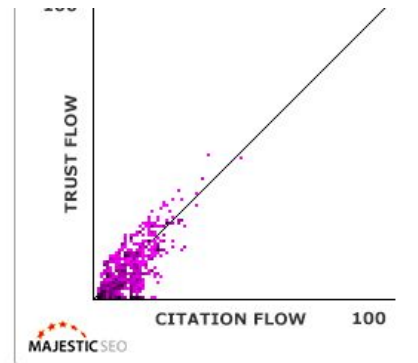
Backlinks



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

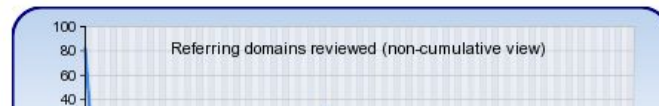
backlinks	Domains	IPS	Subnets
<b>45,537</b>	<b>457</b>	<b>396</b>	<b>374</b>
In the last 5 years			
<b>106,276</b>	<b>1,540</b>	<b>938</b>	<b>793</b>

<b>Homepage Title</b>	<b>SEDICI - Repository of National University of La Plata</b>
<b>Educational Ref. Backlinks</b>	<b>2,073</b>
<b>Governmental Ref. Backlinks</b>	<b>111</b>
<b>Educational Ref. Domains</b>	<b>74</b>
<b>Governmental Ref. Domains</b>	<b>13</b>
<b>Indexed URLs</b>	<b>374,512</b>
<b>Position in Majestic Million</b>	This site is NOT in the Majestic Million.



External Backlinks Link Profile

### Backlink History



# Visibilidad Web - Herramientas

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

Como todo en internet, hay muchas herramientas más.



# Visibilidad Web - Herramientas

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

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



## Search Queries

Top queries

Top pages

Filters

Web X

Jul 13, 2014

to

Aug 12, 2014

Queries

15,794

Impressions

1,197,467

-24%

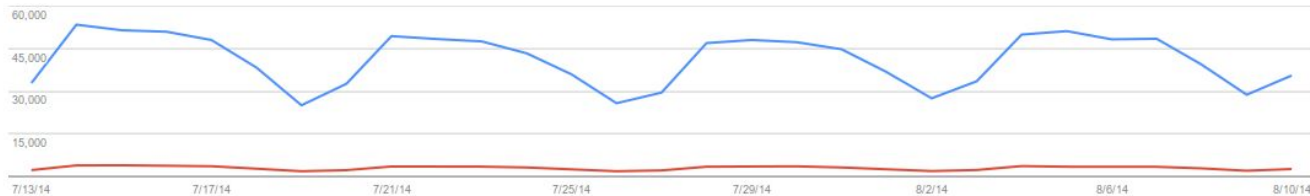
Displaying 80,818

Clicks

86,311

-28%

Displaying 10,200



Download this table

Download chart data

Basic

With change

Show

25 rows

1-25 of 15,794

&lt;

&gt;

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

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



## Links to Your Site

Total links  
**13,441,617**

### Who links the most

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

### How your data is linked

<a href="#">acceder</a>
<a href="#">sedici servicio de difusión de la creación intelectual</a>
<a href="#">libros de cátedra</a>
<a href="#">haga click aquí para acceder al objeto digital</a>
<a href="#">eventos</a>
<a href="#">More »</a>

### Your most linked content

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

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

# Visibilidad Web - Herramientas

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

Google tiene su servicio Google Analytics, gratuito y muy completo

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

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



# Visibilidad Web - Herramientas

Piwik

Dashboard | [Alle Webseiten](#) | [Widgets](#) | [API](#) | [E-Mail-Berichte](#) | [Feedback geben!](#) | Deutsch ▾ | Hallo, anonymous! | [Anmelden](#)

Übersicht | Besucher | Aktionen | Verweise | Ziele | Webseite: Piwik Forums ▾

Zeitspanne: 2011-04-07 | [Widget hinzufügen...](#)

Sie betrachten gerade die Demo von Piwik

### Graph der letzten Besuche

1,190 Besuche

30-60s 4-6 min 2-4 min 11-15 min 1-2 min 8-11 min 15+ min 0-30s  
8-8 min

### Besuchslänge

### Besucher-Browser

### Beste Suchmaschinen

Suchmaschine	Besuche
Google	436
Ask	5
Ecosia	1
Bing	1

1-4 von 4

### Liste der Suchbegriffe

Stichwort	Besuche
piwik forum	7
child xxx	6
piwik	5
piwik performance	5
evaluation awstats piwik	4
piwik geolip	3
piwik datenschutz	3
piwik smtp	2
piwik apache log import	2
piwik bounce rate	2

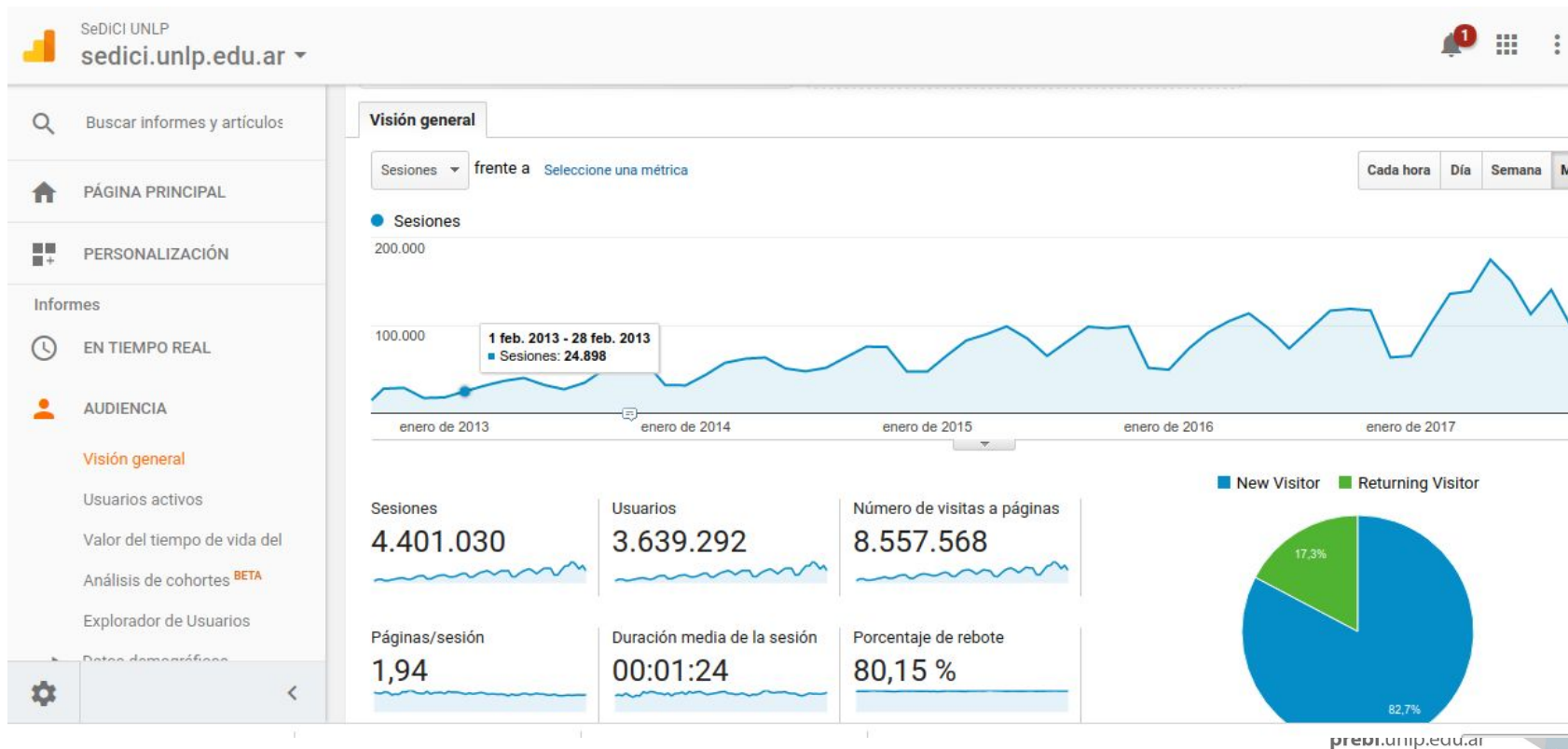
1-10 von 402 nächste >

### Besucherkänder (Weltkarte)

### Besuche nach Server-Zeit

70 Besuche

# Visibilidad Web - Redes sociales



# Visibilidad Web - Herramientas

Las herramientas de análisis de tráfico web externas (PIWIK, Google Analytics) e internas (análisis de registros de acceso) no conocen acerca de la **estructura lógica** del repositorio

Muchas veces necesitamos agrupar métricas (ej. enlaces de una comunidad) y unir métricas de distintas fuentes (descargas de un libro desde el repositorio y desde el portal de libros)



# Visibilidad Web - Redes sociales

- Permiten alcanzar un público inmediato muy amplio (amigos, followers, groups, pages)
- Permiten alcanzar un público mucho más amplio a partir de re-comparticiones (retwits, shares, +1)

Pero... qué compartimos y dónde?

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

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



# Visibilidad Web - Redes sociales

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

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



# Visibilidad Web - Redes sociales

Aprovechar eventos circunstanciales, efemérides, noticias

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



# Visibilidad Web - Redes sociales

Facebook navigation bar: Sedici UNLP, search, profile picture, Gonzalo, Home, notifications, settings, help.

Page navigation: Page, Inbox, Notifications (18), Insights, Publishing Tools.

Like, Following, Share, ...

**Sedici UNLP**  
Published by Analía Pinto [?] · September 18 at 10:28am ·

**Trabajos de digitalización**  
Gracias a la colaboración de la Biblioteca de la Facultad de Informática (UNLP) estamos digitalizando una serie de tesis de grado y posgrado cuyos registros permanecían sin datos hasta la fecha. Les dejamos una lista de las últimas incorporaciones, ahora sí, a texto completo:

- <http://hdl.handle.net/10915/4122>
- <http://hdl.handle.net/10915/3937>
- <http://hdl.handle.net/10915/3935>
- <http://hdl.handle.net/10915/4111>
- <http://hdl.handle.net/10915/3890>
- <http://hdl.handle.net/10915/3921>
- <http://hdl.handle.net/10915/3862>

**Bibliotecas universitarias y educación a distancia**  
En este trabajo se intenta presentar una revisión del aporte que las bibliotecas universitarias pueden hacer a la educación superior a distancia en un entorno de e-learning. Una biblioteca es una...

SEDICI.UNLP.EDU.AR

869 people reached Boost Post

Like Comment Share

**Page Summary** Last 7 days + Export Data

Results from Sep 13, 2017 - Sep 19, 2017  
Note: Does not include today's data. Insights activity is reported in the Pacific time zone. Ads activity is reported in the time zone of your ad account.

Organic Paid

<p><b>Actions on Page</b> September 12 - September 18</p> <p><b>2</b> Total Actions on Page <span>▲100%</span></p>	<p><b>Page Views</b> September 12 - September 18</p> <p><b>46</b> Total Page Views <span>▲39%</span></p>	<p><b>Page Previews</b> September 12 - September 18</p> <p><b>8</b> Page Previews <span>▼11%</span></p>
<p><b>Page Likes</b> September 12 - September 18</p> <p><b>16</b> Page Likes <span>▲46%</span></p>	<p><b>Reach</b> September 12 - September 18</p> <p><b>2,322</b> People Reached <span>▼16%</span></p>	<p><b>Post Engagements</b> September 12 - September 18</p> <p><b>393</b> Post Engagement <span>▲18%</span></p>
<p><b>Videos</b> September 12 - September 18</p> <p><b>40</b></p>	<p><b>Page Followers</b> September 12 - September 18</p> <p><b>40</b></p>	

sedici.unlp.edu.ar  
prebi.unlp.edu.ar

# Visibilidad Web - Wikipedia

Wikipedia es un excelente lugar para difundir trabajos científicos y académicos

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

Permite ofertar contenido propio en un espacio muy visible:

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



# Visibilidad Web - Wikipedia

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

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

- Using Wikipedia to Enhance the Visibility of Digitized Archival Assets, Michael Szajewski <http://www.dlib.org/dlib/march13/szajewski/03szajewski.html>
- Wikipedia and institutional repositories: an academic symbiosis? Alastair G. Smith, [http://researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/1848/Conference\\_paper.pdf?sequence=1](http://researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/1848/Conference_paper.pdf?sequence=1)
- Increasing the Discoverability of Digital Collections Using Wikipedia - A Review with deep analysis, Ed Galloway, Cassandra DellaCorte <http://palrap.pitt.edu/ojs/index.php/palrap/article/view/60>
- Wikipedia Lover, Not a Hater: Harnessing Wikipedia to Increase the Discoverability of Library Resources, Danielle Elder<sup>a</sup>, R. Niccole Westbrook<sup>b</sup> & Michele Reilly<sup>b</sup> <http://www.tandfonline.com/doi/abs/10.1080/19322909.2012.641806#preview>

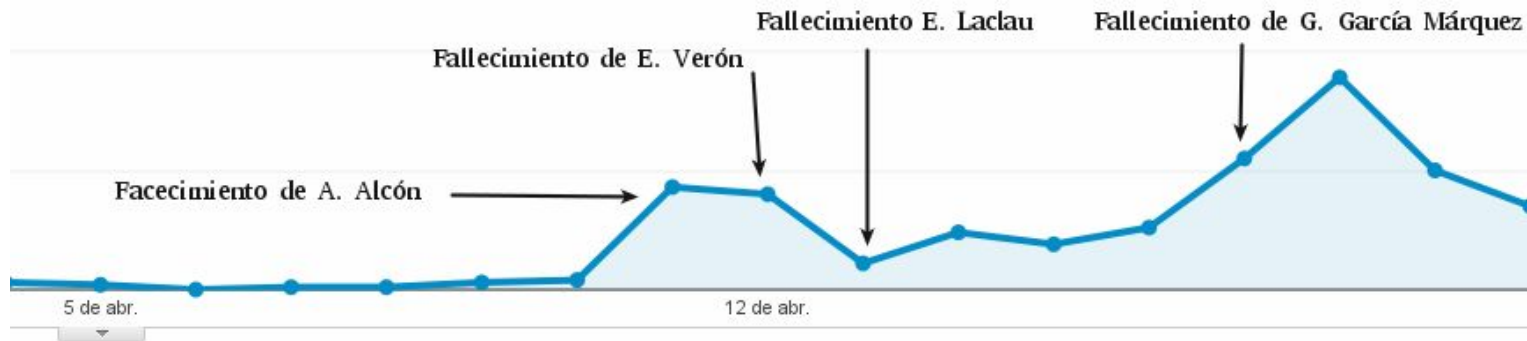


# Visibilidad Web - Wikipedia

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

- Muchas ediciones se vincularon con fallecimientos recientes (Eliseo Verón, Alfredo Alcón, Ernesto Laclau, Gabriel García Márquez)
- Otras surgían a partir de tesis y publicaciones que se depositaban en el repositorio











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

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

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

**Derecha:** Accesos a SEDICI desde sitios externos entre enero y julio de 2014. La versión en español de Wikipedia ascendió a la posición 3 en 7 meses (en realidad 4: abril-julio). Fuente Google Analytics.

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

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



# Visibilidad Web e Interoperabilidad

Mediante interoperabilidad (OpenSearch, RSS/Atom, webservices, OAI), podemos exponer nuestros materiales en múltiples espacios. Ejemplos:

- blogs institucionales (que se comunican con las redes sociales)
- programas institucionales: educación a distancia, radio, editorial
- sitios de laboratorios\*, departamentos, cátedras
- bases de datos, agregadores e índices internacionales

\*Orduña-Malea, Enrique, Cabezas-Clavijo, Álvaro, Delgado López-Cózar, Emilio. Visibilidad e impacto web de los grupos de investigación de información y documentación en las universidades públicas españolas. 2013.

<http://hdl.handle.net/10481/31303> .



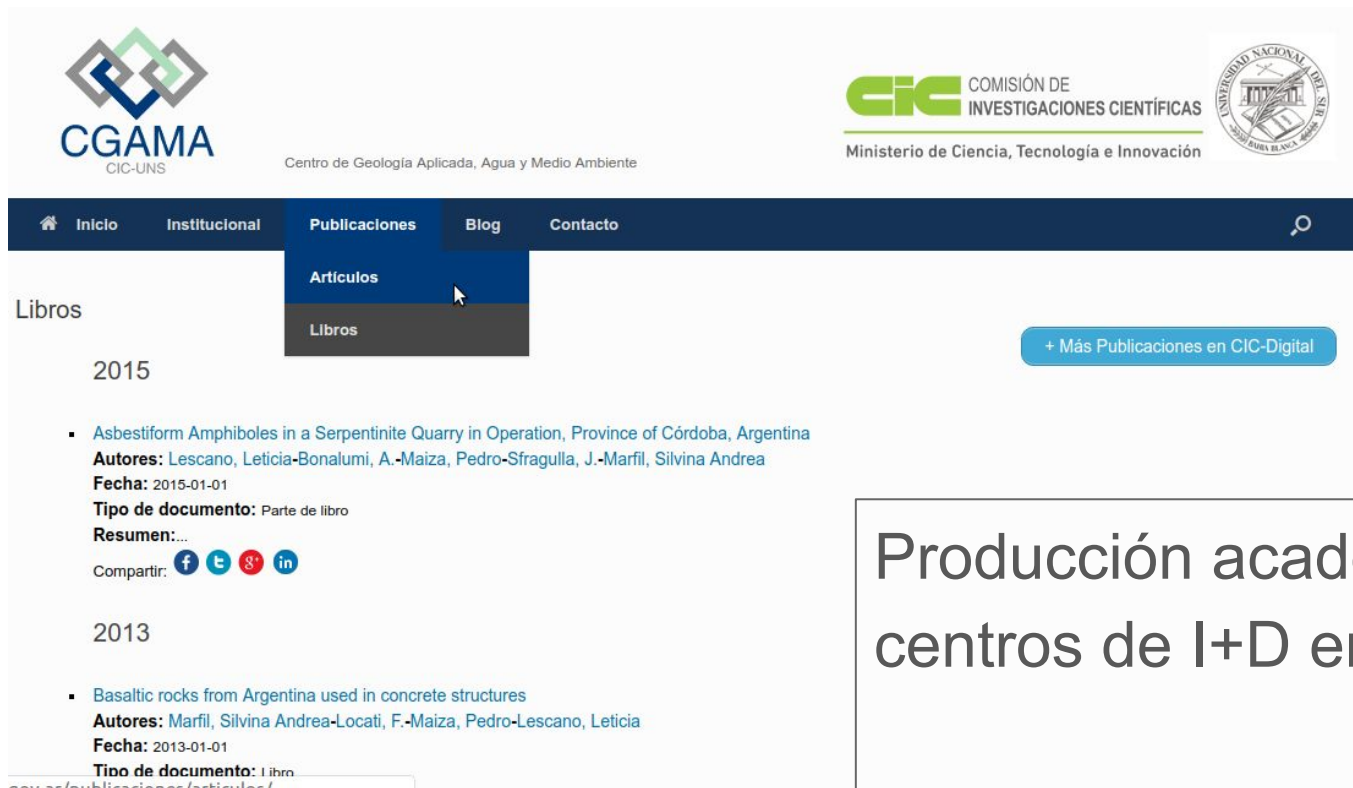
# Visibilidad Web

Algunos artículos de interés en la web:

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



# Identidad digital



**CGAMA**  
CIC-UNS  
Centro de Geología Aplicada, Agua y Medio Ambiente

**CIC** COMISIÓN DE INVESTIGACIONES CIENTÍFICAS  
Ministerio de Ciencia, Tecnología e Innovación

UNIVERSIDAD NACIONAL DEL SUR  
BARCELONA, 1900




Inicio Institucional **Publicaciones** Blog Contacto

Artículos Libros

Libros

+ Más Publicaciones en CIC-Digital

2015

- Asbestiform Amphiboles in a Serpentine Quarry in Operation, Province of Córdoba, Argentina  
**Autores:** Lescano, Leticia-Bonalumi, A.-Maiza, Pedro-Sfragulla, J.-Marfil, Silvina Andrea  
**Fecha:** 2015-01-01  
**Tipo de documento:** Parte de libro  
**Resumen:** ...  
Compartir:    

2013

- Basaltic rocks from Argentina used in concrete structures  
**Autores:** Marfil, Silvina Andrea-Locati, F.-Maiza, Pedro-Lescano, Leticia  
**Fecha:** 2013-01-01  
**Tipo de documento:** Libro

nov.ar/publicaciones/articulos/

Producción académica de los centros de I+D en la web

# Identidad digital

## Perfiles personales en sitio web institucional


Instituto Argentino de Radioastronomía


COMISIÓN DE INVESTIGACIONES CIENTÍFICAS  
Ministerio de Ciencia, Tecnología e Innovación

[Inicio](#) [Institucional](#) [Actividades](#) [Biblioteca](#) [Divulgación](#) [Novedades](#) [¿Cómo llegar?](#) [Contacto](#)

### Dr. E. Marcelo ARNAL

**Tema de Investigación:**

- Morfología del medio interestelar
- Director del Proyecto LLAMA
- Integrante del grupo EsFEGa

**Docencia:**

- Profesor Titular, "Astrofísica de la emisión en el continuo de radio"
- Profesor Titular, "Técnicas Observacionales Radioastronómicas"
- Colaborador, "Espectroscopia atómica y molecular del medio interestelar"

**Contacto:**

- ✉ arnal (at) iar.unlp.edu.ar
-  [LinkedIn](#)
- 
-  SAO/NASA Astrophysics Data System (ADS)

Publicaciones en el repositorio CONICET Digital.

- The H I supershell GS 118+01-44 and its role in the interstellar medium  
**Autores:** Suad, Laura Andrea-Cichowski, Silvina-Noriega Crespo, A.-Arnal, Edmundo Marcelo-Testori, Juan Carlos-Flagey, N.  
**Fecha:** 2016-01-01  
 Compartir:    
- The infrared and molecular environment surrounding the Wolf-Rayet star WR 130  
**Autores:** Cichowski, Silvina-Suad, Laura Andrea-Pineault, S.-Noriega Crespo, A.-Arnal, Edmundo Marcelo-Flagey, N.



- Identificación
- Finalidad
- Breve Reseña Histórica
- El IAR en la actualidad
- Imágenes
- Personal
- ¿Cómo llegar?

**Buscar**



**CONICET DIGITAL**  
REPOSITORIO INSTITUCIONAL

**ÚLTIMAS NOTICIAS EN LOS MEDIOS**



Avanzan en el armado de un telescopio para estudiar el origen del universo – Diario de Cultura (28-06-2017)



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

# Perfiles de investigadores



**Jose Texier**

ORCID:0000-0003-0176-6625

Ciencias de la Computación, Compiladores, Base de Datos, Bibliotecas Digitales, Repositorios Institucionales

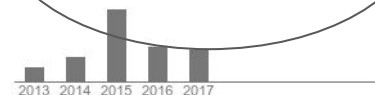
Dirección de correo verificada de unet.edu.ve - [Página principal](#)

Seguir

Título	Citado por	Año
<b>Los repositorios institucionales y las bibliotecas digitales: una somera revisión bibliográfica y su relación en la educación superior</b> J Texier LACCEI 2013, 12	18	2013
<b>El uso de repositorios y su importancia para la educación en Ingeniería</b> J Texier, MR De Giusti, N Oviedo, GL Villarreal, AJ Lira World Engineering Education Forum (WEEF 2012)* Educación en Ingeniería para ...	15	2012
<b>Las actividades y el planeamiento de la preservación en un repositorio institucional</b> MR De Giusti, AJ Lira, GL Villarreal, J Texier BIREDIAL-Conferencia Internacional Acceso Abierto, Comunicación Científica y ...	13	2012
<b>Model-driven software development in the institutional repositories</b> J Texier, M De Guisti, S Gordillo Dyna 81 (184), 186-192	8 *	2014
<b>The Benefits of Model-Driven Development in Institutional Repositories-Los Beneficios del Desarrollo Dirigido por Modelos en los Repositorios Institucionales</b> J Texier, M De Giusti, N Oviedo, G Villarreal, A Lira BIREDIAL 2012	8 *	2012

Google Académico

Índices de citas	Total	Desde 2012
Citas	92	92
Índice h	5	5
Índice i10	3	3



Coautores Ver todos...

Marisa De Giusti(ORCID:0000-0003-2...

Gonzalo L. Villarreal

Ariel J. Lira

Nestor Oviedo

Jusmeidy Zambrano

Fernando Emmanuel Frati

Alberto Eduardo Riba

María Marta Vila (ORCID:0000-0001-9...



# Identidad digital

## ORCID - Perfil del investigador

 **Gonzalo L. Villarreal**

### ORCID ID

 [orcid.org/0000-0002-3602-8211](http://orcid.org/0000-0002-3602-8211)

[Ver versión pública](#)

 [Mostrar su ID en otros sitios](#)

 [Vista de impresión de registro público](#)

 [Obtenga un código QR para su ID](#)

 **También conocido como**

 **País**  
Argentina

 **Palabras clave**  
prebi, sedici, digital libraries, simulation

 **Sitios web**  
PREBI UNLP  
SEDICI UNLP

### Biografía


I've been working for PREBI-SEDICI, in Universidad Nacional de La Plata, since 2004.  
<http://prebi.unlp.edu.ar> , <http://sedici.unlp.edu.ar> .

My interests include digital libraries, repositories, software development and engineering, and discrete-event simulation.

I've obtained a Ph.D. degree in Computer Science, in which I studied about simulation teaching. My theses is available here: <http://sedici.unlp.edu.ar/handle/10915/29753>

I teach Simulation and Modeling, in Computer Science College of UNLP

### ▼ Educación (3)

 Agregar educación

 Ordenar

#### Universidad Nacional de la Plata: La Plata, Buenos Aires, Argentina

2008-03-01 hasta 2013-09-30

Doctor en Ciencias Informáticas (Facultad de Informática)

Fuente: Gonzalo L. Villarreal

Creado: 2014-02-14

#### Universidad Nacional de la Plata: La Plata, Argentina

2001-02-01 hasta 2007-03-17

Licenciado en Sistemas (Facultad de Informática)

Fuente: Gonzalo L. Villarreal

Creado: 2014-02-14

#### Universidad Nacional de la Plata: La Plata, Argentina

1999-02-01 hasta 2004-07-13

Analista en Computación (Facultad de Informática)

# Identidad digital

## ORCID - Producción científica

▼ Obras (79) + Agregar obras Exportar obras Editar en lote Ordenar

Accesibilidad de los contenidos en un repositorio institucional:  
análisis, herramientas y usos del formato EPUB  
e-Ciencias de la Información  
2016-06-30 | journal-article  
DOI: [10.15517/ecl.v6i2.23690](https://doi.org/10.15517/ecl.v6i2.23690)  
Fuente: Crossref Fuente preferida

Celsius 3-Manual de administración  
2016-03-01 | manual  
URL: <http://sedici.unlp.edu.ar/handle/10915/52120>  
Fuente: Gonzalo L. Villarreal Fuente preferida


Análisis de la producción institucional a partir del uso de  
herramientas de data visualization  
BIREDIAL-ISTEC 2015 (Barranquilla, Colombia, 17 al 21 de noviembre de  
2015)  
2015 | conference-paper  
Fuente: Gonzalo L. Villarreal Fuente preferida

Curso: Preservación digital de documentos. Archivos,  
bibliotecas y museos  
XIX Escuela Complutense Latinoamericana (Buenos Aires, 2015)  
2015 | conference-paper

# Identidad digital



ResearchGate



**Claudia F Pons** id 10.98  
PhD  
National University of La Plat..., La Plata · Facultad de In...

[Message](#)

You follow Claudia. [Unfollow](#)

Overview
Contributions
Info
Scores
Research Interests

RG Score ⓘ

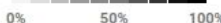
10.98

**Breakdown:**

- 99.74% Publications
- 0.00% Questions
- 0.00% Answers
- 0.26% Followers

**Percentile:**

Claudia Pons's score is higher than 47.5% of ResearchGate members'.



0% 50% 100%

***h*-index** ⓘ

11

***h*-index**

7

excluding self-citations

**Top *h* cited research:**

**Dimensions and Dichotomy in Metamodeling**

Article · Jun 1998

[Follow](#) [See more](#)

# Identidad digital



## Impactstory profiles

**Impactstory**
hi, Gonzalo

**Gonzalo Luján Villarreal**

Universidad Nacional de La Plata Research and development

**open access 85%**

share

OVERVIEW
ACHIEVEMENTS
TIMELINE
PUBLICATIONS

**ACHIEVEMENTS** view all

**Open Access** Top 10%

85% of your research is free to read online. This level of availability puts you in the top 8% of researchers.

**Global South** Top 10%

Of people who save and share your research, 94% are in the Global South. That's a high proportion: only 1% of researchers publish work that

**TIMELINE** view all

5

Online mentions over 1 years

5

**PUBLICATIONS** view all

Accesibilidad de los contenidos en un repositorio institucional: análisis, herramientas y usos del formato EPUB

2016 *e-Ciencias de la Información*

5

**prebi.unlp.edu.ar**

# Identidad digital

Academia.edu

Search... ADVANCED

HOME ANALYTICS SESSIONS READERS MENTIONS

UPLOAD


Message Bell Profile

**Jose Texier**

E Government Adoption +4

9 Followers · 3 Following · 26 Total Views

FOLLOW MESSAGE



Search... ADVANCED

HOME ANALYTICS SESSIONS READERS MENTIONS

UPLOAD

Message Bell Profile

Notifications Mark All as Read

Carlos Nusch started following you. July 31 FOLLOW

Jose Texier invited you to be a co-author on DSpace as a Repository Tool for Administrative Documents in the National Experimental University of Tachira. May 9 ACCEPT

Jose Texier invited you to be a co-author on The Benefits of Model-Driven Development in Institutional Repositories -Los Beneficios del Desarrollo Dirigido por Modelos en los Repositorios Institucionales. May 9 ACCEPT

VIEW ALL NOTIFICATIONS

Los repositorios institucionales bibliográfica y su relación en la

Bookmark Download 2

Representación de los recursos

Bookmark Download 1

La Representación de Recursos SEDICI, Repositorio Institucio

Bookmark Download 1

New paper added to Digital Libraries:

How are on-line Digital Libraries changing Theatre Study and M

The digital era is transforming Performing Art studies: increasingly, digitized records of a performance easily accessed on-line. The recent introduction of new audiovisual and hypermedia tools is deeply tra landscape on three different levels: the status of PA documents, the way Performing Arts researches a Performing Arts repositories. The on-line access to digital video allows for a new management of Perf

BOOKMARK DOWNLOAD by Maia Giacobbe Borelli top 3% 76

# Altmetrics

- El tradicional Índice de Impacto (Impact Factor) se aplica sobre las revistas, pero no analiza puntualmente del impacto de cada artículo
  - Los índices tradicionales (ej. h-index) son lentos, a veces el impacto se refleja muchos años después
  - ¿podemos obtener métricas sobre lo que está sucediendo **en este momento** con una publicación?
- No, si nos limitamos a contar citas

# Altmetrics

Nuevas herramientas para difundir de la producción científica:

- revistas digitales de OA: pueden registrar las descargas y los accesos,
- repositorios institucionales: idem anterior, y a la vez interoperar y diseminar,
- gestores de bibliografía (Mendeley, Zotero), servicios orientados a ciertos documentos (SlideShare, figshare, Scribd,) y redes de investigadores (Research Gate, Academia.edu),
- blogs (científicos) y redes sociales: FB, G+, Twitter, LinkedIn



# Altmetrics

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

## Algunos enlaces de interés

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

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

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

Plumx <https://plumanalytics.com/>



# Altmetrics - Integración con revistas

plos.org create account sign in

**PLOS** GENETICS

Browse Publish About

Search  Q

advanced search

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

## Six1 is essential for differentiation and patterning of the mammalian auditory sensory epithelium

Ting Zhang, Jinshu Xu, Pascal Maire, Pin-Xian Xu

Published: September 11, 2017 • <https://doi.org/10.1371/journal.pgen.1006967>

Article Authors **Metrics** Comments Related Content

0 Save	0 Citation
784 View	1 Share

Download PDF

Print Share

Check for updates

ADVERTISEMENT

Viewed

Total Article Views	HTML Page Views			PDF Downloads	XML Downloads	Totals
	PLOS	PMC				
<b>784</b>	691	0	82	11		<b>784</b>
Sep 11, 2017 (publication date) through Sep 20, 2017 *			0	0	11	<b>0</b>
	<b>691</b>		<b>82</b>	<b>11</b>		<b>784</b>

Totals: 691 82 11 784

11.87 % of article views led to PDF downloads

\*Although we update our data on a daily basis, there may be a 45-hour delay before the most recent numbers are available. PMC data is posted on a monthly basis and will be made available once received.

Cited



Subject Areas

- Cochlea
- Cell differentiation
- Organ of Corti
- Outer hair cells
- Embryos
- Inner hair cells
- Epithelium
- Precursor cells

<http://journals.plos.org/plosgenetics/article/metrics?id=10.1371/journal.pgen.1006967>



# Altmetrics - Integración con IR


 Producción CSIC ▾ Pasarela Estadísticas Contacto
 

 Servicios ▾

DIGITAL.CSIC / Servicios Centrales CSIC / Unidad de Recursos de Información Científica para la Investigación (URICI) / (URICI) Artículos

English [español](#)

Por favor, use este identificador para citar o enlazar a este ítem: <http://hdl.handle.net/10261/79872>

Compartir / Impacto:



 Mendeley
  BASE
 






 Ver citas en Google académico

Visualizar otros formatos: [MARC](#) | [Dublin Core](#) | [RDF](#) | [ORE](#) | [MODS](#) | [METS](#) | [DIDL](#)

 Estadísticas

Exportar otros formatos: [EndNote](#) [BIBTeX](#) [CSV](#)


  
 Comparte tu historia  
de Acceso Abierto

**Título :**  **Open Access and the Changing Landscape of Research Impact Indicators: New Roles for Repositories**

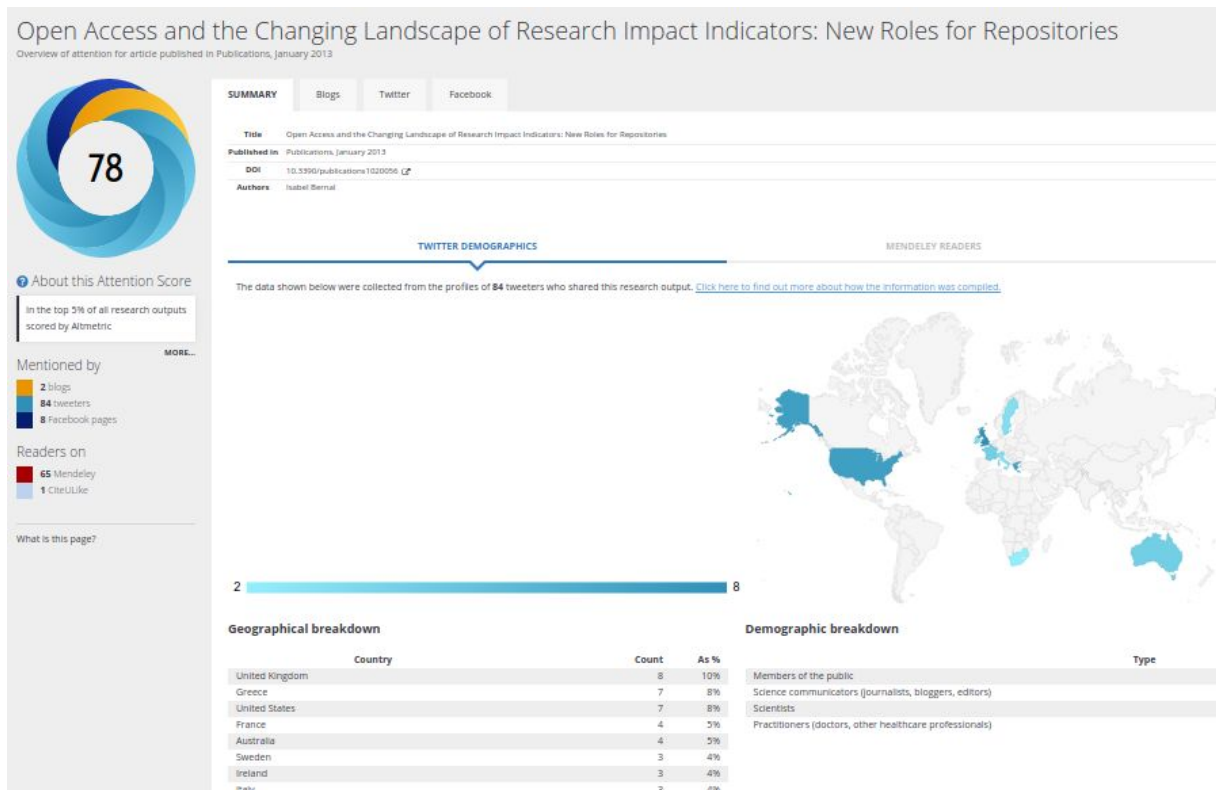
**Autor :** Bernal, Isabel  ; Oficina Técnica de Digital.CSIC

**Palabras clave :** Open access  
Research assessment systems  
New metrics and indicators  
Expanded roles for repositories  
Repositories and research visibility optimization

<https://digital.csic.es/handle/10261/79872>



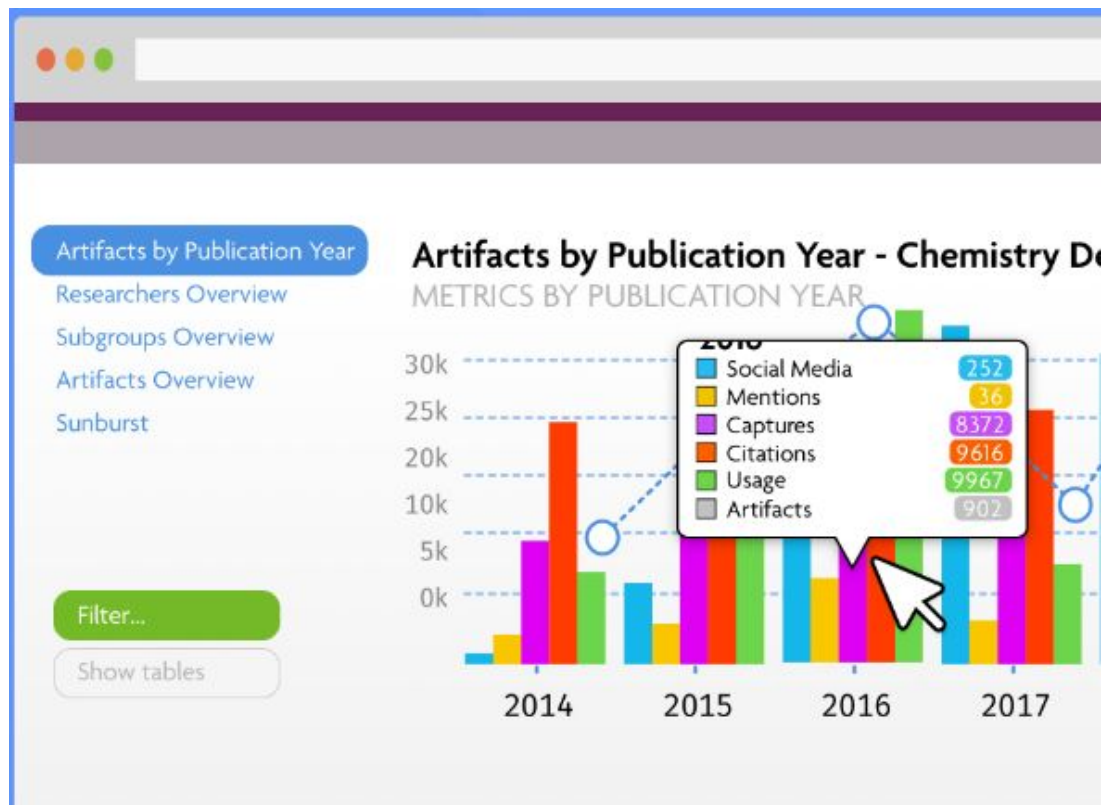
# Altmetrics - Integración con IR



<https://www.altmetric.com/details/1637818>



# Altmetrics



# Altmetrics - Integración con IR



[Products](#) [Learn](#) [Integrate](#) [Interact](#) [About](#)

Stay in touch

Plum Analytics / Products / PlumX for Institutional Repositories

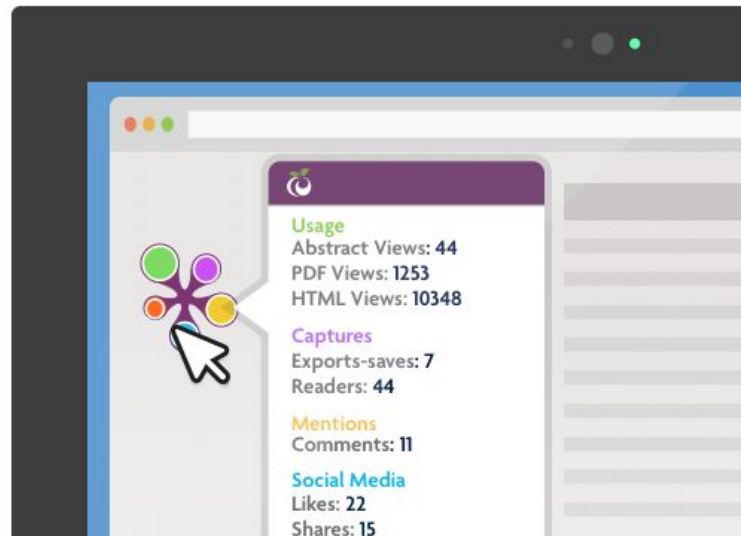
## PlumX for Institutional Repositories:

Get more value from your institutional repository.

PlumX for Institutional Repositories allows you to get more value from an institutional repository (IR) by embedding the research metrics about your research directly into the IR. These research metrics are about the different versions of the same article – not just the one in the IR, so your authors can see the impact and reach of their research in just one place. This makes your IR a more valuable place for research and encourages your authors to deposit their articles and other research.

### Embed metrics where you need them.

The PlumX metrics come to you. The Plum Print shows all of the research metrics about your research that you can easily embed anywhere.



# Altmetrics - Integración con IR

The screenshot shows a web browser window displaying a repository page for the document "Asymmetries in Business Cycles". The browser's address bar shows the URL: <https://repository.urosario.edu.co/handle/10336/8912>. The page header includes the logos for Universidad del Rosario and CRAI (Centro de Recursos para el Aprendizaje y la Investigación). The main content area displays the title "Asymmetries in Business Cycles" and a brief abstract in Spanish. A PlumX metrics overlay is visible, showing the following data:

- Usage: 207
- Abstract Views: 207
- Downloads: 112
- Social Media: 2
- Tweets: 2

The PlumX Metrics Detail Page is also visible at the bottom of the overlay. The page also features a sidebar with navigation options and a footer with the URL: [https://plu.mx/urosario/a/?repo\\_url=http://repository.urosario.edu.co/handle/10336/8912](https://plu.mx/urosario/a/?repo_url=http://repository.urosario.edu.co/handle/10336/8912).

# Altmetrics

- Lagotto le permite al usuario agregar datos relevantes sobre la performance de los artículos de investigación, incluyendo qué tan seguido fueron vistos, descargados, citados, discutidos y recomendados.
- Iniciada en Marzo 2009 por PLOS, se llamó Article-Level Metrics (ALM) hasta Septiembre de 2014.
- Recolecta métricas sobre artículos desde fuentes externas
- Open Source (MIT licence), desarrollada sobre Ruby on Rails 3.2.x (migrando a 4.x), MySQL y CouchDB
- Usada por PLOS, PKP, CrossRef, entre otros.

Información general: <http://articlemetrics.github.io/>

Para desarrolladores y proveedores de servicios: <http://articlemetrics.github.io/docs/>

Lagotto en Github: <https://github.com/lagotto/lagotto>





# Clase 4

# Administración

Tareas cotidianas en el repositorio

# Análisis documental | Catalogación

Es un conjunto de operaciones destinadas a **representar** la forma y el **contenido** de un documento según **normas estandarizadas**. Es el proceso que permite el reconocimiento de un documento a partir de una **descripción unívoca** y sin ambigüedades proporcionando los elementos necesarios para su identificación.



# Análisis documental | Catalogación

## Objetivos

- 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



# Normalización

La normalización persigue **regularizar objetos** que, en la práctica, no suelen ser necesariamente uniformes, con el fin de **facilitar las operaciones** documentales, disminuyendo costes, mejorando la **calidad**, **acelerando procesos** y posibilitando intercambios.

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.



# Lista de normas

- Reglas de catalogación
  - [ISBD](#)
  - [AACR2](#)
  - [RDA](#)
  - [FRBR](#)
  
  - **Números normalizados**
  - ISSN
  - ISBN
  
  - **Sistemas de clasificación**
  - [Dewey](#)
  - [CDU](#)
- **Lenguajes documentales**
  - Tesauro
  - Lista de epígrafes
  
  - **Lenguajes naturales**
  - Palabras clave
  
  - **Normas ISO**



# Análisis documental

**Catalogación, descripción bibliográfica:** describe el documento para su **identificación física**. Análisis de la **forma**: título, autor, páginas, editorial, idioma, etc.

**Indización, clasificación:** se escogen los conceptos más adecuados para representar el **contenido** de un documento (de qué trata) y la **traducción** de estos conceptos a **términos** extraídos de un **lenguaje documental**. Se pretende desarrollar un sistema que permita la búsqueda de los documentos por los temas que tratan.

**Resumen:** breve descripción del contenido de un documento

# Indización, clasificación (análisis de contenido)

Consiste en:

- La búsqueda, análisis y selección de los conceptos presentes en el documento para **describir su contenido**, con miras a su almacenamiento y recuperación.
- La **traducción** de esos conceptos para pasar del lenguaje natural del autor (el que utiliza el discurso científico, técnico o literario) al **lenguaje documental o controlado**, para evitar metáforas, sinonimias y homonimias.

# Indización, clasificación (**Herramientas**)

**Tesauros:** listas de términos (palabra o conjunto de palabras) llamados “**descriptores**” que guardan entre sí relaciones semánticas de **jerarquía, asociación y equivalencia**.

**Listas de encabezamientos de materias:** 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.

# Indización, clasificación (otros recursos)

## Palabras clave

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

**Lista de descriptores libres:** listas de términos (palabra o conjunto de palabras) que elabora el **bibliotecario** según las necesidades de su biblioteca y de sus usuarios.



# Análisis de contenido - Tesauro Eurovoc (SOC)



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 ]

# Análisis de contenido - Sistemas de clasificación

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
Tesaurus	ESCUELAS ESPACIOS EDUCATIVOS PLANIFICACION DE LAS INSTALACIONES NORMAS DE CONSTRUCCION DISEÑO ARQUITECTONICO
Lista de términos libres	ESCUELAS-RANCHO



# Resumen o *abstract*

Representación abreviada del contenido de un documento sin interpretación ni crítica.

Los objetivos del resumen son:

- **Informar al usuario** sobre el **contenido** preciso de los documentos. Proporciona al usuario los **elementos decisorios** sobre la conveniencia de consultar el original íntegro o de desechar su selección por no ser de interés. Determina la pertinencia del documento.
- Es utilizable para la **recuperación automatizada**
- Para las personas con un **interés marginal** por el tema les es suficiente con leer el resumen.

# Normalización de nombres de autores

La falta de normalización de los nombres de los investigadores y de sus centros en las publicaciones científicas y en las principales bases de datos bibliográficas es un hecho conocido, que disminuye la visibilidad de los autores y de sus centros a nivel nacional e internacional y dificulta la recuperación de las publicaciones y de las citas por ellas recibidas.

**VIAF** (Fichero de Autoridades Virtual Internacional) es un proyecto conjunto de varias bibliotecas nacionales, implementado y alojado por OCLC (Online Computer Library Center) <https://viaf.org/>

**IRALIS** (International Registry of Authors-Links to Identify Scientists). Es un sistema de estandarización de las firmas de los autores científicos. <http://www.iralis.org/es>



# Normalización de nombres de autores

[ORCID](#) (Open Researcher and Contributor ID). Sistema internacional para asignar una identificación de autor mediante una URL para facilitar la recuperación de información bibliográfica sobre sus trabajos en diferentes recursos. Es un proyecto apoyado por la mayoría de las grandes editoriales a nivel mundial.

[Researcher ID](#). Permite generar un código identificador de un autor de forma inequívoca para facilitar las búsquedas y la creación de perfiles curriculares y datos bibliométricos como las citas recibidas.

[Mendeley](#). Además de gestor bibliográfico, permite crear un perfil curricular a modo de red social académica donde gestionar las publicaciones de un autor.

[Research Gate](#). Red social académica que permite mostrar y compartir publicaciones.

[Google Scholar Citations](#). Herramienta de Google para crear un perfil profesional de publicaciones y hacer un seguimiento a las citas recibidas.



# Metadatos



# Metadatos

Son la información descriptiva que se aplica sobre un recurso para facilitar su organización, son datos que describen otros datos, información estructurada para describir, explicar, localizar o facilitar la obtención, uso o administración de un recurso de información.

Según la norma ISO 15489-1 del 2001 define los metadatos, en el contexto de la gestión de documentos, como: "datos que describen el contexto, contenido y estructura de los documentos, así como su gestión a lo largo del tiempo (...) Como tales, los metadatos son información estructurada o semiestructurada que posibilita la creación, registro, clasificación, acceso, conservación y disposición de los documentos a lo largo del tiempo y dentro de un mismo dominio o dominios diferentes."

# Papel de los metadatos

- Recuperación de la información
- Administración de documentos
- Gestión de derechos, autoría y propiedad intelectual
- Estado de archivo
- Control y descripción de procesos



# Papel de los metadatos

- Seguridad y autenticación
- Valoración de contenidos
- Preservación y conservación
- Visibilidad de la información
- Actualización de la información



# Tipos de Metadatos

Tipo	Uso	Ejemplos
<b>Administrativo</b>	Usados en la identificación, gestión y administración de recursos de información	Adquisición de información Derechos y reproducción Requerimientos legales para el acceso Localización de información Criterios de selección para la digitalización Control de la versión
<b>Descriptivo</b>	Utilizados para representar recursos de información	Registros catalográficos Proporcionar ayuda en la búsqueda Índices especializados Hiperenlazar relaciones entre recursos Anotaciones de los usuarios
<b>Preservación</b>	Para salvaguardar los recursos de información	Informar sobre las condiciones de uso de los recursos físicos Informar sobre las acciones llevadas a cabo para preservar versiones físicas y digitales de recursos
<b>Técnico</b>	Relativos a cómo funcionan los sistemas o el comportamiento de los metadatos	Documentación de hardware y software Digitalización de la información (formato, ratio de compresión...) Autenticación y datos de seguridad (encriptación, passwords, etc.) Control de tiempo de respuesta de sistemas
<b>Uso</b>	Relativos al nivel y tipo de uso que se hace con los recursos informativos	Información sobre versiones Reutilización del contenido del recurso

# Metadatos – Dublin Core (DCES)

Dublin Core DC es un esquema de metadatos elaborado por la DCMI (Dublin Core Metadata Initiative). Actualmente es el esquema de datos más utilizado. Consta de 15 elementos básicos.

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



# Metadatos – Dublin Core cualificado (DCTERMS)

El Dublin Core cualificado es una extensión de esquema Dublin Core básico donde algunos de sus elementos están especializados:

- **Título:** dc.title ⇒ dcterms.alternative, dcterms.title
- **Relación:** dc.relation ⇒ dcterms.isVersionOf, dcterms.isPartOf, ...
- **Fecha:** dc.date ⇒ dcterms.created, dcterms.available

Otros tipos de esquema son:

- Networked Digital Library of Theses and Dissertations (ETD)
- Metadata Objects Description Schema (MODS)
- Categories for the Description of Works of Art (CDWA )
- Cataloguing Cultural Objects (CCO)



## Networked Digital Library of Theses and Dissertations (ETD)

- Universidad de Utah State
  - Versiones electrónicas de tesis o disertaciones
  - Metadatos utilizados en SEDICI
- 
- `thesis.degree.grantor` (Institución garante)
  - `thesis.degree.name` (Grado alcanzado)



# Metadata Objects Description Schema (MODS)

- Desarrollado por la Oficina de Desarrollo de Redes y Normas MARC de la Biblioteca del Congreso (LC)
- MODS es un conjunto de metadatos que tiene el mismo objetivo que DC, pero que busca superar los problemas y limitaciones detectados en la aplicación de DC. Para ello, toma como punto de partida el formato MARC, del que selecciona diferentes En SEDICI se utilizan:
  - **mods.location** (URL de acceso al catalogo de la biblioteca, o información de acceso al recurso físico)
  - **mods.origenInfo.place** (Institución donde se creó el documento)
  - **mods.recordInfo.recordContentSource** (Registra el nombre de la entidad y/o sitio web desde donde se obtuvo la información del registro)

## Categories for the Description of Works of Art (CDWA)

Directrices para la documentación de los bienes patrimoniales desarrolladas por un grupo de trabajo dirigido por el Getty y la College Art Association of America (CAA),

## Cataloguing Cultural Objects (CCO )

Manual para documentar arte, arquitectura y objetos culturales desarrollado y diseminado por el Visual Resources Association (VRA).

# Metadatos - Esquema de metadatos SEDICI

SEDICI utiliza un perfil de metadatos propio: dichos metadatos están traducidos al inglés y al portugués, y el sistema los transforma al formato [Dublin Core](#) para que sean accedidos por otros repositorios vía [OAI-PMH](#).

Ej.: [tesis](#)

dc.format.extent	110 p.	es
dc.language	es	es
dc.title	Familias y estrategias de crianza	es
dc.type	Tesis	es
sedici.title.subtitle	Prácticas de tránsito de niños y niñas por diversos grupos de crianza en sectores populares de Viedma	es
sedici.creator.person	Belmartino, María Belén 🇺🇦	es
sedici.subject.materias	Trabajo Social 🇺🇦	es
sedici.subject.decs	Familia 🇺🇦	es
sedici.subject.decs	Crianza del Niño 🇺🇦	es
sedici.subject.decs	Población Suburbana 🇺🇦	es
sedici.description.fulltext	true	es
mods.originInfo.place	Facultad de Trabajo Social 🇺🇦	es
sedici.subtype	Tesis de maestría	es
sedici.rights.license	Creative Commons Atribución-NoComercial-CompartirIgual 2.5 Argentina (CC BY-NC-SA 2.5)	
sedici.rights.uri	<a href="http://creativecommons.org/licenses/by-nc-sa/2.5/ar/">http://creativecommons.org/licenses/by-nc-sa/2.5/ar/</a>	
sedici.contributor.director	Colangelo, María Adelaida 🇺🇦	es
sedici.contributor.juror	Colangelo, María Adelaida	es
sedici.contributor.juror	Fuentes, Pilar	es
sedici.contributor.juror	Velurtas, Marcela	es
thesis.degree.name	Magister en Trabajo Social 🇺🇦	es
thesis.degree.grantor	Facultad de Trabajo Social 🇺🇦	es
sedici.date.exposure	2012-06-04	



## Tipologías



# Tipologías y formatos

**Artículo:** Un artículo es un texto en prosa no ficcional en el que se desarrolla un tema en particular. Está escrito por un autor o varios autores, quienes llevaron adelante una investigación y escriben para publicar los resultados de investigación en una revista científica o académica.

**Sonido:** Recurso diseñado principalmente para ser oído. Ejemplos: formato de archivo de audio, CD-ROM y grabaciones de sonido. (AUDIO EN CIC)

**Comunicación de congreso:** Comunicación de congreso que se presenta en el congreso y se expone ante la audiencia. La comunicación del congreso se publica en las actas.

Fuente: Centro Argentino de Información Científica y Tecnológica [Vocabulario de tipos de recursos de información](#)



**Contribución a revista:** Texto (por ejemplo: artículo, columna, editorial u otra contribución) escrito para ser publicada en una revista académica o científica.

**Imagen fija:** Representación visual estática. Este tipo de imágenes incluye diagramas, dibujos, diseños gráficos, planos, mapas, fotografías e impresiones.

**Informe:** Se denomina informe al registro de hallazgos de una investigación terminada o en progreso, o de otros hallazgos técnicos. Por lo general, está identificado por un número de informe y a veces también por un número de subsidio asociado a la agencia de financiamiento.

**Libro:** Una publicación no seriada que está completa en un volumen o en un número definido de volúmenes. En su formato físico, un libro es una serie de hojas de papel, pergamino u otro material encuadernadas en uno de sus extremos con tapas. Los libros se suelen identificar con un ISBN.



**Capítulo de libro:** Un capítulo o sección de un libro, por lo general con un título o número de capítulo que lo identifica.

**Publicación Periódica:** Publicación seriada con su propio título identificador, caracterizada por una variedad de contenidos y contribuciones (por ejemplo, artículos, editoriales, reseñas, columnas), que se publica con una periodicidad regular.

**Reseña:** Una reseña del trabajo de un tercero.

**Tesis:** una tesis o disertación es un documento presentado para postularse a un título académico o a un cargo profesional, en la que el autor presenta un tema de investigación y los resultados. Pueden ser de grado, maestría o doctorado.



## Formatos de Archivos



# Formatos Digitales

Un formato digital es un sistema particular de codificación de la información para su almacenamiento o tratamiento en un soporte informático o digital.

Los formatos más conocidos son el *.doc* para texto, *.tiff* y *.jpeg* para imágenes, *.mp3* y *.wav* para audio, *.avi*, *.mpg* para video, etc.



Los objetivos perseguidos por un formato específico pueden ser diversos:

- Almacenar un solo tipo de contenido plano sin ninguna codificación adicional. Ejemplo: .txt
- Incorporar especificaciones para codificar la información (principalmente para su compresión, transmisión o cifrado) Ejemplo: .pdf
- Combinar y sincronizar varios tipos de contenido en un solo archivo. Ejemplo: los archivos .mpeg o .AVI que incluyen pistas de audio, vídeo, subtítulos, metadatos, etc.



# Cuestiones a tener en cuenta

- **¿Su uso está generalizado?** ¿Existen varios programas para leer este formato? ¿Está utilizado por otras instituciones como formato de preservación?
- **¿Está abierto?** El uso del formato no debe ser regido por patentes.
- **¿Está documentado?** ¿Fue publicada su documentación (un formato puede estar documentado sin estar necesariamente abierto)? ¿Fue normalizado por instituciones como W3C o ISO? Esta documentación permitirá construir nuevos programas para leer el formato si sus vendedores ya no se encargan de su mantenimiento.
- **¿Existen programas para validarlo y caracterizarlo?**



# Dificultades para la elección de Formatos

- Cuando uno elige el formato más adecuado al contenido, debe valorar los riesgos de **obsolescencia** que corre el formato.
- También se debe tomar en cuenta si las políticas del repositorio o la legislación de la institución impondrán que sean aceptados todos los datos, sea cual sea su formato.
- La finalidad del archivo digital y su ubicación en la cadena de valor: formato inicial de **creación**, formato intermedio para **edición**, formato final para preservación o **difusión**. A veces el formato mostrado al público no es el mismo en el que se preserva.

# Niveles de Riesgo de un Formato (BnF)

Nivel	Descripción
<b>Almacenado</b>	Formato no identificado para el cual sólo se asegura la preservación de la cadena de bits.
<b>Identificado</b>	Formato identificado por una herramienta de identificación para el cual no fue definido ninguna estrategia de preservación.



<b>Nivel</b>	<b>Descripción</b>
<b>Conocido</b>	Formato para el cual la BnF tiene por lo menos una herramienta de referencia, realiza una vigilancia tecnológica y define una estrategia de preservación, sea emulación o migración a un formato controlado.
<b>Controlado</b>	Formato documentado por el cual la BnF tiene herramientas de referencia, realiza una vigilancia tecnológica e impone requisitos a los productores.



## Factores de sostenibilidad a tener en cuenta en la selección de los formatos

**Divulgación/Transparencia:** hace mención al grado en el que las especificaciones técnicas de los formatos son accesibles (es decir son de código abierto).

**Apertura:** se refiere a la dependencia de un formato determinado a una patente. En el caso de las entidades dedicadas a la preservación y/o difusión del patrimonio cultural, en general, se tiende a utilizar formatos abiertos, es decir formatos de archivo que no se hallan sujetos a patentes o derechos de autor. Ejemplo: Empresas como Apple, Amazon, Microsoft conforme con criterios comerciales y no necesariamente de preservación, pueden imponer formatos propietarios, es decir, protegidos por patentes o derechos de autor.

**Dependencia/Interoperabilidad:** se trata del grado de dependencia de un formato determinado respecto a un hardware, a un software o a un sistema operativo específico.

**Estabilidad/Compatibilidad:** grado en el que un formato mantiene su funcionalidad e integridad con versiones anteriores o posteriores.

**Aceptación:** hace referencia al grado de utilización de los formatos por parte de los creadores, distribuidores y usuarios de los recursos.

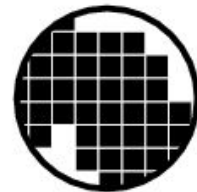
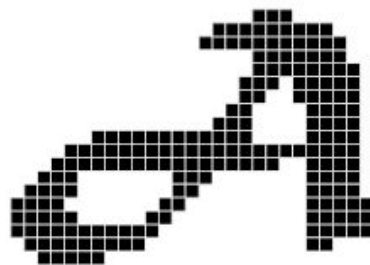
**Estandarización:** adecuación formal a los procesos o especificaciones establecidos por un organismo de normalización (ISO, NISO, W3C...) con el objetivo de garantizar la calidad de los archivos y su interoperabilidad.

**Mecanismos de protección técnica:** los mecanismos de protección técnica como por ejemplo el cifrado, utilizados habitualmente para proteger la propiedad intelectual, no deben dificultar la recuperación de datos, migración de los contenidos o su adaptación a nuevas necesidades derivadas de la evolución tecnológica.

# Formatos de Imagen

Existen dos tipos básicos de formatos de imagen:

- Gráficos o imágenes pixeladas o de **mapa de bits** (raster images): las imágenes están representadas por píxeles. Es el tipo de formato que se utiliza comúnmente para captar imágenes con cámaras digitales y escáneres.
- imágenes **vectoriales**: están definidas matemáticamente y utilizan formas geométricas para representar las imágenes. En este caso la información que almacenan no hace referencia a los píxeles. Se utilizan principalmente en actividades como el diseño gráfico, ingeniería, etc. El tamaño de los ficheros es muy pequeño.



Las dos características básicas que determinan la calidad de la imagen, principalmente en las imágenes pixeladas, son la **resolución** y la **profundidad de color** o profundidad del bit.

La **resolución** es la cantidad de puntos que componen una imagen, es decir, la capacidad de distinguir los detalles espaciales más finos. Habitualmente se expresa en píxeles por pulgada (ppp o ppi en inglés). Lo importante es que a mayor número de píxeles, mayor calidad de la imagen y mayor peso del archivo.

La **profundidad de color** o **profundidad del bit**, se refiere al número de bits utilizados para describir el color de cada píxel. Cuanto mayor sea la profundidad de bits mayor será la cantidad de tonos representados

# Número de colores visibles por bit

Profundidad de Color	Nº de colores visibles
1 bit (monocromo)	2 tonos
4 bit	16 tonos
8 bit (indexed color)	256 tonos
24 bit (true color)	16.777.216 tonos

# Formatos de Imagen Recomendados (IFLA)

Profundidad de Color	Nº de colores visibles
Preservación documentos impresos y manuscritos y otros materiales especiales (fotos, planos, grabados, mapas...)	Formato TIFF sin compresión / Resolución mínima 400 ppp / Color: Escala grises 256 gamas / 16,7 millones colores
Preservación prensa histórica	Formato TIFF sin compresión / Resolución mínima 400 ppp / Color: Escala grises (256 gamas)
Preservación prensa moderna	Formato TIFF sin compresión / Resolución mínima 300 ppp / Color: Escala grises (256 gamas)
Difusión	archivos con compresión JPEG

\* En la digitalización de imágenes, el TIFF sigue siendo el formato más utilizado en la preservación de documentos, aunque el JPEG 2000 está comenzando a ser usado siempre que la compresión sea sin pérdida.



# Formatos de Audio



Se distinguen dos categorías elementales de formatos:

- Por un lado, aquellos que recogen en forma de datos el sonido creado directamente a través de dispositivos electrónicos como ordenadores, sintetizadores, etc. para que después sean interpretados por el software utilizado para su reproducción.
- Por otro, los formatos más comunes y que interesan en este caso, los utilizados en las grabaciones musicales, de voz o de cualquier otro tipo, como por ejemplo el WAV, que almacena muestras de la onda de sonido para su posterior reproducción (formatos waveform).



# Formatos de Audio

En contenidos de audio son dos las características principales a considerar:

- **Frecuencia de muestreo:** número de muestras que se toman de la onda de sonido en un período de tiempo, se expresa en Khz.
- **Profundidad de bits:** número de bits utilizados para representar la onda sonora.

*A mayor frecuencia de muestreo y profundidad de bits, mayor calidad del sonido. Por ejemplo, para obtener una calidad de CD de audio, la grabación se realiza a 44,1 Khz y 16 bits. En cambio, cuando sólo se va a difundir, se eliminan espacios vacíos y algunas frecuencias del archivo.*



# Formatos de Audio Recomendados

## Recomendaciones IASA (International Association of Sound and Audiovisual Archives) para preservación

Formato WAVE y BWF con fines archivísticos

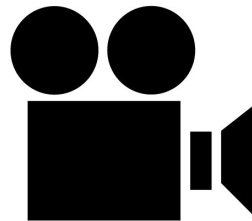
Calidad recomendable: 96 kHz / 24 bit

Calidad mínima: 48 kHz / 24 bit

Las mismas especificaciones para las grabaciones de voz



# Formatos de Video



Los formatos de vídeo son en general, **formatos contenedores** o formatos de compresión de información audiovisual ya que contienen varios tipos de información: pistas de vídeo, pistas audio, en ocasiones texto o incluso imágenes incrustadas para los subtítulos y los metadatos necesarios para su sincronización.

En el caso de los contenidos audiovisuales resulta más complejo establecer unas directrices de uso común para la preservación por los diferentes tipos de información (audio, imagen...) y los productos y canales de difusión específicos para los que están diseñados: grabación, edición, cine, DVD, internet.

**Resolución de la imagen:** definida por el número de píxeles («puntos») que componen la imagen.

**Velocidad de la imagen (frame rate):** número de imágenes estáticas o fotogramas por segundo, se expresa en fps (frames per second). Existen distintos estándares para cine (24 fps) y los distintos sistemas de televisión, etc.

**Tasa de bits (bit rate):** velocidad de transferencia de los datos contenidos en el archivo audiovisual. Se expresa en bits por segundo (bps) y determina la nitidez de la imagen: a mayor velocidad más información puede contener la imagen.

**Profundidad de color:** al igual que en las imágenes estáticas, se refiere al número de bits utilizados para describir el color de cada píxel.

**Relación de aspecto o tamaño del cuadro:** proporción entre el ancho y la altura de la imagen.

# Formatos de Video Recomendados

<b>Preservación</b>	<b>Difusión</b>
JPEG2000 (ISO/IEC 15444-4) (.mj2)	JPEG 2000 MPEG4 AVI MOV MP4 MKV OGG



## Formatos de Texto

El formato más difundido para la publicación de textos es el formato de documento portátil o **PDF**, pero no es raro encontrar autoarchivos de materiales hechos en formatos **.doc** o **.docx** u otro tipo de formatos de texto editable como **.odt**. En estos casos siempre es recomendable la transformación del material al formato PDF. El formato PDF fue creado por Adobe, y es ahora un estándar abierto y oficial reconocido por la Organización Internacional para la Estandarización (ISO). A los fines de la preservación digital el formato recomendado es el PDF/A. El **PDF/A** es el estándar más común para los documentos de texto con formato, pero muchas entidades que ofrecen contenidos en formatos de texto electrónico en formato **EPUB**. Ambos formatos están basados en XML.



## Formatos recomendados para Texto

El **PDF/A** se presenta como el estándar aceptado para la creación de documentos digitales accesibles online y susceptibles de ser impresos, tanto aquellos basados en texto como los que incluyen imágenes, gráficos, etc. que requieren de un diseño preciso.

El **EPUB** es el estándar de facto recomendado para el texto electrónico. Aunque puede soportar imágenes está más orientado a la publicación de texto, por ello no es el formato más adecuado para documentos que requieren un diseño preciso o están basados en imágenes.

## Ventajas del PDF/A

PDF/A es, de hecho, un subconjunto de PDF obtenido excluyendo aquellas características superfluas para el archivado a largo plazo de forma similar a como se ha definido el subconjunto PDF/X para la impresión y artes gráficas. Además, el estándar impone una serie de requisitos a los programas para la visualización de archivos PDF/A.

Un programa de visualización que se ajuste a los requisitos debe seguir ciertas reglas incluyendo la conformidad con las directrices en cuanto a la gestión de color, el uso de fuentes integradas a la hora de la visualización, o la posibilidad de realizar anotaciones por parte del usuario.


## En síntesis:

- Para la preservación se utilizan en mayor medida los formatos no propietarios, reconocidos como estándares. Cuando se trata de difundir existe una mayor flexibilidad.
- Los formatos presentados como válidos para la preservación, también lo son para la difusión de los contenidos aunque se suele aplicar algún tipo de compresión o se reduce de calidad en pos de su funcionalidad.
- La principal divergencia encontrada en los formatos utilizados en la industria y en el ámbito de la preservación, reside en la utilización de medios de protección técnica.

# Formatos Sugeridos

Tipo de contenido	Preservación	Difusión
<b>IMAGEN</b>	TIFF (sin compresión) JPEG2000 (sin pérdida)	JPEG JPEG2000 (con pérdida) TIFF (con compresión)
<b>TEXTO</b>	EPUB (libro electrónico) PDF/A1 (ISO 19005-1)	EPUB PDF/A PDF HTML XHTML Open office Texto plano (ISO 8859-1)
<b>AUDIO</b>	BWF WAVE	MP3
<b>VIDEO</b>	JPEG2000 (ISO/IEC 15444-4) (.mj2)	JPEG 2000 MPEG4 AVI MOV

*Fuente: "Formatos de Difusión y Formatos de Preservación de Contenidos Digitales" Kultura 2.0-Cuadernos de Formación Gobierno Vasco - Departamento de Educación, Política y Cultura*



# Comunidades y Colecciones



# Comunidades y colecciones

- Las comunidades pueden contener subcomunidades o colecciones, pero no ítems
- Las colecciones sólo pueden contener ítems
- Un ítem pertenece a una sola colección, pero puede estar asociado a otras colecciones

# Servicios de un repositorio digital

- Búsqueda y recuperación
- Exploración
- Copyright
- Autoarchivo
- Servicios a otros sistemas



# 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.



# 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.

# Búsqueda y recuperación: criterios mínimos

**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.

# 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.

# 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.



# Trabajo Cotidiano



# Administración: ingesta

¿Cómo llegan los materiales?

- 👉 **Autoarchivo**
- 👉 **Carga mediada**
- 👉 **Sword (OJS, interoperabilidad)**
- 👉 **Otros repositorios y portales**
- 👉 **Sitios específicos de las revistas/congresos**
- 👉 **Convenios con otras instituciones**
- 👉 **Personalmente**
- 👉 **Otros**



# Administración: ingesta

## Autoarchivo

Los materiales son enviados por los usuarios a través de un formulario de carga sumamente sencillo. En el caso de SEDICI es usado mayoritariamente por los tesisistas de **posgrado**, quienes por **mandato (resolución UNLP 78/11)** deben depositar sus tesis para obtener su título. No obstante, también lo utilizan tesisistas de **grado** así como investigadores que ni bien publican suben sus trabajos al repositorio. También es utilizado por miembros de laboratorios e institutos de la UNLP que muestran sus materiales en el repositorio (caso [LEEM](#)).



# Administración: ingesta

## Pre-carga mediada

En este caso, se trata de colecciones particulares cuyos **administradores** son **externos** al repositorio. Luego de haber recibido una **capacitación** en el uso de DSpace y algunas nociones de catalogación, envían sus materiales cargándolos directamente en la colección que les corresponde y luego se completa la catalogación por parte de los administradores. Es el caso de las colecciones pertenecientes a los **museos de la UNLP, REA y Radio Universidad**.

# Administración: ingesta

## Sword (OJS)

Mediante el protocolo Sword se **importan** al repositorio los recursos de las revistas que se gestionan a través del **Portal de Revistas** de la UNLP. Este **plugin** permite que los administradores del repositorio reciban tanto los archivos como los **metadatos básicos** de los recursos y sólo deban **normalizarlos**, y en algunos casos agregarles otros metadatos que no contempla OJS. Este plugin no se está utilizando en la actualidad debido a la última actualización de OJS.

# Administración: ingesta

## Otros repositorios y portales UNLP

Periódicamente, los administradores del repositorio recorren sitios de la UNLP como **Memoria Académica**, **Revistas de la FAHCE** o **Libros de la FAHCE** para actualizar las colecciones correspondientes. En este caso la carga se realiza en forma manual.

# Administración: ingesta

## Sitios específicos de revistas/congresos u otros

Del mismo modo, se recorren periódicamente los sitios de las revistas, congresos, unidades académicas y otros dentro del ámbito de la UNLP para actualizar las colecciones correspondientes. Ejemplo: **Papel Cosido** (Bellas Artes).

# Administración: ingesta

## Convenios con otras instituciones

Otras instituciones, vinculadas de un modo u otro a la UNLP, alojan sus contenidos en el repositorio debido a su impacto y visibilidad. De este modo, el repositorio obtiene otra vía de ingesta de material, como las colecciones pertenecientes a la **RedUNCI** o la **Academia Nacional de Agronomía y Veterinaria**.

# Administración: ingesta

## Personalmente

Los usuarios pueden concurrir al repositorio en forma personal y llevar sus materiales en un pen-drive o CD, siempre que su filiación corresponda a la universidad como alumnos, docentes, becarios, no docentes o investigadores.

También se reciben materiales por e-mail, aunque no es la vía más aconsejable (se recomienda siempre utilizar el autoarchivo).

# Administración: ingesta

## Otros

El repositorio posee en muchos casos los datos bibliográficos de un recurso pero no su texto completo. En ese caso, siempre que sea posible, se procura **contactar al autor** y solicitarle el texto para completar esos registros. En otros casos, cuando se trata de materiales más antiguos, de los que se posee su ubicación en alguna **biblioteca** de la UNLP, se solicita a la administración de la biblioteca que se facilite el ejemplar para su **digitalización** y posterior puesta en línea (ejemplo, tesis [Patricia Zalba](#)).



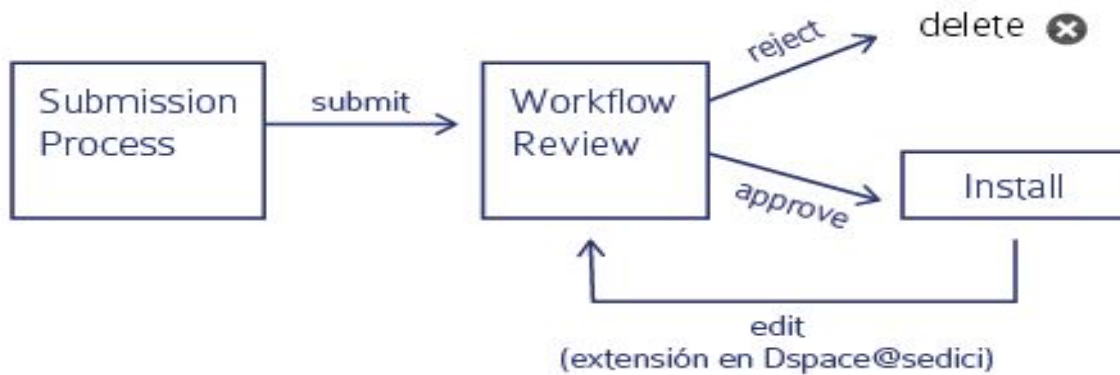


# Autoarchivo

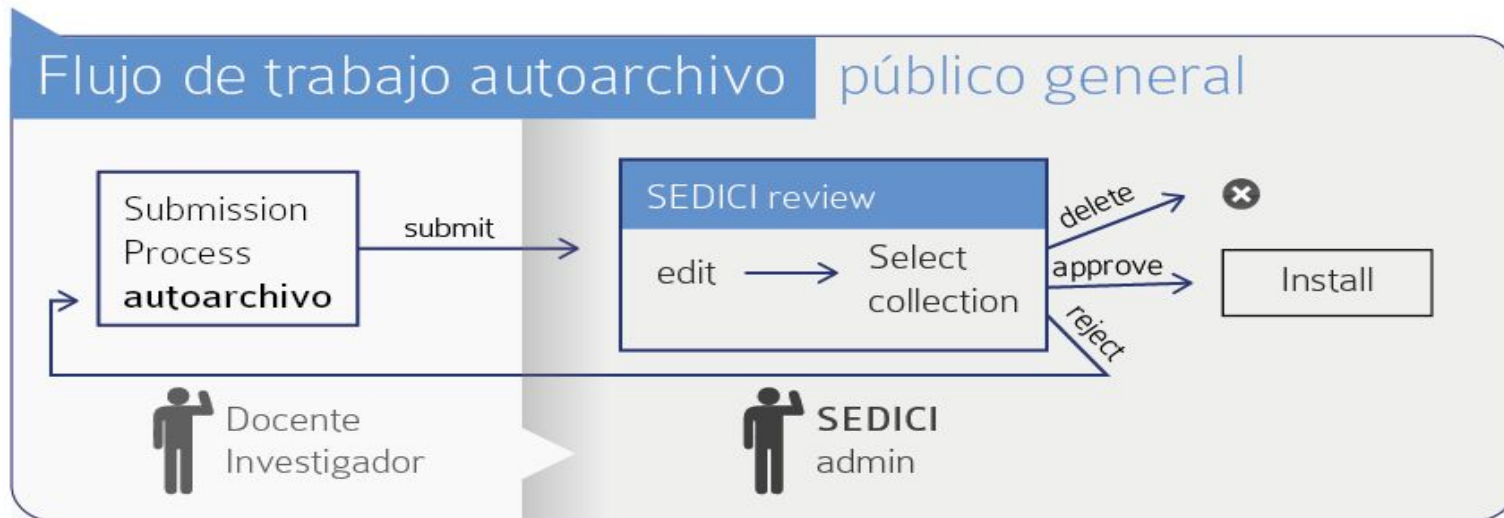


# SEDICI | Flujo de trabajo

## Flujo de trabajo general



# SEDICI | Flujo de trabajo

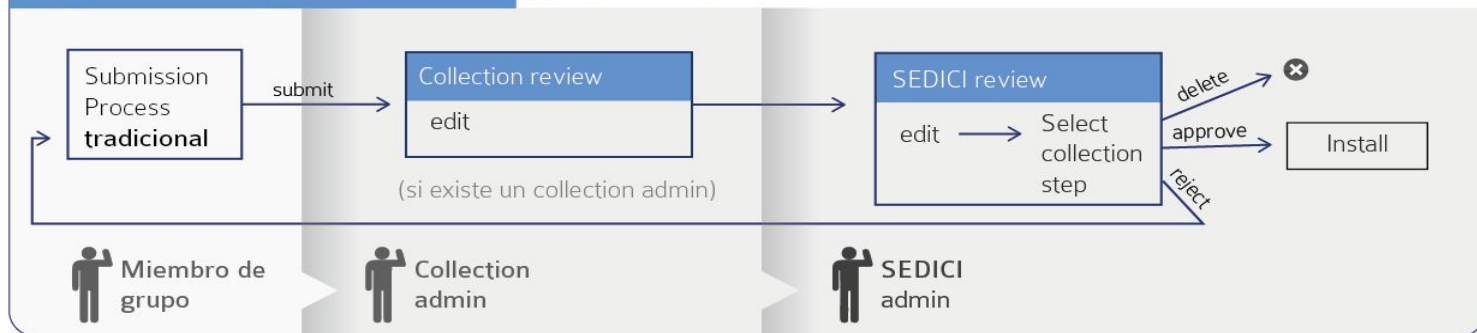


Autoarchivo



# SEDICI | Flujo de trabajo

## Flujo de trabajo tradicional grupos específicos



Pre-carga



# SEDICI | Flujo de trabajo diario

## Carga de documentos



[Analía Pinto](#)

Salir

Editar perfil

Mis documentos

## Cargas Pendientes

Envíos pendientes. Podría **Nuevo envío**.

Título	Colección	Remitente
No hay envíos incompletos.		

Perfil de administrador

## Tareas del flujo de trabajo

Estas tareas son ítems que están esperando aprobación antes de ser añadidas al repositorio. Hay dos colas de tareas, una para tareas que ha aceptado y otra para tareas que todavía no han sido asumidas por nadie.

### Sus tareas

Tarea	Ítem	Colección	Remitente
<input type="checkbox"/> Revisión SEDICI	Guía de buenas prácticas para la administración de ...	Libros	Analía

**Devolver las tareas seleccionadas a la cola**

### Tareas en cola

Tarea	Ítem	Remitente
Autoarchivo (2)		
<b>Asumir las tareas seleccionadas</b>		

Tareas en cola del administrador



# SEDICI| Flujo de trabajo diario

Tareas en cola

Tarea	Ítem	Remitente
Autoarchivo (2)		
<b>Asumir las tareas seleccionadas</b>		

Tareas en cola general



# SEDICI | Supervisión

Las tareas de supervisión dentro de un repositorio incluyen:

- 👉 **Distribución** de las tareas a realizar entre los administradores. Esto supone desde la correcta recepción del material por las diferentes vías (autoarchivo, recolecciones desde distintos portales dentro y fuera de la universidad, envíos por e-mail, digitalizaciones propias, convenios con terceros) hasta la asignación de qué material va a cargar cada administrador, de acuerdo a su experiencia y habilidades.
- 👉 **Revisión** exhaustiva de registros cargados en el pasado, especialmente del momento en que SEDICI no funcionaba sobre DSpace.

# SEDICI | Supervisión

Las tareas de supervisión dentro de un repositorio incluyen:

- 👉 Revisión de **registros** como resultado de nuevas políticas adoptadas, por actualizaciones del sitio o por nuevas normalizaciones.
- 👉 Revisión constante de las cargas realizadas en la actualidad, para ejercer un estricto **control de calidad** sobre los datos y metadatos expuestos al mundo.
- 👉 **Correcciones** sobre los registros de diverso tipo.
- 👉 Tareas de revisión y corrección asignadas desde el **Sistema de Gestión de Incidencias** de SEDICI (también conocido como “trac”).

# SEDICI | Comunicación

La comunicación dentro del repositorio puede dividirse en dos grandes grupos:

## ❑ **Comunicación interna**

- ❑ Soporte a los usuarios del sitio (consultas al mail, telefónicas y a través de **UserVoice**)
- ❑ **Consultas** de usuarios/tesistas sobre derechos de autor, embargo, autoarchivo
- ❑ Redacción de **tutoriales** para el uso apropiado del sitio (cómo realizar autoarchivo, por ejemplo)
- ❑ **Capacitación** de administradores internos/externos
- ❑ Redacción y edición de los manuales de procedimientos y de todo otro tutorial de uso interno para los administradores del repositorio

# SEDICI | Comunicación

## ☐ Comunicación externa

- ☐ Presencia de SEDICI en las **redes sociales** (Twitter, Facebook, LinkedIn, SlideShare y otras).
- ☐ Redacción de posteos para el **blog de SEDICI**.
- ☐ Preparación de posteos para la fanpage de SEDICI
- ☐ Estrategias de **difusión y visibilidad**.
- ☐ **Publicidad** mediante folletos y boca a boca.



# SEDICI | Comunicación interna

## 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



# SEDICI | Manual de procedimientos

Unidades académicas → Presidencia → PREBI-SEDICI → Cursos, presentaciones, tutoriales y otros

## Manual de procedimientos del Servicio de Difusión de la Creación Intelectual (SEDICI)

Autor: Servicio de Difusión de la Creación Intelectual (SEDICI)

2017

Tipo de documento: Libro



### Resumen

Este manual se organiza del siguiente modo: en la primera parte, titulada «Administración de...», detallan todas las tareas que hacen a la catalogación de los ítems, los metadatos utilizados, las... reciben los materiales, cómo se ordenan las colecciones y comunidades, y todo aquello que su... principal del repositorio, que es la puesta en línea, en forma adecuada y jerarquizada, ... académica, científica y artística de la universidad. En la segunda parte, titulada «Otras tareas...», diferentes roles que un administrador puede asumir así como otras tareas que se desarrollan e... repositorio, como la comunicación, la generación de contenidos para redes sociales y otros asp... en el apartado de «Anexos» se incluyen documentos de consulta para los administradores. E... decurso del manual se incluyen cuadros y recordatorios sobre distintos aspectos de la carg... facilitar la tarea diaria de los administradores; también se ofrecen enlaces para dar a cono... conceptos que provienen de otros campos del saber, principalmente de la informática.

### Información general

Fecha de publicación: 2017

Editor: Pinto, Analía Verónica

Colaborador: Folegatto, Lucas (diseño de tapa) | De Giusti, Marisa (revisión técnica) | Villarrea... (revisión técnica) | Nusch, Carlos (revisión técnica) | Pelloche, Silvia (revisión técnica)

Idioma del documento: Español

Editorial: Servicio de Difusión de la Creación Intelectual (SEDICI)

Institución de origen: Servicio de Difusión de la Creación Intelectual (SEDICI)



# SEDICI | Manual de procedimientos

El nuevo manual, publicado en 2017, mejoró aspectos del anterior e incorporó otros (nuevos metadatos, nuevos tipos documentales incorporados al repositorio en los últimos años, nuevas normalizaciones) y a diferencia del anterior está puesto en línea en formato **libro**, para que pueda ser difundido tanto entre usuarios comunes como entre administradores de otros repositorios que quieran tomarlo como modelo para los suyos.

# SEDICI | Edición

Si bien esta no es una tarea específica de un repositorio ni tampoco cotidiana, en SEDICI se han realizado varias experiencias de edición de libros electrónicos. Los primeros de ellos fueron realizados en 2004 ([Notas sobre federalismo fiscal](#), de Horacio L. P. Piffano) y 2005 ([Microeconomía aplicada a la educación universitaria](#), del mismo autor), pero el proyecto más ambicioso encarado hasta el momento ha sido el del libro *Cirugía. Bases clínicas y terapéuticas*, único en su género.



# SEDICI | Edición

Tanto en su primera como en su segunda edición, el libro incluye, además de texto, gráficos e imágenes, videos de diversas operaciones quirúrgicas relatadas por destacados profesionales de la cirugía argentina e internacional. El libro es uno de los materiales más consultados y descargados de SEDICI y a fines del año pasado se editó y publicó su tercera edición, revisada y ampliada.

Otros libros editados en SEDICI:

Análisis económico del derecho tributario, de Horacio L. P. Piffano

XVIII Ibero-American Science and Technology Education Consortium General Assembly (compilación de ponencias)

Anales de jAUTI 2012 (compilación de ponencias)



# SEDICI | Autoarchivo

*Demostración de cómo se realiza un autoarchivo por un usuario común y luego el camino que sigue el envío cuando es tomado por los administradores del repositorio.*





# Objetivos

- Crear conciencia en los profesionales, usuarios, funcionarios de bibliotecas y archivos, políticos, investigadores, etc., sobre los riesgos que conlleva mantener en el tiempo los objetos digitales y dar accesibilidad permanente a los mismos.
- Analizar estándares e implementaciones para cumplir con el objetivo de preservación.
- La digitalización sus dificultades y la generación de nuevos materiales para preservar.



# Introducción

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

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



# La preservación de los contenidos

En los documentos en papel se habla de “negligencia benigna”: el olvido de un manuscrito en un arcón, puede que lo preserve. En los digitales, la negligencia benigna no sirve: un disco olvidado 5 años... no sirve.



- No a la negligencia benigna.
- No a la preservación basada en las condiciones ambientales.
- No se conserva para cualquier usuario futuro sino para una comunidad designada: el conjunto de los consumidores que tienen que entender la información almacenada.
- No necesariamente se conserva la integridad externa del documento sino las propiedades significativas.
- Se debe asegurar la integridad y autenticidad del recurso



# Problemas en la preservación de OD

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



# Preservación digital

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

**(Keefer; Gallart, 2007).**



# Etapas en la preservación

## 1. Archivar los documentos digitales

gestión documental

## 2. Preservar el *bitstream*

## 3. Garantizar el acceso a largo plazo



# La preservación supone que:

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



# Preservación digital

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



# Noción de preservación de UNESCO



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

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



# Objeto digital

## Acciones en su ciclo de vida para mantener el acceso

### OD Y METADATOS DE PRESERVACIÓN

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

Deben guardarse las relaciones que vinculen al objeto con otros

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

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



#### Autenticidad

Mediante la documentación de su procedencia

Debe conocerse su **creador**

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

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

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



“UNA METODOLOGÍA DE EVALUACIÓN DE REPOSITORIOS DIGITALES PARA ASEGURAR LA PRESERVACIÓN EN EL TIEMPO Y EL ACCESO A LOS CONTENIDOS”

Autora: Ing. Marisa R. De Giusti

Directora: Dra. Silvia Gordillo



# Preservación de los contenidos de un RI

Criterios nuevos para los recursos digitales:

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



# Metadatos y metadatos de preservación

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



# Preservación del contenido de los RIs

¿Qué materiales hay en los RIs?

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

¿Qué materiales se tienen que preservar a largo plazo?

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

· Otros condicionantes: misión, disponibilidad de recursos humanos, económicos, materiales, obligaciones legales o contractuales.



# Preservación del contenido de los RIs

## Selección de recursos para su preservación

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

¿Qué atributos se quieren preservar?

datos, funcionalidad

apariencia, esencia

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



## Un avance: estándares

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

# Problemas en la preservación: software

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

# Preservación de contenido

- Hay una muy importante necesidad de preservar el contenido digital en el tiempo, con el objetivo de conservarlo accesible frente a riesgos como: incendios, inundaciones, robos, problemas de hardware (rotura de discos, etc.) y cambios tecnológicos constantes.
- *Es un proceso continuo*
- Además de lo técnico, los esfuerzos de preservación incluyen retos legales, económicos e institucionales.



# Obsolescencia

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



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

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



# Preservación de contenido. “Obsolescencia digital”

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.



# Preservación de contenido. Estrategias

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

- Migración
- 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

# Preservación de contenido

- Los riesgos de pérdida de datos por eventos desafortunados siempre son posibles.
- Para disminuir esos riesgos es necesario contar con un sistema de backups (datos, configuración, documentación, etc) como se hablara en clases precedentes.
- También es necesario elegir los formatos de acuerdo a los criterios que se mencionaron: uso de una gran comunidad, apertura, licenciamiento libre...

# Selección de formatos: generalidades

Para asegurar la preservación de la información, el formato elegido debe ser legible por una aplicación durante el mayor tiempo posible. Esto implica evitar formatos propietarios cerrados, como documentos en Word (en cualquiera de las versiones). Lo recomendable es utilizar formatos propietarios pero abiertos, como el formato TIFF, o mejor aún, formatos no propietarios como el ASCII.



# Selección de formatos: generalidades

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



# Selección de formatos : generalidades

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

# Selección de formatos : generalidades

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



# Sobre PDF/A

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



# Sobre PDF/A

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

## Otros elementos de la compatibilidad con PDF/A

El contenido de audio y video está prohibido.

Java script y enlaces a archivos ejecutables están prohibidos.

Todas las fuentes deben estar embebidas, y también deben ser legalmente embebibles para renderización ilimitada y universal. Esto significa para un usuario poder abrir el documento y que los caracteres se muestren de manera correcta (de aquí a X años) aunque no tenga esa tipografía en su computadora.

Los espacios de colores deben ser especificados de una manera independiente del dispositivo.

Se prohíbe la encriptación.

El uso de metadatos basados en estándares se mantiene.



## Otros elementos de la compatibilidad con PDF/A

Las referencias a contenidos externos están prohibidas.

La compresión de imágenes LZW y JPEG2000 están prohibidas en PDF/A1, pero JPEG 2000 se permite en PDF/A2.

Capas y objetos transparentes están prohibidos en PDF/A1 pero no en PDF/A2.

Firmas digitales provisionales se permiten en PDF/A2.

Los archivos embebidos están prohibidos en PDF/A1, pero PDF/A2 permite embeber archivos PDF/A.

PDF/A3 permite embeber cualquier formato como XML, CSV, CAD, archivos de Word, planilla de cálculo, otros PDF/A, etc. como objetos archivados completos.



## Niveles de cumplimiento

PDF/A posee dos niveles de cumplimiento:

**PDF/a** aplica corrección semántica y estructura. Cada carácter debe tener su equivalente Unicode. La estructura se expresa por medio de etiquetas.

**PDF/b** aplica integridad visual.



# National Archives of Australia



Australian Government



NATIONAL ARCHIVES OF AUSTRALIA

Explore:

Search the collection	Go >
Search the website	Go >

Your story, our history

Home
The collection ▾
Records management ▾
Visit us ▾
About us ▾

[Home](#) > [About us](#) > [Our organisation](#) > [Accountability and reporting](#) > [Operations and Preservation policies](#)

Our organisation

Advisory Council

Leadership team

Accountability and reporting

Access to information under the Freedom of Information Act

Annual Reports

Budget

Contracts

Corporate Plan

Departmental indexed file list

Grants

Information Publication Scheme

## Digital Preservation Policy

**Preserving Archival Digital Records Transferred from Commonwealth Agencies**

July 2009

**Version**  
1,2, July 2011

**Contents**

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

# Preservation file formats. Australia

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

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

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



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



Format Category	Preferred Preservation open file format <sup>1</sup> Files in these formats are converted to our preferred preservation open file format (preserved as they are)	Acceptable open file format <sup>2</sup> (Not converted, preserved as they are)
	Plain text in Unicode or ASCII.	<ul style="list-style-type: none"> <li>•</li> <li>•</li> <li>•</li> </ul>
	<p>Not yet decided.</p> <p>Video files are currently preserved as is.</p>	<ul style="list-style-type: none"> <li>•</li> <li>•</li> <li>•</li> <li>•</li> </ul>

- Flash Movie file (swf)
- Motion JPEG/JPEG 2000
- RealVideo (rv, rmvb)
- Windows Media Video (wmv)



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



## Preservation Formats for Electronic Records Smithsonian

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

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

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



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

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



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



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



# University of Texas



Digital Repository

## Recommended File Formats

### Recommended File Formats for Preservation

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

#### Textual Formats

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

#### Image/Graphic Formats

JPEG  
JPEG2000  
PNG  
SVG 1.1 (no Java binding)  
TIFF

#### Audio Formats

AIFF  
WAVE

#### File Extensions

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

#### File Extensions

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

#### File Extensions

.aif, .aiff  
.wav

[https://repositories.lib.utexas.edu/recommended\\_file\\_formats](https://repositories.lib.utexas.edu/recommended_file_formats)



# Algunas recomendaciones en digitalización

Como guía en procesos de digitalización pueden usarse:

"Technical Guidelines for Digitizing Cultural Heritage Materials" generado en 2010 por la Federal Agencies Digitization Guidelines Initiative (FADGI).

"Directrices para proyectos de digitalización de colecciones y fondos de dominio público" , IFLA (2002).

"Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files - Raster Images" , NARA (2004).

“Recomendaciones para la digitalización de los documentos en archivos”. Junta de Castilla y León (2011).



# Algunas recomendaciones en digitalización

De cada obra digitalizada es recomendable tener al menos 3 versiones:

- a) imagen maestra resultante del escaneo,
- b) documento maestro con imagen derivada de la anterior y el OCR generado sobre ella y
- c) otra(s) derivada(s) a los fines de uso y distribución en la web.



# Enlaces de interés

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

Directrices IFLA para digitalización

Recomendaciones para la digitalización de archivos. Junta de  
Castilla y León



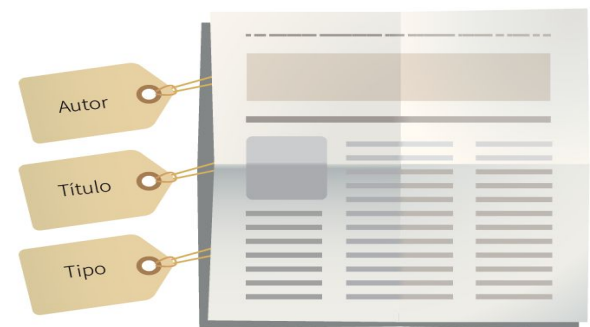
# Formatos. ¿Cómo conocer lo que tiene un RI?

**Perfilamiento automatizado de los objetos del repositorio:** esto involucra al objeto de contenido (CDO) con sus propiedades significativas y a la información de representación de ese objeto (RI). Realizar el perfil con DROID que contrasta con el registro PRONOM y brinda un reporte.

El punto 1 es una de las 3 partes que se consideran importantes a la luz de cumplir con la ISO 14721 y realizar una evaluación del repositorio en los aspectos de preservación y accesibilidad

# ¿Qué acciones se proponen?

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



Información descriptiva  
(DI)

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



# Resumen PD en RIs

Regulación de todos los procedimientos.

Regulación de los derechos de preservación digital sobre los documentos.

Regulación de los formatos admisibles.

Control de formatos en la ingestión.

Formatos de visualización y de preservación

Almacenaje de metadatos técnicos.

Copias sistemáticas externas.

Creación de procedimientos de contingencia ante desastres.

Auditoría interna/externa de seguridad.

Plan de preservación...



# Bibliotecas y repositorios digitales

## Capítulo 8:

La preservación en el repositorio institucional: el Modelo OAIS ISO 14721. Comparación con las facilidades que ofrece DSPACE.



# Estándares

El estándar 14721 (OAIS), los metadatos PREMIS y las directrices para la preservación, 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: 2012

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



# El Modelo OAIS

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



# El modelo de Referencia OAIS

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



# 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.



# Antecedentes

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



# Funciones del Modelo de referencia

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

# Propósito y campo de Aplicación

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



# Propósito y campo de Aplicación

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

# Sección 2



# Conceptos en OAIS

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

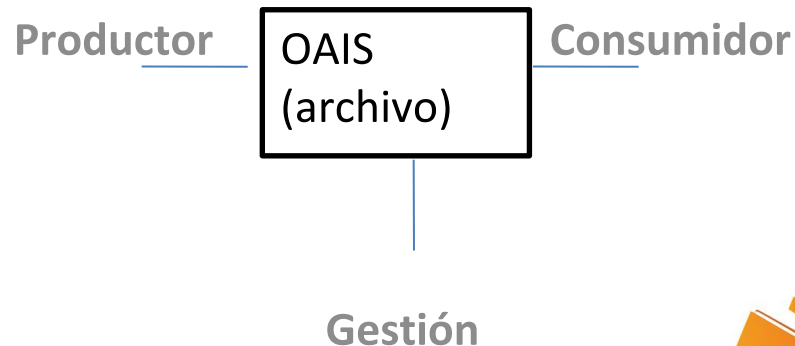


# Medioambiente OAIS

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

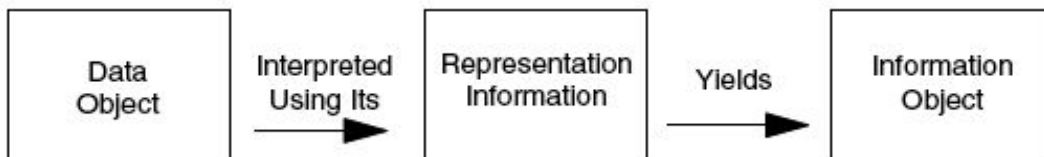
## Actores en el modelo

- **Producer-Consumer-Management**



# Conceptos en OAIS

Una definición clara de información es central para la capacidad del OAIS para preservar esa información. Una persona o un sistema, tienen una base común de conocimientos (KB) que le permite comprender la información. Se considera información en este campo a cualquier tipo de conocimiento que puede intercambiarse y que se expresa a través de algún tipo de datos: la información en un artículo periodístico, se expresa por caracteres (datos), los cuales bajo el paraguas de un lenguaje (KB), se convierten en información relevante. Si el receptor desconoce la lengua, entonces el artículo tendrá que ser acompañado por información extra, por ejemplo, un diccionario o una gramática.



## Conceptos en OAIS

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

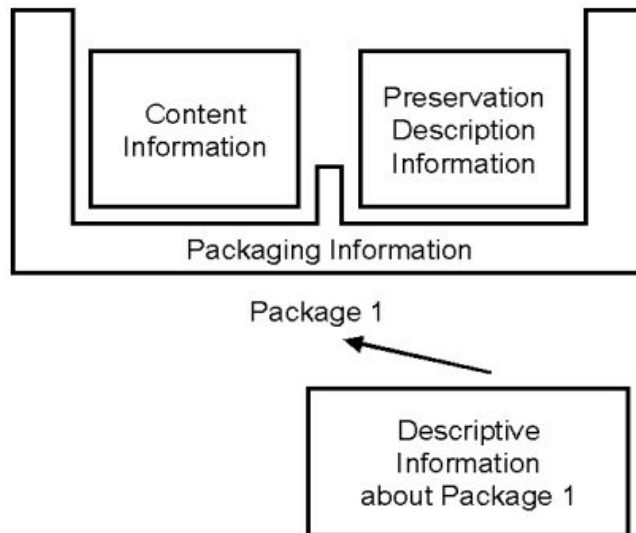


## Conceptos en OAIS

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



# Conceptos en OAIS



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



## Conceptos en OAIS

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



# Estructura del Paquete de Información



## El paquete de información -IP

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



## Elementos de la PDI

La **procedencia**, más allá de describir la fuente, incluye los procesos que se han realizado sobre la información: la historia del objeto, cambios, versiones y responsables. El **contexto** muestra las relaciones con otras fuentes de información o contenidos. La **referencia** provee una identificación única del contenido. La **integridad (o fijeza)** provee una protección para que la información no sea alterada de manera intencional /no. Los **derechos de acceso** proveen información sobre los términos de acceso incluyendo preservación, distribución y uso de la información de contenido.

# Conceptos en OAIS

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



## Clases de IPs según su función

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



## Clases de IPs según su función

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



## Clases de IPs según su función

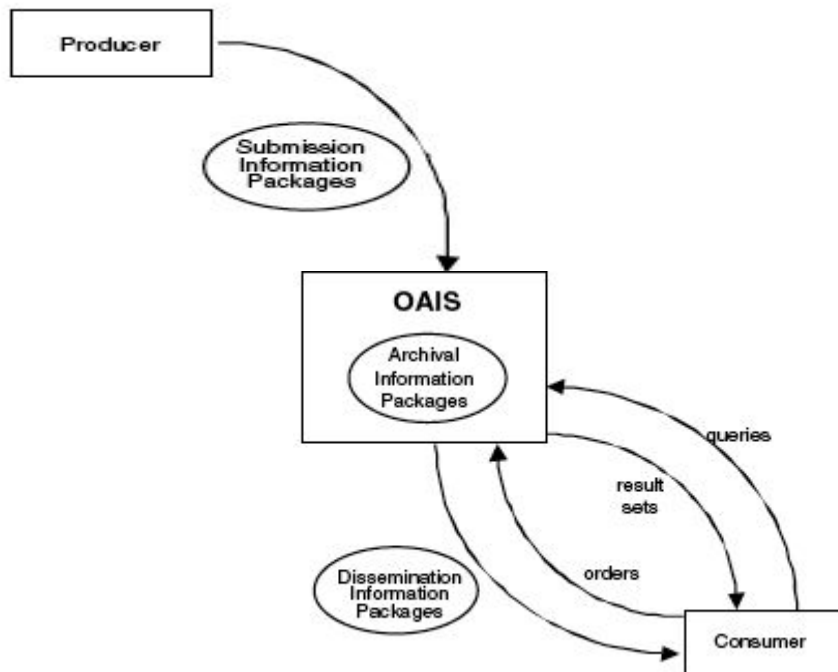
Dissemination Information Package (DIP): es el paquete que se entrega a un consumidor en respuesta a una solicitud. La información de empaquetado toma muchas formas dado que los usos de OAIS son diversos, puede ser tan completo como los AIP a partir de los cuales se construye o ser sólo una breve descripción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



# OAIS interacciones externas de alto nivel

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

# OAIS interacciones externas



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

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



# Sección 3: Responsabilidades del OAIIS



# 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.



# OAIS Responsabilidades obligatorias

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



# Mecanismos de deslinde de responsabilidades

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



# Sección 4

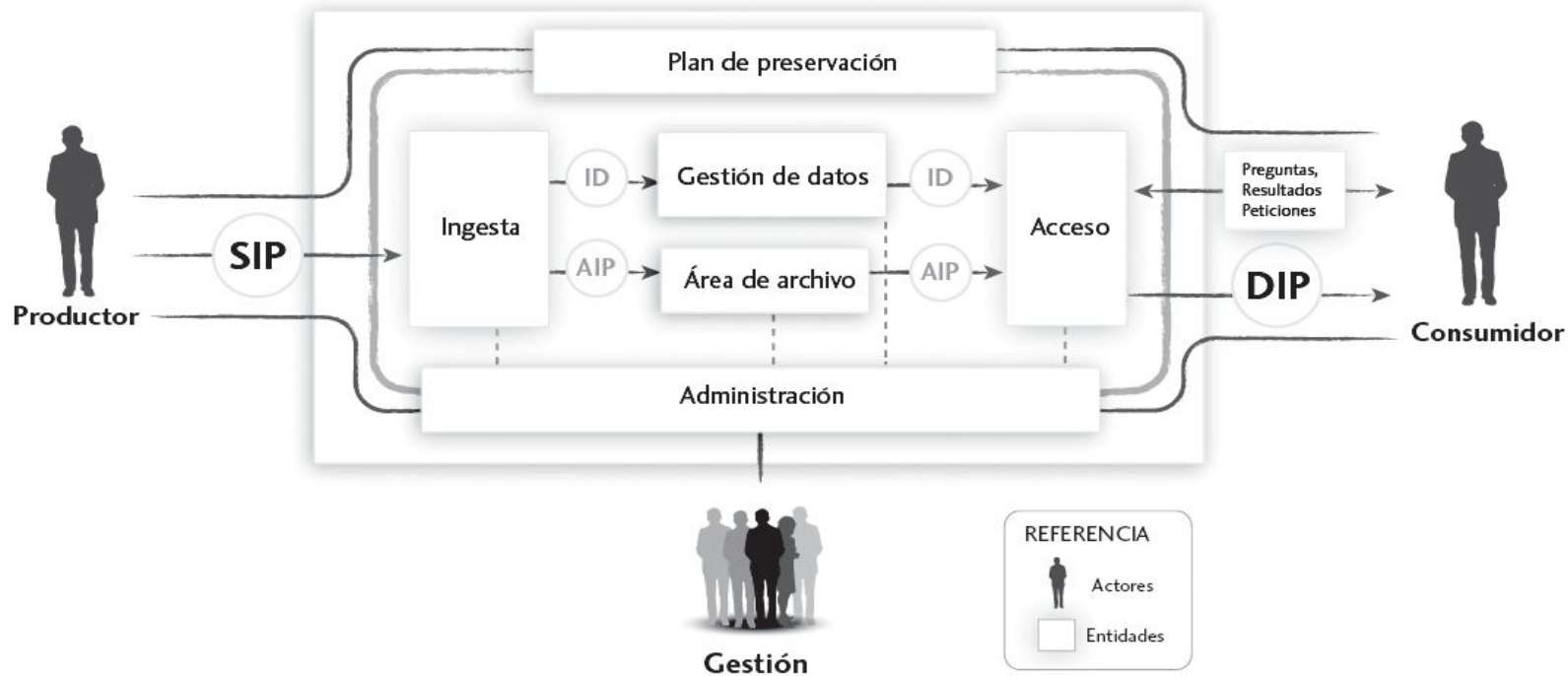


# OAIS

## 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



## Modelo OAIS

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

# Modelo OAIS

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



# Modelo OAIS

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

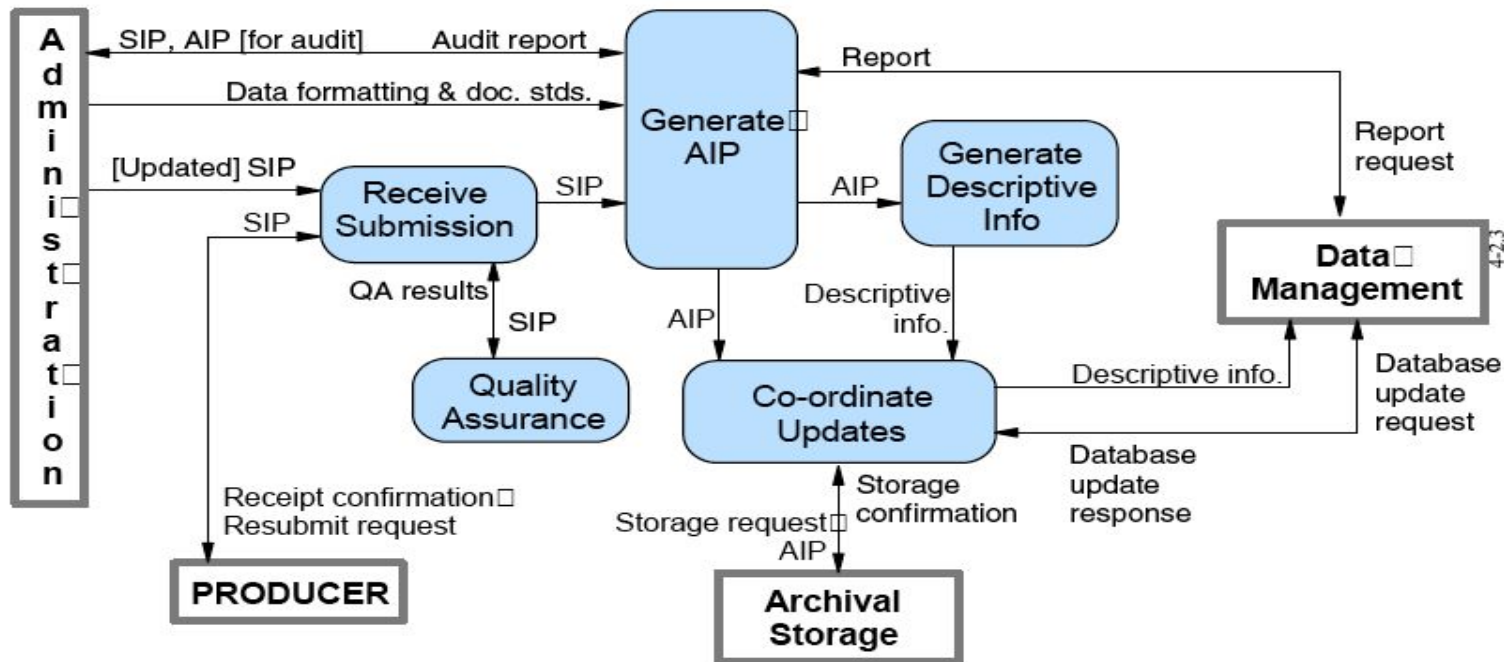


# Ingesta/Ingest/presentación

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



# Functions of Ingest

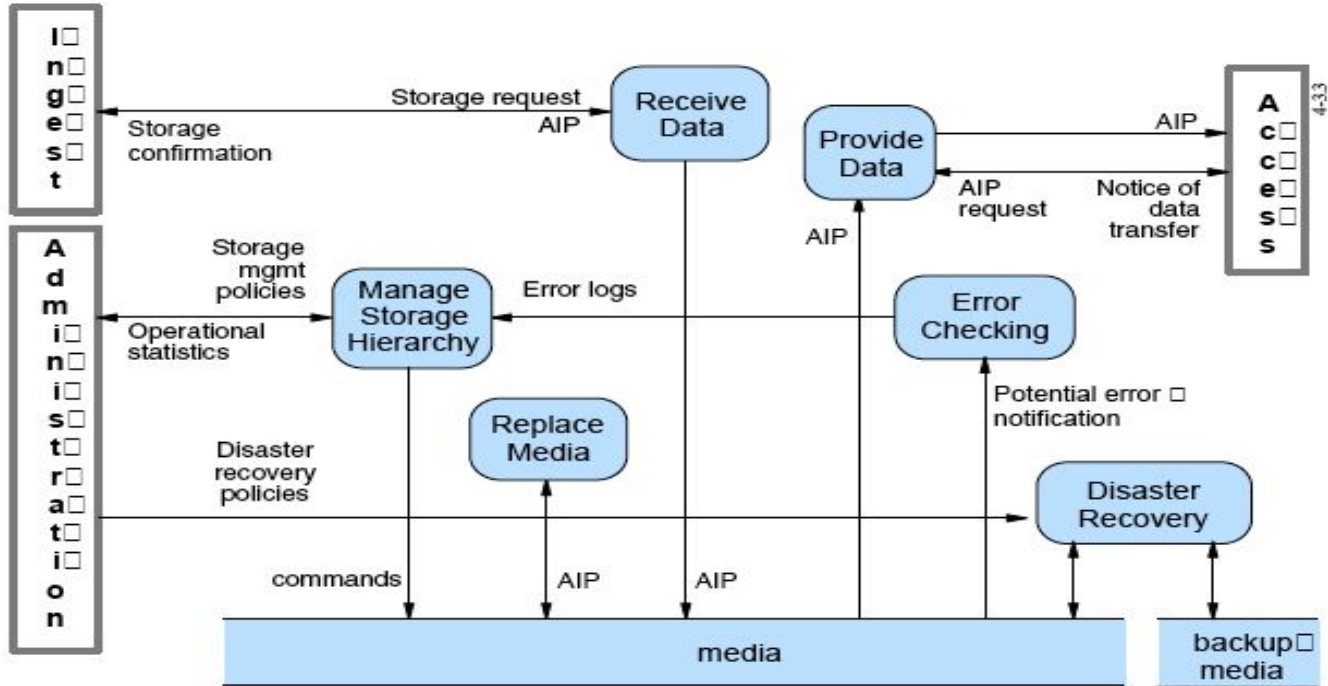


# Entidad OAIS Ingest

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



# Functions of Archival Storage

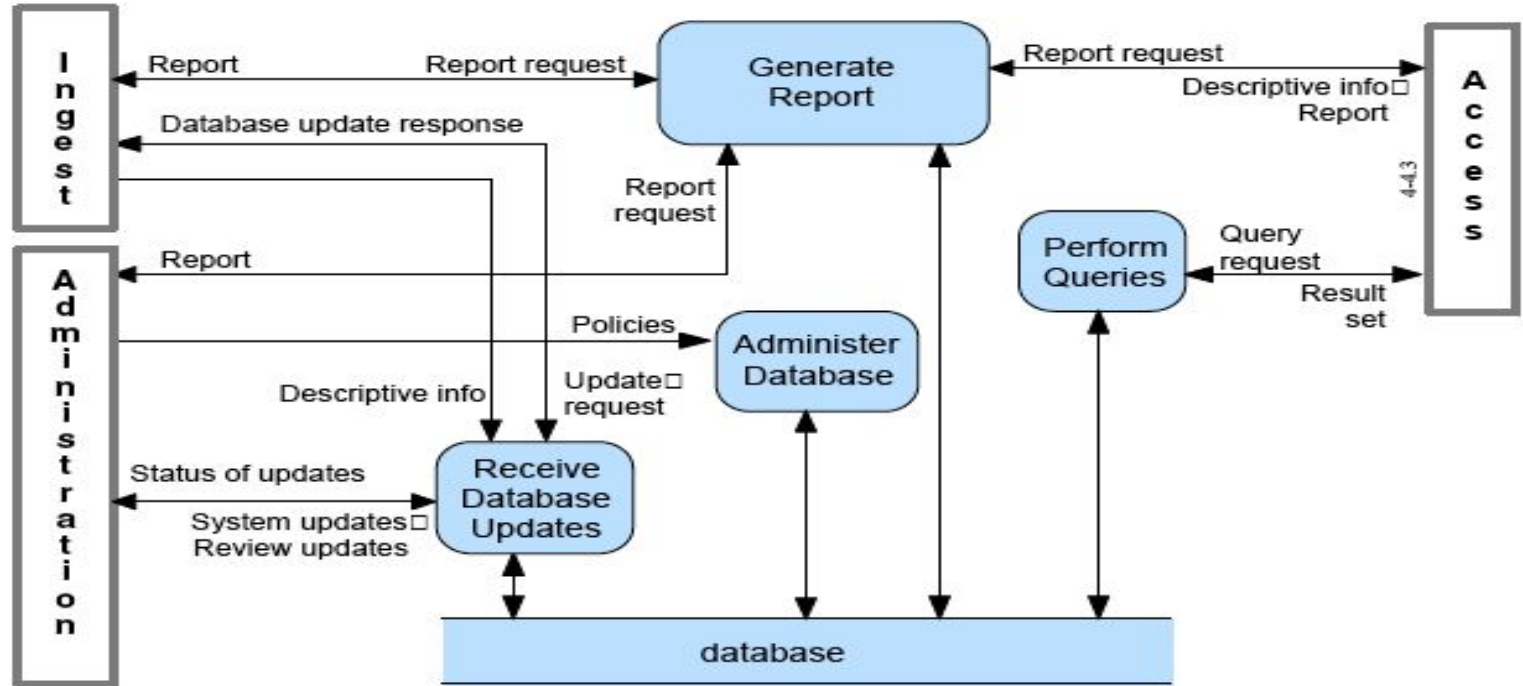


# Entidad OAIS Archival Storage

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



# Functions of Data Management

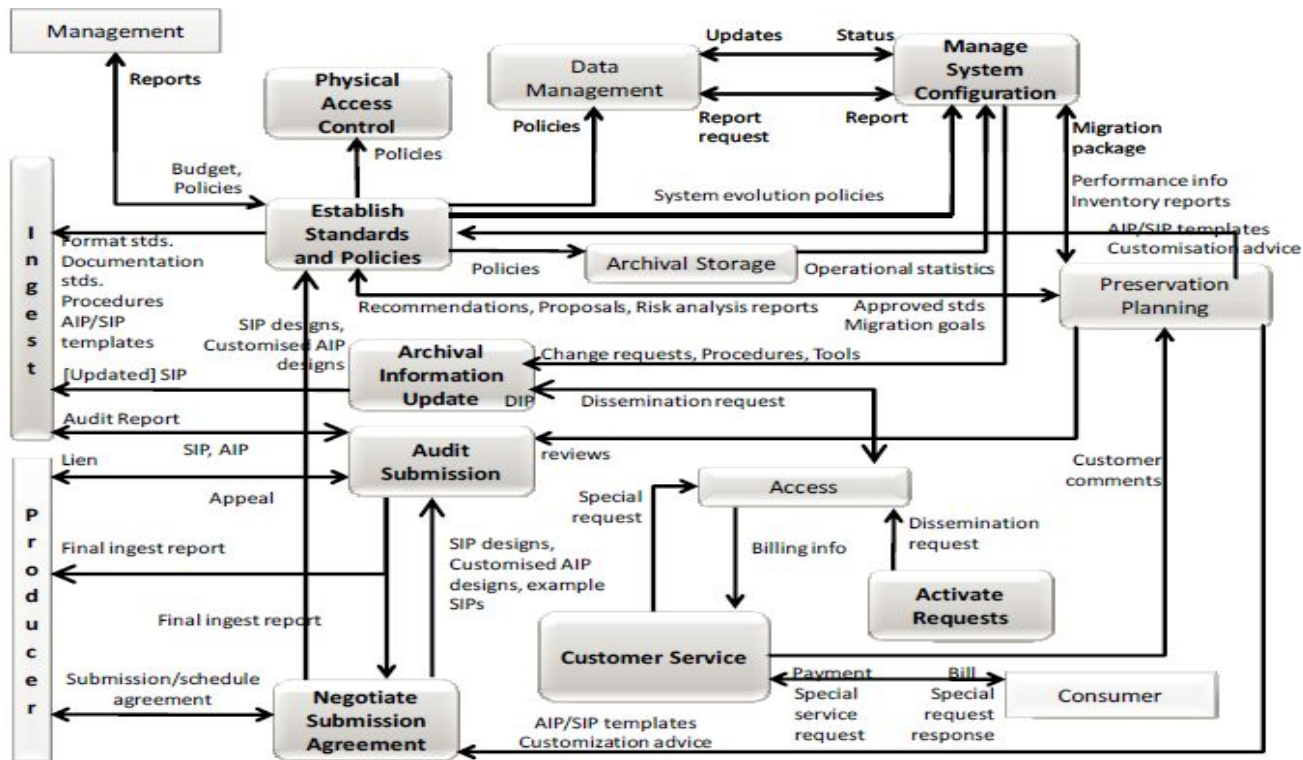


# Entidad OAIS Data Management

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



# Functions of Administration



# Entidad OAIS Administration

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

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

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



# Entidad OAIS administration

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

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

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

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



# Entidad OAIS administration

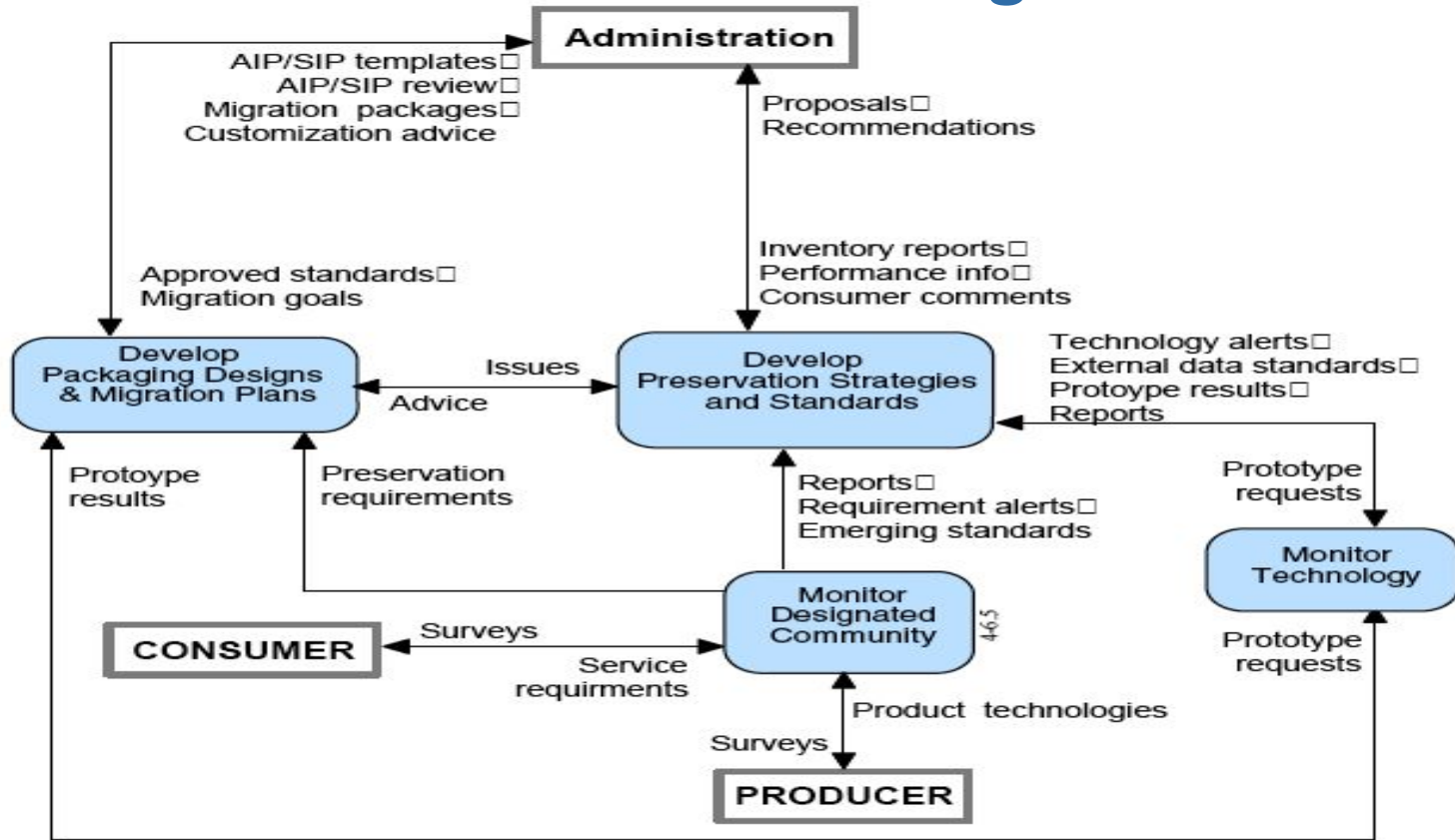
Verifica que los AIP y SIP suministrados sigan las especificaciones. En el caso de SIP y de AIP verifica la comprensión por parte de la comunidad designada. Verifica que la Información de representación y la PDI son adecuadas y comprensibles para la comunidad designada.

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

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



# Functions of Preservation Planning



# Entidad OAIS preservation planning

**Descripción:** Interactúa con los consumidores y productores de archivos. Proporciona reportes, alertas de requisitos y estándares independientes.

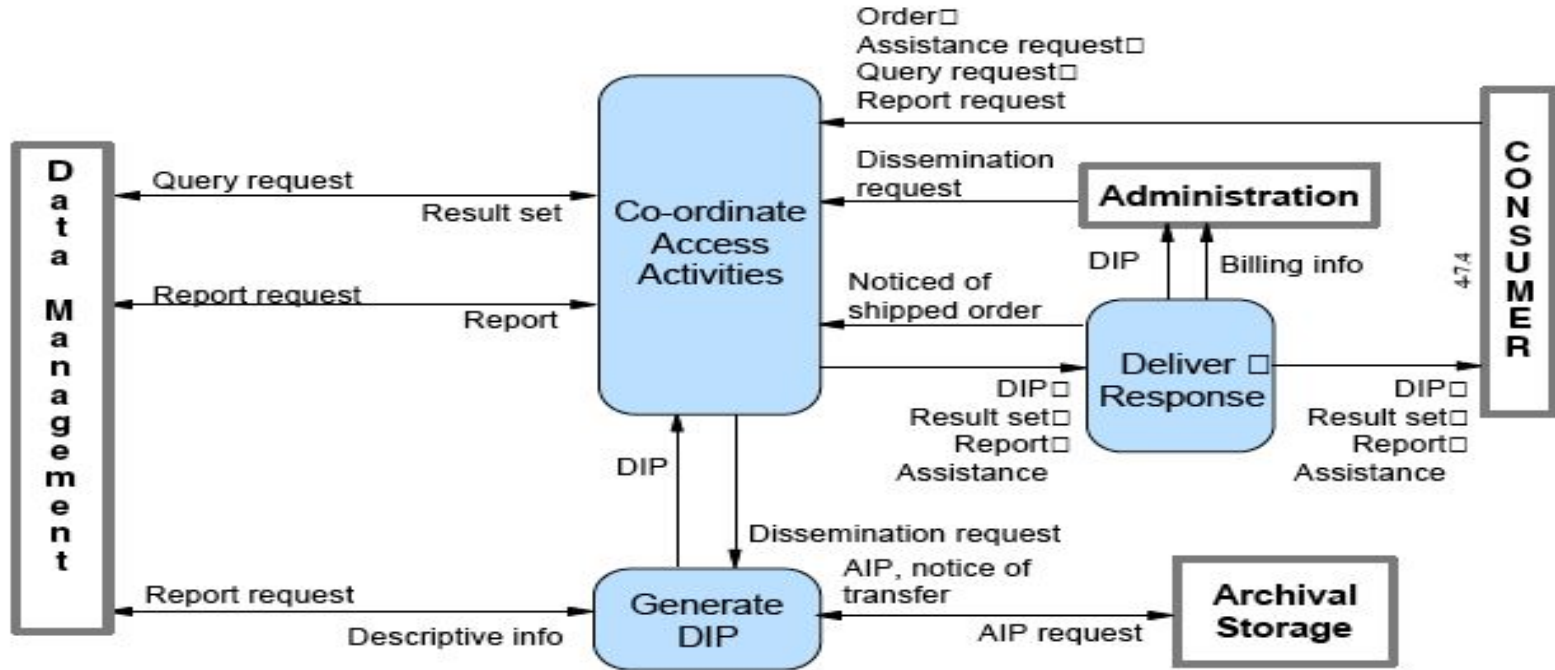
Identifica tecnologías que pueden causar obsolescencia.

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

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



# Functions of Access



# Entidad OAIS Access

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

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

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



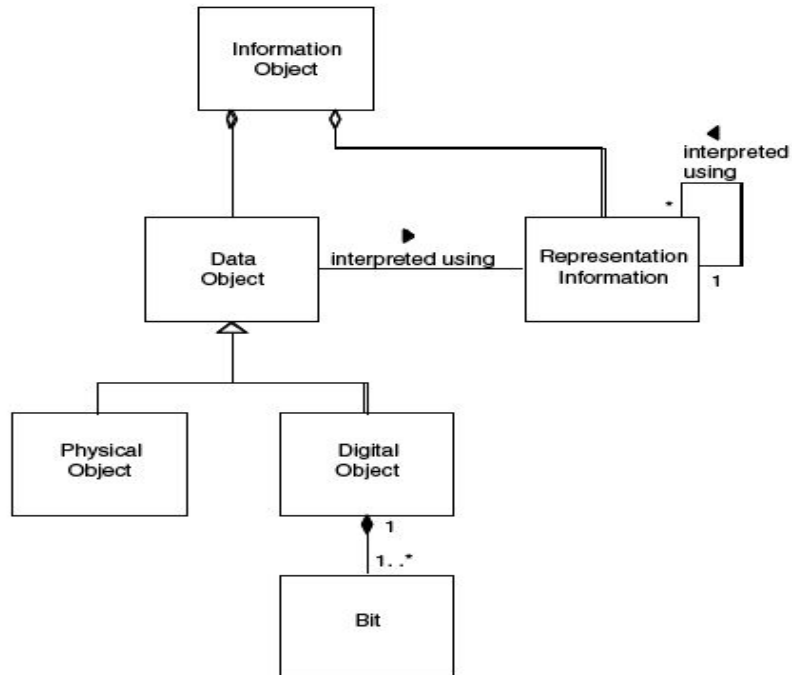
# OAIS

Modelo de Información

Sección 4.2



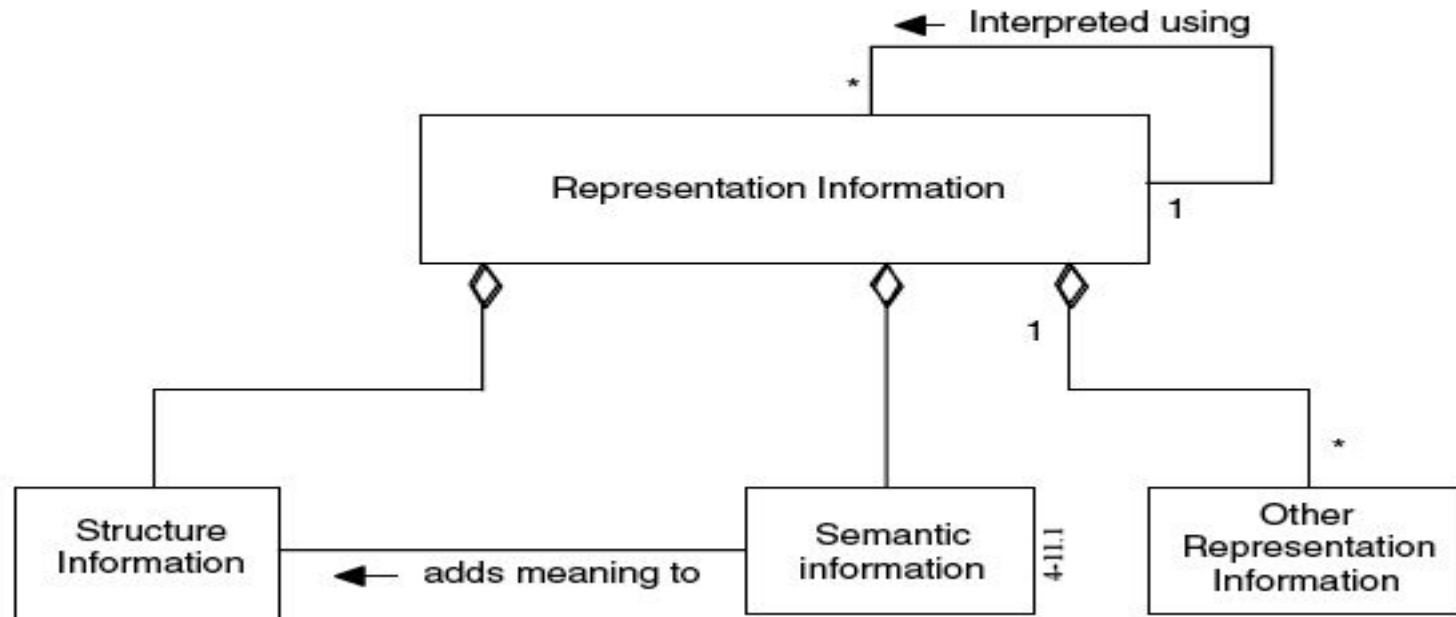
# OAIS Objeto de información



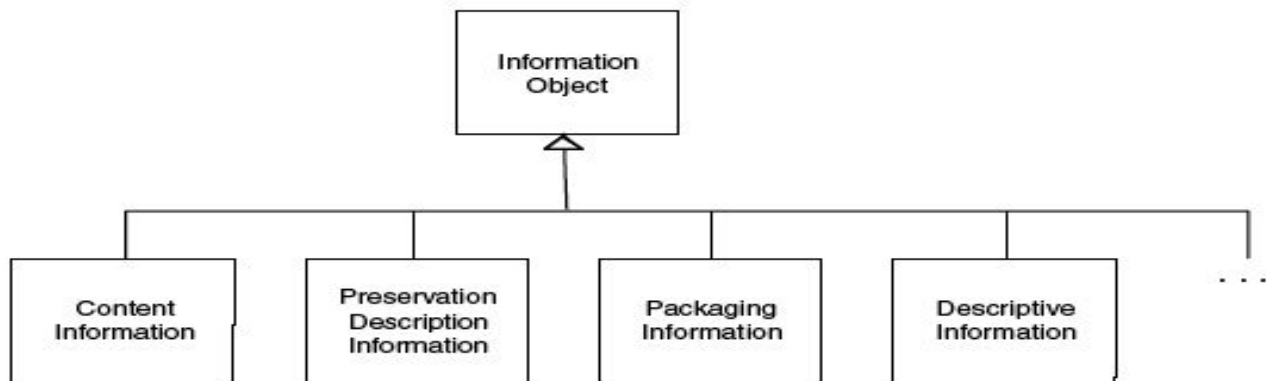
- El **Objeto de Información** está compuesto de un Objeto de Datos, que puede ser físico o digital, e Información de Representación que permite la interpretación completa de los datos.



# Representation Information Object



# Tipos de objetos de información



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

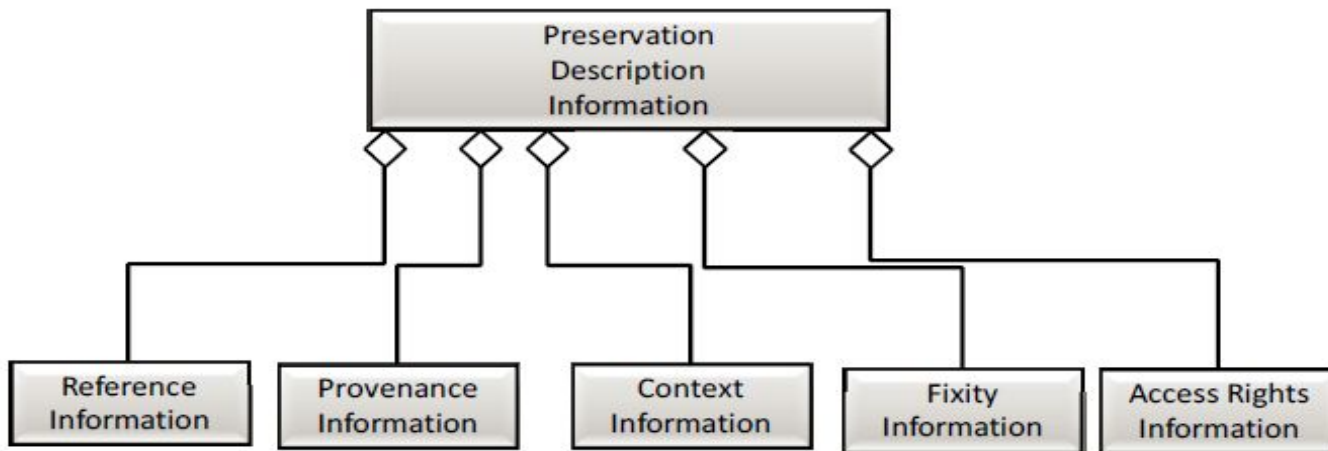


# Información de contenido

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



# Información descriptiva de preservación (PDI)



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



# Información descriptiva de preservación

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

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

# Información descriptiva de preservación

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

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

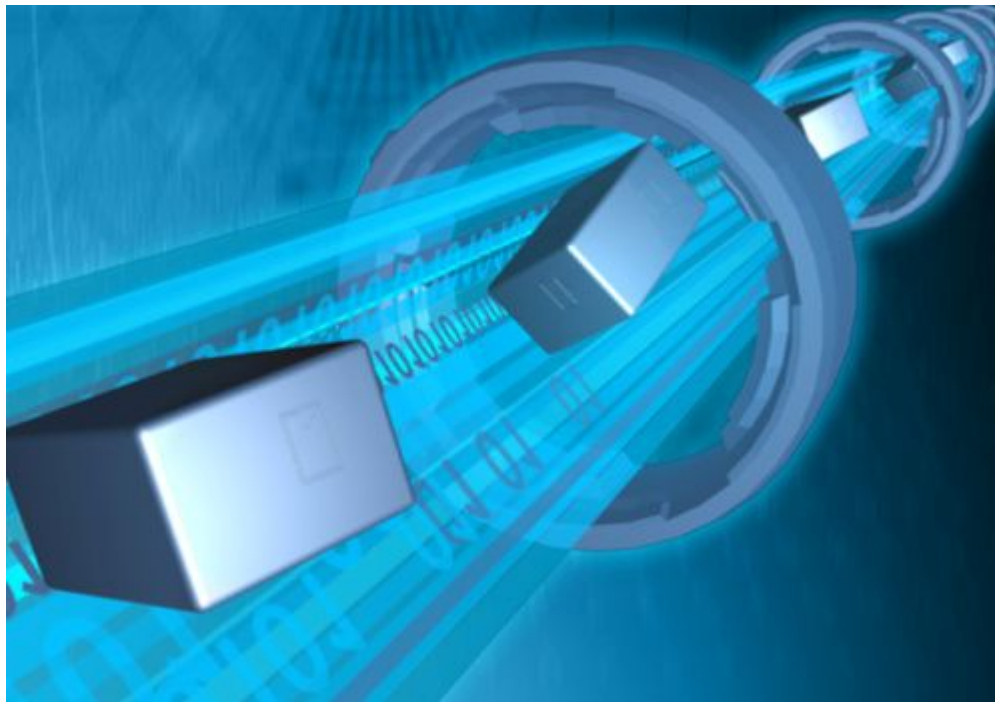
**Información de sobre derechos de acceso:** proporciona los permisos de uso de la información de contenido.

# Paquetes de información en OAIS

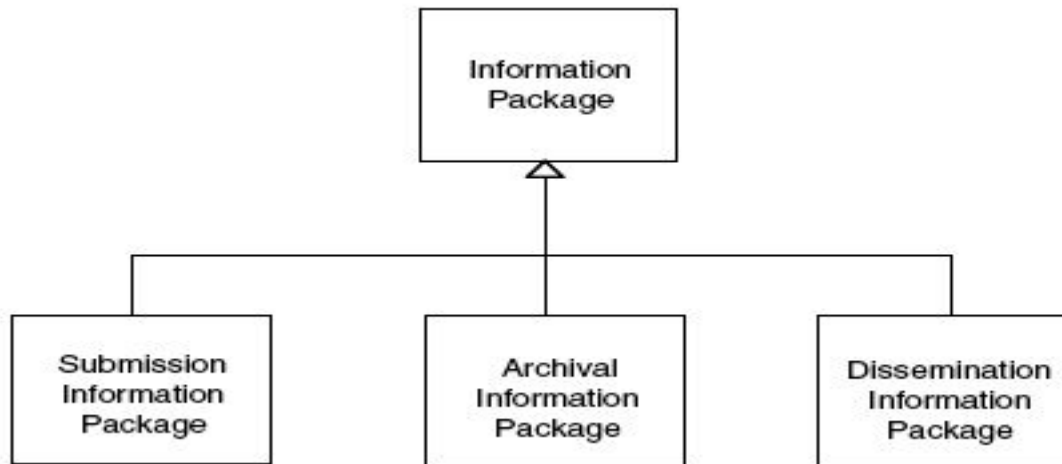
- Las estructuras de información conceptual necesarias para cumplir las funciones OAIS.
- Un paquete de información es un contenedor.
- Hay varios tipos de paquetes de información que se utilizan en el proceso de archivo. Estos paquetes de información pueden ser utilizados para:
  - estructurar y almacenar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



# Paquetes de información en OAIS



# Tipos de paquetes de información



# SIP

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

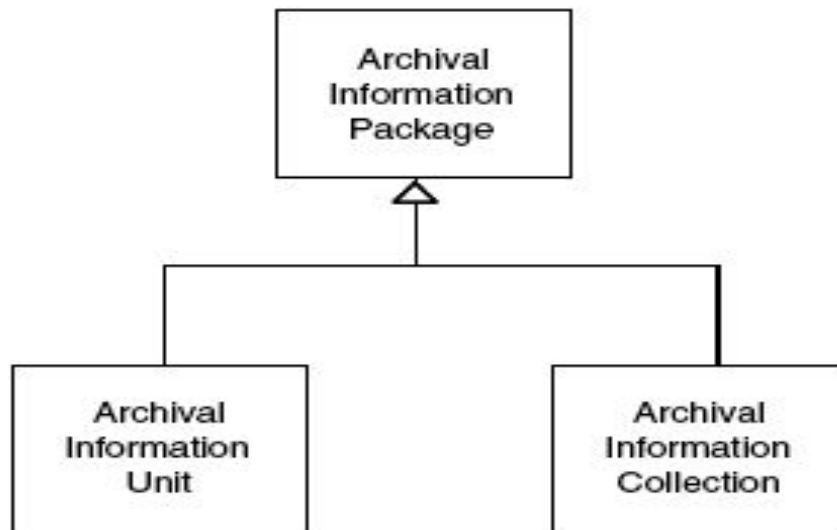


# AIP

Un Paquete de Información de Archivo es una especialización del Paquete de Información. El PIA se define para proporcionar una forma concisa de referirse a un conjunto de información que tiene, en un principio, todas las cualidades necesarias para una Conservación a Largo Plazo de un determinado Objeto de Información, de forma permanente o indefinida. El PIA es en sí mismo un Objeto de Información que contiene otros Objetos de Información.



# 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.
- OASIS 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

# Preservación en el repositorio

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

# Metadatos

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

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

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



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

# Metadatos de preservación

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



# PREMIS

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

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

# Diccionario de datos PREMIS

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

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



# Modelo de Datos PREMIS

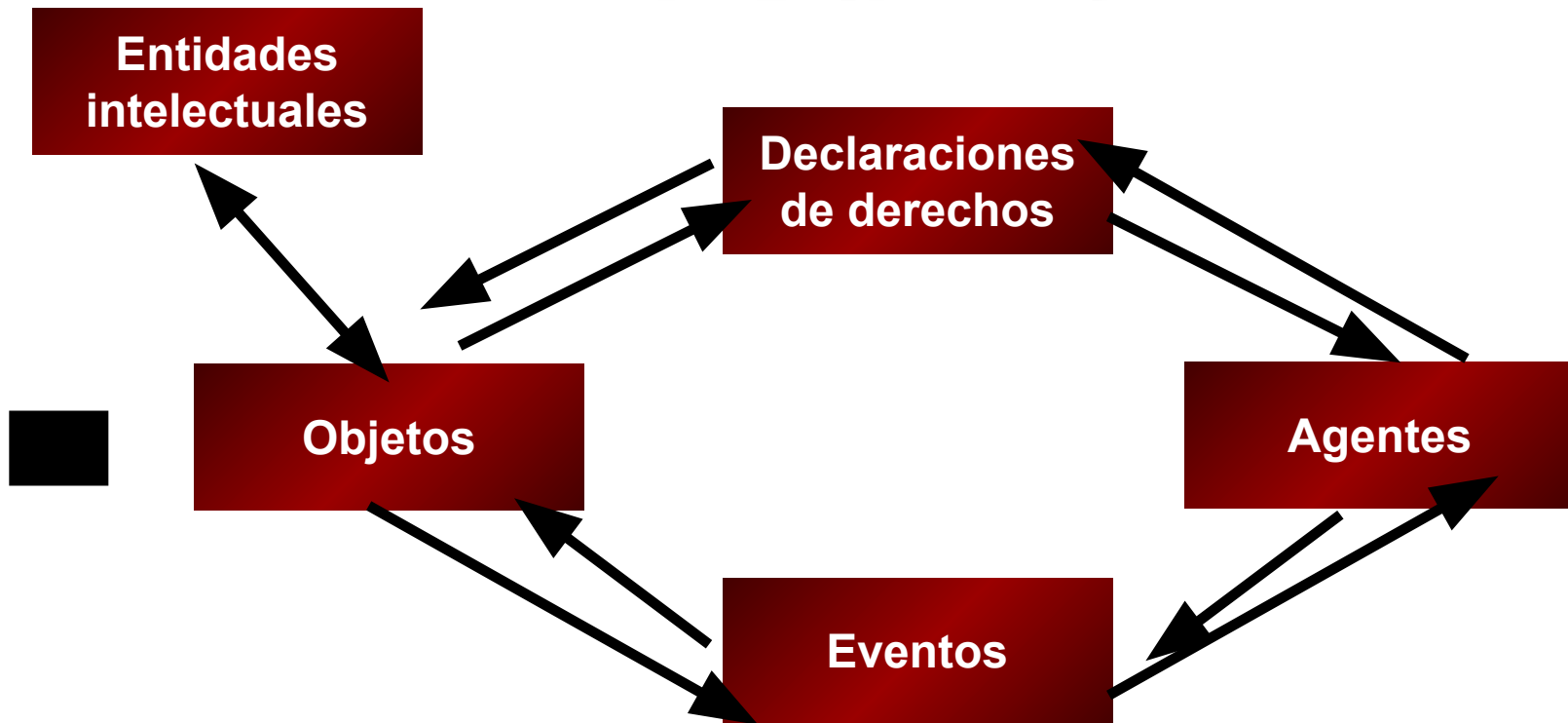
Las entidades que este modelo define se denominan:

- Entidades intelectuales
- Objetos
- Derechos
- Agentes
- Eventos



## Modelo de datos

PREMIS



# PREMIS

**Entidad intelectual:** conjunto coherente de contenido que se describe como una unidad: por ejemplo, un libro, un mapa, una fotografía, una publicación periódica,... etc. Una entidad intelectual puede incluir otras entidades intelectuales: por ejemplo, un sitio web, puede incluir una página web, una página web puede incluir una fotografía.

Una entidad intelectual puede tener una o más representaciones...

**Objeto** en PREMIS difiere de la definición de objeto digital normalmente utilizada en la comunidad de las bibliotecas digitales, que entiende el término “digital object” como una combinación de identificador+datos+metadatos. No es en absoluto un conflicto. La entidad objeto en el modelo de PREMIS es una abstracción definida sólo para agrupar atributos (unidades semánticas) y clarificar relaciones.



# PREMIS

**Evento:** Acción que incluye al menos un Objeto Digital y/o un agente conocido en el repositorio de preservación

**Agente:** Actor (humano, máquina o software) asociado con uno o más eventos asociados a un objeto digital

**Derechos:** Afirmación de uno o más derechos o permisos que pertenecen a un objeto digital y/o a un agente



# Entidad intelectual

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



# Objetos

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



# Eventos

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



# Agentes

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



# Derechos

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





## VIII Workshop

*Proyectos digitales*

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

# Metadatos para la preservación digital: PREMIS

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

emendez@bib.uc3m.es

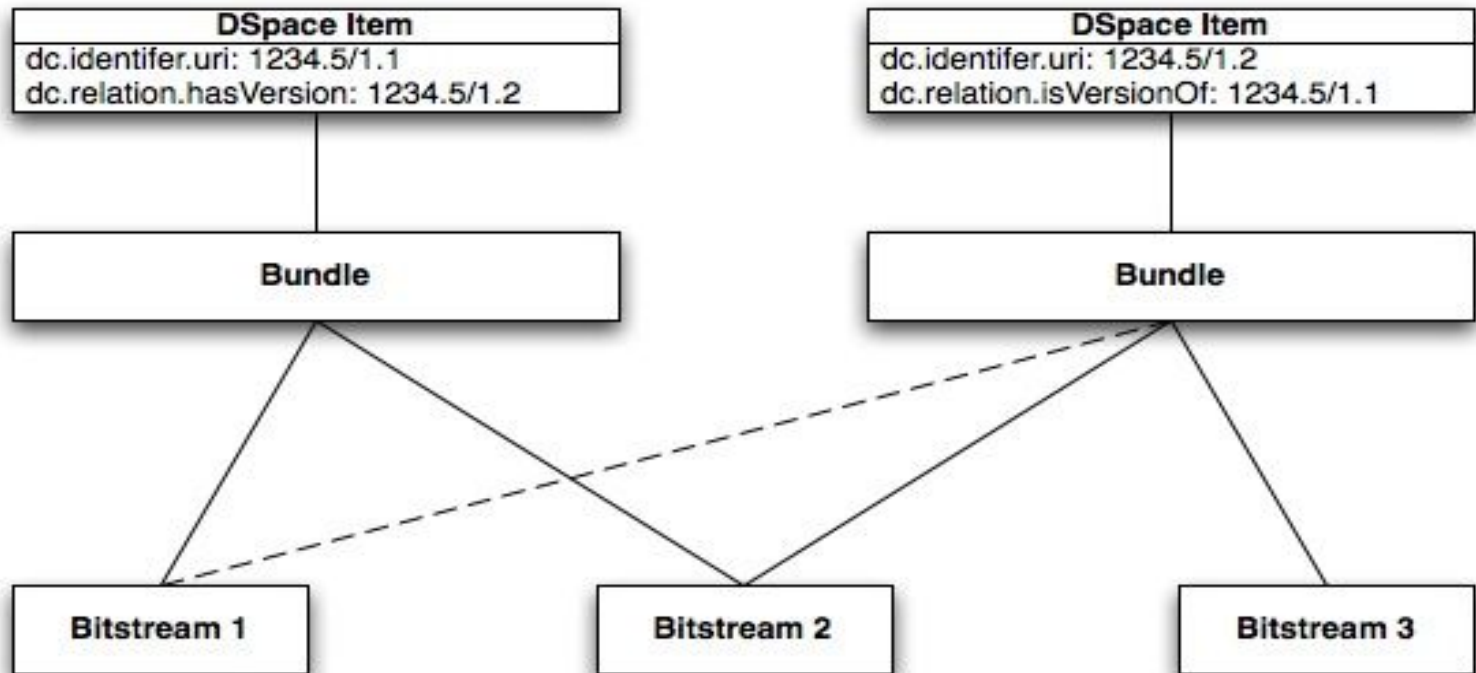
Dpto. Biblioteconomía y Documentación



Universidad Carlos III de Madrid



# DSPACE: Modelo de datos



# OAIS y DSPACE

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

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

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



# OAIS y DSPACE

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

# OAIS y DSPACE

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

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



# OAIS en DSPACE

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



# OAIS en DSPACE

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

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



# OAIS en DSPACE

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

# OAIS en DSPACE

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

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



# OAIS en DSPACE

## Tareas de Curación en Dspace

### Definición

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

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

### Tareas de curación en Dspace

Dspace proporciona herramientas para implementar distintas tareas de curación automáticas.



# OAIS en DSPACE

## Tareas de Curación en Dspace

### ¿Qué se puede realizar?

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



# OAIS en DSPACE

## Tareas de Curación en Dspace

### ¿Cómo se pueden realizar?

#### Dspace provee:

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



# OAIS en DSPACE

## Tareas de Curación en Dspace

### ¿Cómo se pueden realizar?

#### Se pueden implementar:

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



# OAIS en DSPACE

## Tareas de Curación en Dspace

### Ejemplos:

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



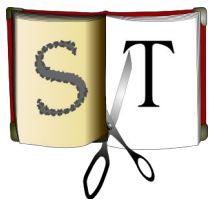


# Herramientas



# Herramientas

Una vez que se recibe material en el repositorio es necesario hacer un correcto relevamiento de los metadatos a partir de la obra, así como analizar los típicos problemas que pueden presentarse a partir de la variedad de materiales y formatos. A continuación se presentarán brevemente algunos casos y posibles soluciones que permiten llevar a cabo de un modo más simple la tarea de descripción documental.



**Video Downloader GetThemAll**

ofrecido por FVD Inc



Adobe Acrobat DC



sedici.unlp.edu.ar  
prebi.unlp.edu.ar

# Tratamiento y mejoras de archivos PDF

Suele ocurrir que el repositorio recibe archivos PDF de revistas o eventos que deben divididos para que cada artículo sea cargado por separado. Una de las opciones disponibles es el Acrobat DC.

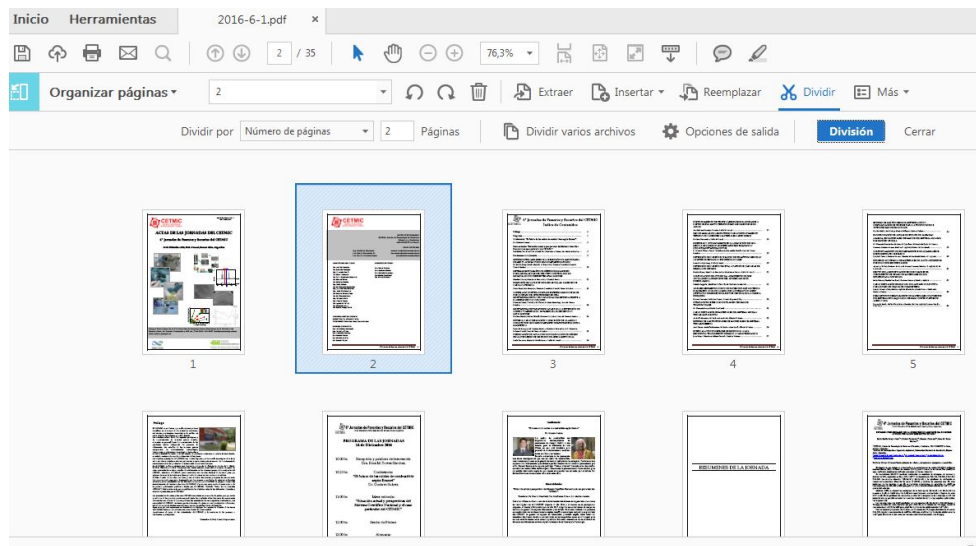


Adobe Acrobat DC



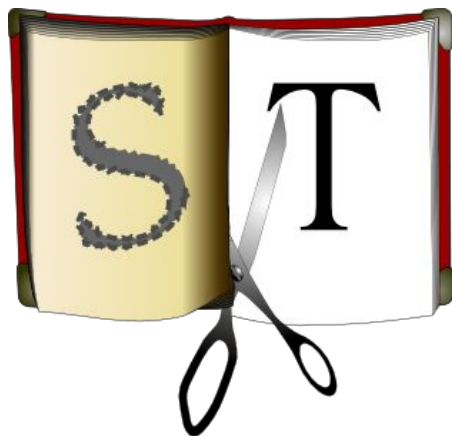
# Acrobat DC - División de Archivos

Este programa permite dividir por cantidad de páginas o por marcadores. En el caso de que el archivo no cuente con marcadores es posible agregarlos manualmente de manera fácil para poder dividir el archivo correctamente.



# Scan Tailor

Otra ocurrencia común es que lleguen pdfs con hojas escaneadas en modo apaisado o que los editores de una revista o un libro envíen la versión para imprimir. En este caso debemos contar con un programa que como el Scan Tailor permita dividir al medio las hojas y organizar el pdf de manera correcta.



# Scan Tailor

Se puede elegir manualmente el área correspondiente a cada hoja o el programa puede hacerlo de manera automática.

Archivos Herramientas Ayuda

- 1 Corregir Orientación
- 2 Dividir Páginas
- 3 Alineación
- 4 Selección de Contenido
- 5 Margenes
- 6 Salida

Debrucción de Pagina

Debrucción Automatica

Cambiar...

Línea Divisoria

Auto Manual

**1<sup>o</sup> Jornadas de Pasantes y Docentes de Escuela Media y Grado del CETMIC.**  
**16 de Noviembre de 2011, M.B. Corrientes, Buenos Aires, Argentina**

**CETMIC**  
**CARACTERIZACIÓN POR DRYS Y CARA ELECTRICA PERIFERICA DE SISTEMAS AG-Cu-Si/NOVOBLENDA**

Javier Lantaras\*, Mariela Fernandez† y Rosa Torres Sánchez†  
 †CETMIC, Centro de Tecnología de Recursos Minerales y Carbonácos, CIC-CONICET La Plata, Corrientes, Argentina  
 ‡Departamento de Química, Facultad de Ciencias Exactas UNLP

**Palabras Clave:** *caracterización y Pb-Monitoreo, Plata, Cobre, Uso urbano, hidrocloruro.*

Estudios previos, realizados en el CETMIC, demostraron la aplicabilidad de la Ag separada en monometaluros (Ag-MMT) como sistema de floculantes para reducir la contaminación de boratos en el agua. En este trabajo se estudia la especificidad de sistemas Cu-MMT y Ag-Cu-MMT para su uso en la actividad general del proceso, la de adsorción y floculación que brinda el Cu y Ag-Cu-MMT. Ag-Cu-MMT y Ag-MMT, para evaluar la inversión de los metales en el proceso de la MMT y los riesgos asociados significativos en el proceso por efecto de esta inversión. Los análisis de DRYS, permitieron determinar la relación de los distintos estados de oxidación de ambos metales. Su floculación fue estudiada por medio de análisis por AAS, en el sobrenadante de las distintas separaciones.

El momento del pico de reflexión (001) (Fig. 1), respecto de la muestra inicial MMT (13.42 Å), evidenció la incorporación de los cationes en la estructura de la MMT.

Figura 1. Espectro de DRYS del pico de reflexión (001).

Figura 2. Curvas de potencial ante vs pH.

Las curvas de potencial ante, en el rango de pH estudiado (pH= 5-7), muestran valores más negativos, para las muestras caracterizadas con el metal Ag o Cu, potencial  $E_{ante} = -47$  mV que las obtenidas para MMT (30 mV), mientras que las muestras con ambos metales (Ag-Cu-MMT y Ag-Cu-MMT) MMT) dieron valores intermedios entre las anteriores, la muestra MMT evidenció la incorporación de ambos metales en su superficie externa. Los espectros de DRYS, permitieron la incorporación de Ag y Cu en la estructura de la MMT, con presencia simultánea de Ag<sub>2</sub>O y Cu<sub>2</sub>O tanto a sus especies oxidadas, en relación cercana a 1:5 y un menor desplazamiento de la Ag que el aumento de concentración de Cu en las muestras nuevas. El análisis de difracción de Ag-Cu-MMT, indicó una menor liberación de Ag que Cu, en todos los estados. La presencia de cationes en el agua de los análisis, puede generar una menor liberación de Ag y Cu de la MMT.

**Prólogo**

El CETMIC es un Centro que realiza investigaciones científicas en el campo de los materiales avanzados, desarrollando y aplicando tecnologías de punta, en otros entornos investigados para la industria. Los estudios sobre átomos de los aspectos tecnológicos y de caracterización de materiales nuevos (metales, aleaciones) son generados bajo la supervisión de los profesores (Bases) mediante los procesos de colaboración. Los estudios de los materiales nuevos comprenden la caracterización mineralógica, fisicoquímica y aplicaciones tecnológicas destinadas a su uso industrial. Referente a la elaboración de productos nuevos, se estudia la síntesis, desarrollo, control de calidad y distribución de los productos.

Los objetivos principales del CETMIC son la investigación y el desarrollo tecnológico de todo lo referente a los materiales avanzados, innovadores y sus nuevas formas de uso. La transferencia de tecnología al medio productivo y la formación de recursos humanos.

El CETMIC, se lleva adelante desde el desarrollo y de grado de alumnos becados de la UNLP y la UBA. Actualmente, varios alumnos avanzados de cursos de grado de estas universidades realizan estudios de colaboración en los diversos grupos de investigación del CETMIC. Asimismo el CETMIC, posee convenios con escuelas medias de la zona, por estar quienes alumnos de los mismos realizan procesos de entrenamiento de sus propios estudiantes.

Con el Objetivo de continuar mejorando, y fortaleciendo los servicios y resultados de estas tareas, se decide realizar por los alumnos en que el CETMIC se propone realizar la Primera Jornada de Pasantes y Docentes de Escuela Media y Grado del CETMIC. Como parte comprendiendo el primer semestre de la presente participación pública dirigida por el CETMIC, titulada "Actos de las Jornadas CETMIC", Publicación en la que se publicarán los distintos actos de los diferentes Jornadas o eventos organizados por el CETMIC.

Los **Jornadas de Pasantes y Docentes de Escuela Media y Grado del CETMIC** consisten en una reunión de grupo para que todos y cada uno de los pasantes pueda compartir entre los resultados obtenidos en las experiencias vivenciadas con el resto de sus compañeros. Los materiales de sus respectivas instituciones, y la experiencia del CETMIC, se muestran para la mejor de las partes. Investigadores del CETMIC, tendrán un par de horas para el sistema Científico Argentino en Corrientes, del Centro de Tecnología de Recursos Minerales (CETMIC) en particular, con el objetivo de orientar a los Estudiantes las posibilidades de inversión laboral que presenta el sistema.

Esperamos que esta primera experiencia sea fructífera y multiplique los espacios de diálogo de las tareas del CETMIC, para pasar adelante sobre todo con el resto de la comunidad.

Agradecemos el apoyo de los autoridades del CETMIC, y la conformación de las instituciones participantes.

Noviembre de 2011, Corrientes Organizador.

**COORDINADORES**

**DRYSS TITULO**  
 Dr. Alberto Sola

**CAROLINA BARRERA TITULO**  
 Dra. Rosalva Lombardi

**DR. PABLO ALBERTO**  
 Dr. Pablo Albert

**DR. CLAUDIO GARCIA**  
 Dr. Claudio Garcia

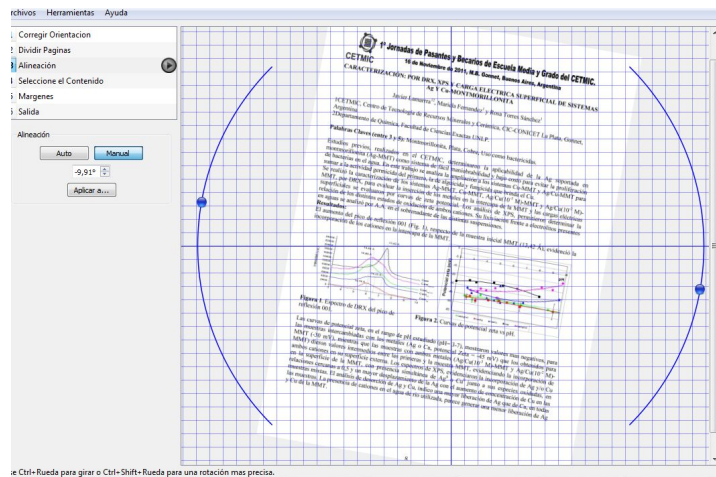
**COMITÉ ORGANIZADOR**  
 Dr. Pablo Sola  
 Dr. Carlos Lombardi  
 Dr. Nicolás Lombardi  
 Dr. Claudio Sola

3

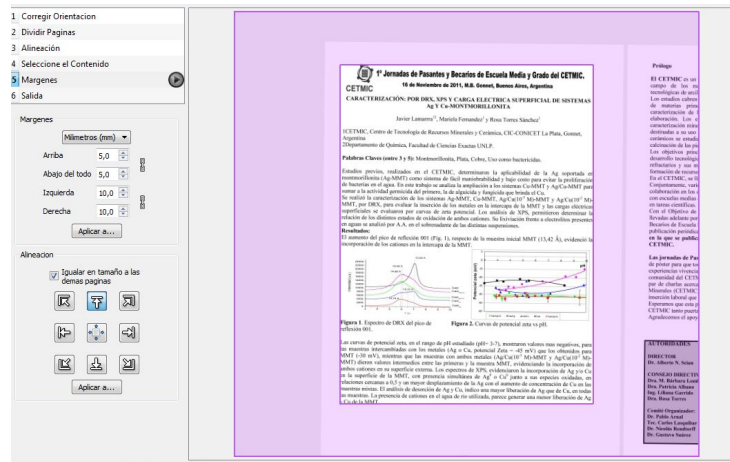
Use the mouse wheel or +/- to zoom. When zoomed, dragging is possible.

# Scan Tailor

También es posible corregir la orientación de las páginas, seleccionar los márgenes y la sección del contenido para realizar OCR y por último ajustar el color y contraste de las páginas.

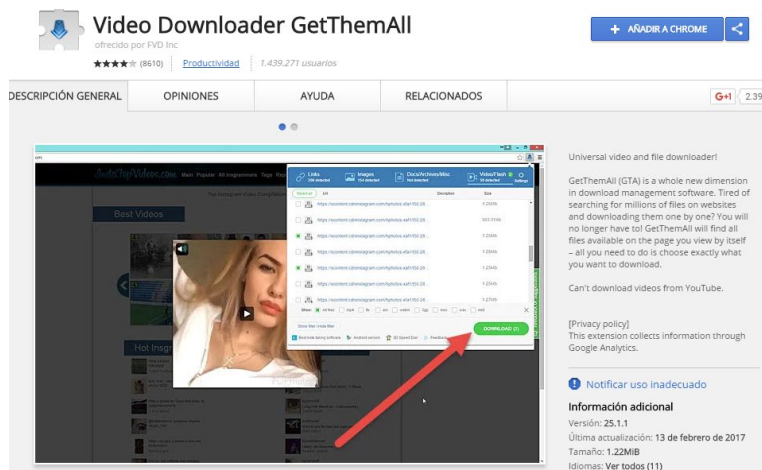


Ctrl+Rueda para girar o Ctrl+Shift+Rueda para una rotación más precisa.



# Descarga masiva de archivos PDF

Muchas veces ocurre que debemos descargar archivos PDF de manera masiva, por ejemplo, cuando vamos a cargar en el repositorio un volumen de una revista que se encuentra en una página web. Existen numerosas herramientas, en SEDICI utilizamos una extensión de Chrome llamada GetThemAll



**Video Downloader GetThemAll**  
otricado por FVD Inc.  
★★★★☆ (8610) | Productividad | 1.439.271 usuarios

DESCRIPCIÓN GENERAL OPINIONES AYUDA RELACIONADOS

Universal video and file downloader!

GetThemAll (GTA) is a whole new dimension in download management software. Tired of searching for millions of files on websites and downloading them one by one? You will no longer have to! GetThemAll will find all files available on the page you view by itself - all you need to do is choose exactly what you want to download.

Can't download videos from YouTube.

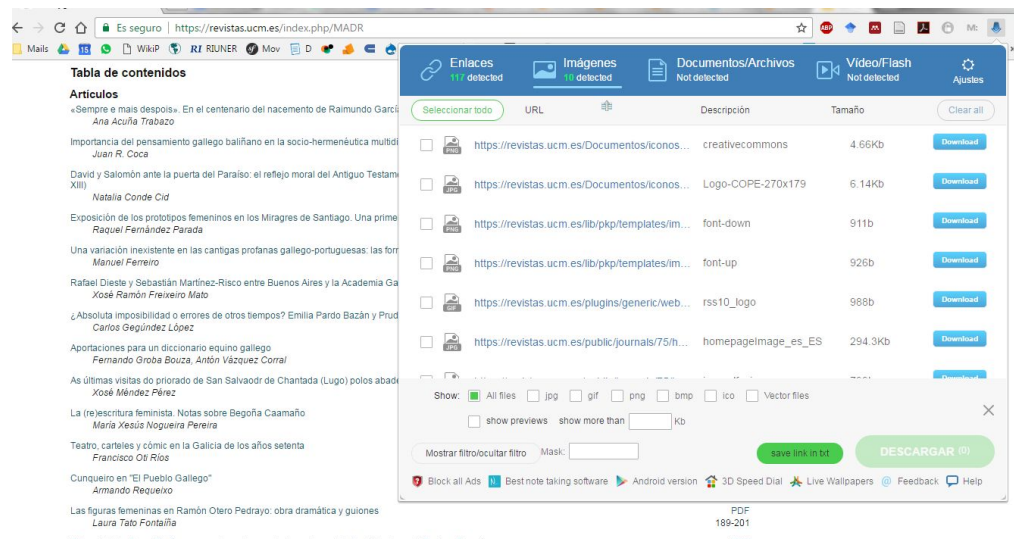
[Privacy policy]  
This extension collects information through Google Analytics.

Notificar uso inadecuado

**Información adicional**  
Versión: 25.1.1  
Última actualización: 13 de febrero de 2017  
Tamaño: 1.22MiB  
Idiomas: Ver todos (11)

# Descarga masiva de archivos PDF

La extensión GetThemAll detecta automáticamente los archivos pdf de una página web y permite la descarga en breves pasos. Ejemplo: [Journal of Computer Science and Technology](#).



The screenshot shows a web browser window with the URL <https://revistas.ucm.es/index.php/MADR>. The page content includes a table of contents with various articles. Overlaid on the page is the GetThemAll extension interface, which displays a list of detected items:

Seleccionar todo	URL	Descripción	Tamaño	Clear all
<input type="checkbox"/>	<a href="https://revistas.ucm.es/Documentos/iconos...">https://revistas.ucm.es/Documentos/iconos...</a>	creativecommons	4.66Kb	Download
<input type="checkbox"/>	<a href="https://revistas.ucm.es/Documentos/iconos...">https://revistas.ucm.es/Documentos/iconos...</a>	Logo-COPE-270x179	6.14Kb	Download
<input type="checkbox"/>	<a href="https://revistas.ucm.es/lib/pkp/templates/im...">https://revistas.ucm.es/lib/pkp/templates/im...</a>	font-down	911b	Download
<input type="checkbox"/>	<a href="https://revistas.ucm.es/lib/pkp/templates/im...">https://revistas.ucm.es/lib/pkp/templates/im...</a>	font-up	926b	Download
<input type="checkbox"/>	<a href="https://revistas.ucm.es/plugins/generic/web...">https://revistas.ucm.es/plugins/generic/web...</a>	rss10_logo	988b	Download
<input type="checkbox"/>	<a href="https://revistas.ucm.es/public/journals/75/n...">https://revistas.ucm.es/public/journals/75/n...</a>	homepagelimage_es_ES	294.3Kb	Download

Below the table, there are options to show previews and a 'DESCARGAR (0)' button.

# Normalización de Texto

Es muy común que el título de un artículo, su resumen y otros metadatos sean copiados directamente desde el archivo PDF. Esta acción suele presentar problemas de diferente tipo que alteran el texto a pegar, por ejemplo: codificación inválida, exceso de saltos de línea y tabulaciones, textos en Mayúsculas, caracteres especiales que no se veían en el PDF, entre otros.

⇒ Normalizador de Texto. <http://sedici.unlp.edu.ar/norm.php>

Es una aplicación PHP que a partir de una mínima configuración revisa y corrige el texto ingresado. Por ejemplo:

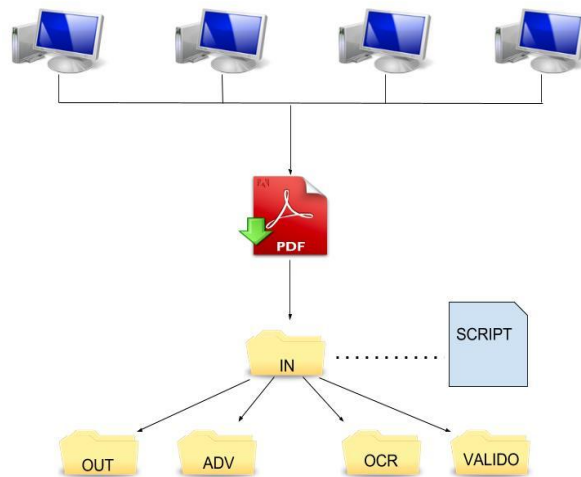
- multiplicidad indeseada de espacios y saltos de línea
- detección de saltos de línea indebidos
- Normalización de caracteres (UNICODE NORM)



## 3-height

Para trabajar con Documentos PDF se generó un script que utiliza un programa estilo **servidor/cliente**, encargado de:

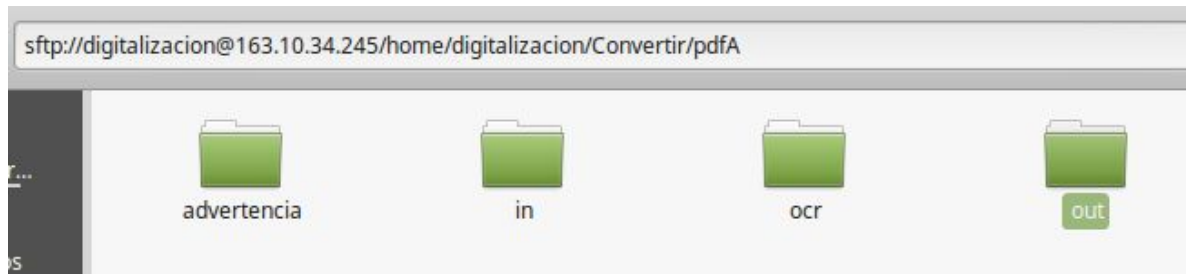
- Detección de archivos
- Análisis
- Conversión
- Verificación



## 3-height

Simplemente tenemos una carpeta compartida con el nombre PDFA que consta de 4 directorios donde los administradores podrán transformar los archivos PDF en PDFA. Los directorios son:

- Una carpeta “in” para ingresar los archivos a procesar
- Una Carpeta “out” donde se depositarán los archivos resultantes.
- Y dos carpetas destinadas a diferentes tipos de errores llamadas “advertencia” y “ocr”



# Generación de PDF/A - 3-Height

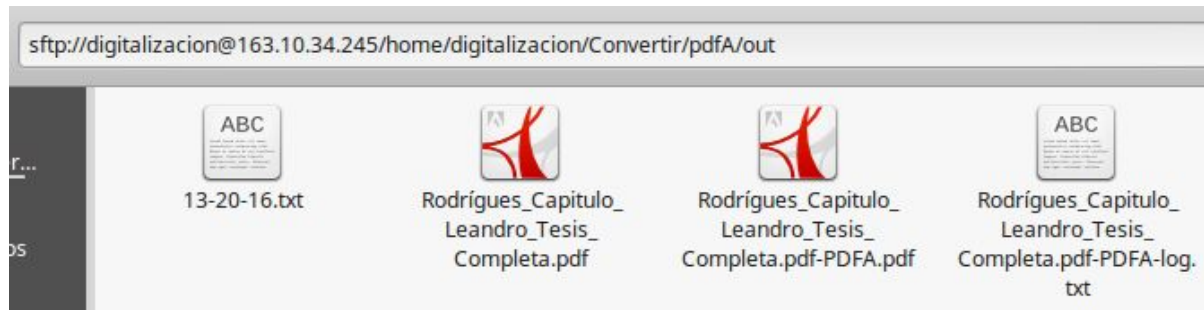
El 3 height analiza el pdf y elige en qué versión va a convertirlo. Si la conversión sale bien, en la carpeta out tendremos los siguientes archivos:

El archivo con la fecha 13-20-16.txt presenta el log de la ejecución del script..

El archivo pdf original.

El archivo convertido con la terminación: -PDF/A.pdf

El último archivo txt da más detalles de la conversión del archivo original

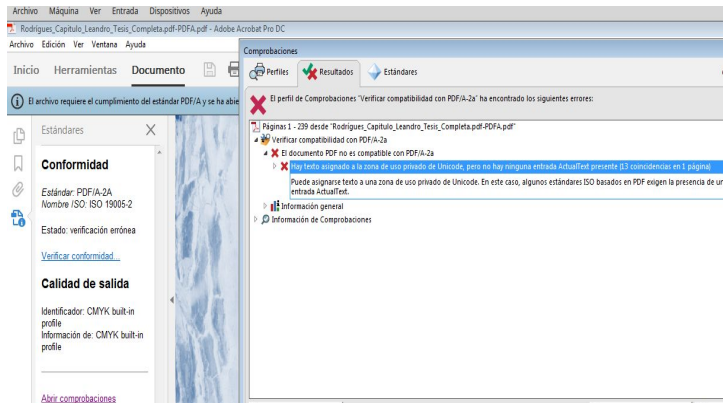


# Validación de PDF/A - Acrobat DC

Una vez obtenido el pdf/A de 3-Height es necesario validarlo también en Acrobat DC. Si la verificación es errónea dependiendo el caso de error podemos arreglarlo desde el mismo Acrobat. Por ejemplo: cuando un archivo no pasa la verificación porque el texto no es unicode en todo el pdf. Generalmente este problema se soluciona transformando el archivo en la versión de pdfA llamada pdf/A-2u.

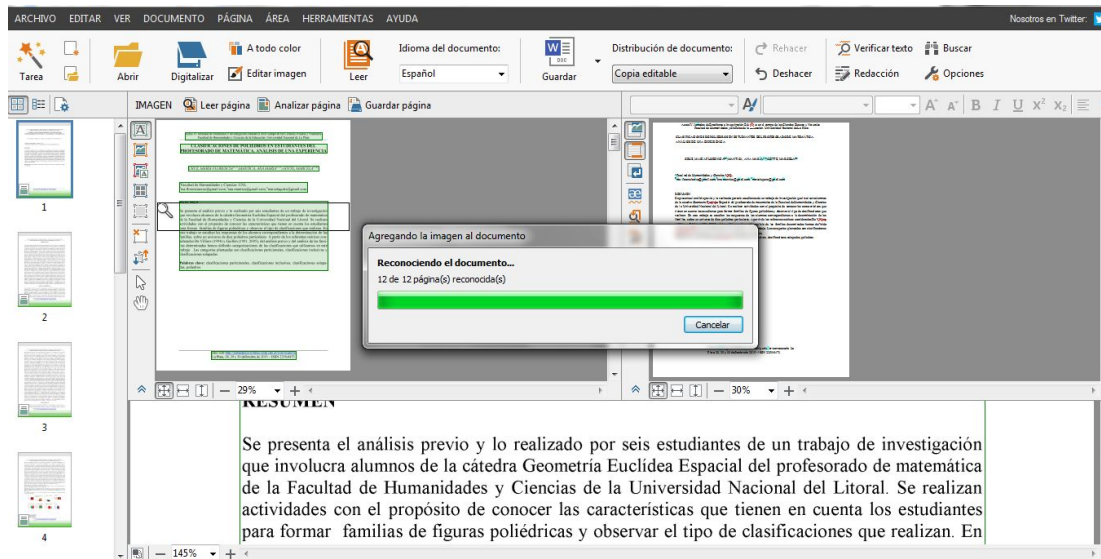


Adobe Acrobat DC



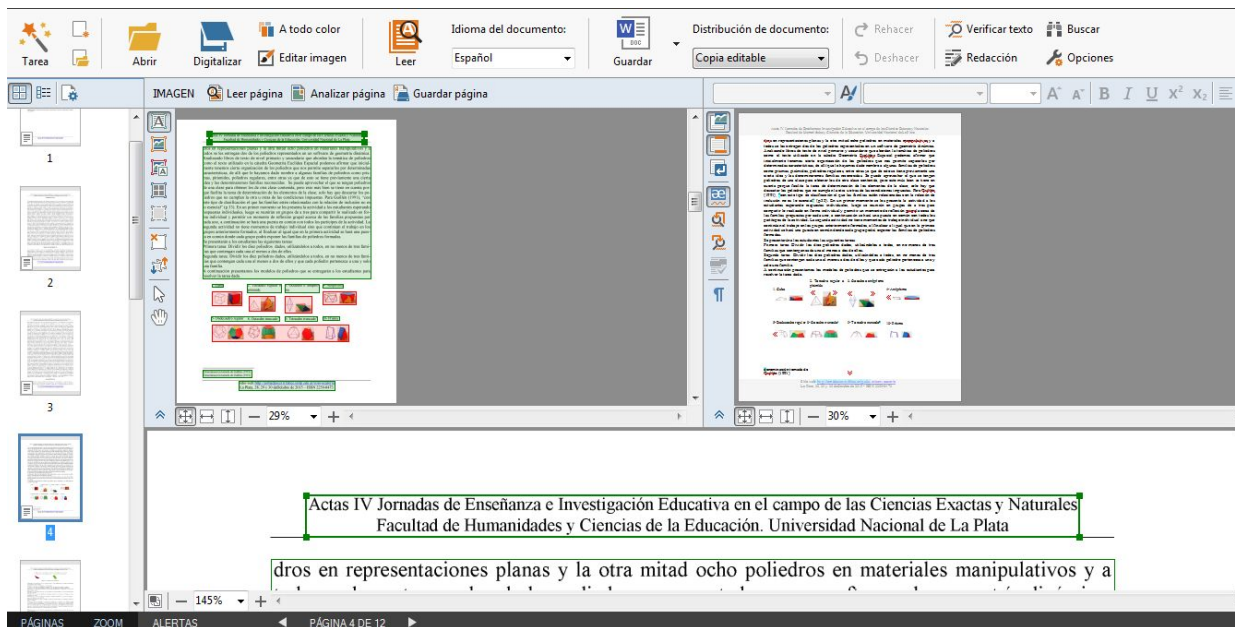
# Generación de OCR con ABBYY Fine Reader

ABBYY Fine Reader es un software de OCR que permite trabajar y editar pdfs de manera rápida y confiable.



# ABBYY Fine Reader

El software permite seleccionar imágenes e indicar qué parte de la página debe ser reconocido y qué no. Es de los softwares con motor de reconocimiento óptico de caracteres más precisos del mercado.



# Digitalización

*Digitalización y preservación*

*Digitalización de documentos*

*Escáner DAL*

*Proceso de imágenes*

*Digitalización para dispositivos móviles*

*Formatos*

*PDF/A*



# Introducción a la Digitalización

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

Una estrategia de preservación es la de adoptar estándares internacionales, es decir, apoyarse en la afirmación de que los estándares internacionales son relativamente estables en el tiempo.

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.



# Generación de PDF/A

Un formato de preservación para documentos de texto es el estándar PDF/A, descrito en las normas ISO 19005-(1-2-3).

Este formato está basado en el estándar PDF 1.4, al que le incorpora algunos requerimientos adicionales, por ejemplo:

- Especificaciones sobre los metadatos y la estructura del archivo.
- La paleta de colores (incluyendo escala de grises y blanco/negro) no deben ser representados en un espacio de color de dispositivo (DeviceRGB, DeviceCMYK, DeviceGray).
- Las fuentes usadas en texto visibles deben estar embebidas (incluidas dentro del archivo).

Uno de los propósitos de los requerimientos del estándar PDF/A es de proveer soporte para personas con capacidades diferentes, por ejemplo, incorporando la información requerida y necesaria para aplicaciones que hagan el pasaje de texto a voz.



# Digitalizaciones en SEDICI

- 1) Documentos de papel físicos (libros, revistas, tesis, informes, etc.)
- 2) Documentos .doc, .docx, .odt, etc.
- 3) Documentos PDF



# 1) Digitalización de documentos

- Escáner HP scanjet 7500 / Fujitsu fi-7160

Softwares de escaneos: HP Smart Document Scan Software 3 - PaperStream

- Escáner DAL

Captura de imágenes

Rotación de la imagen

Renombrado

Combinación

Proceso de imágenes mediante ScanTailor/Photoshop

Impresión en PDF, OCR en Abbyy FineReader 12



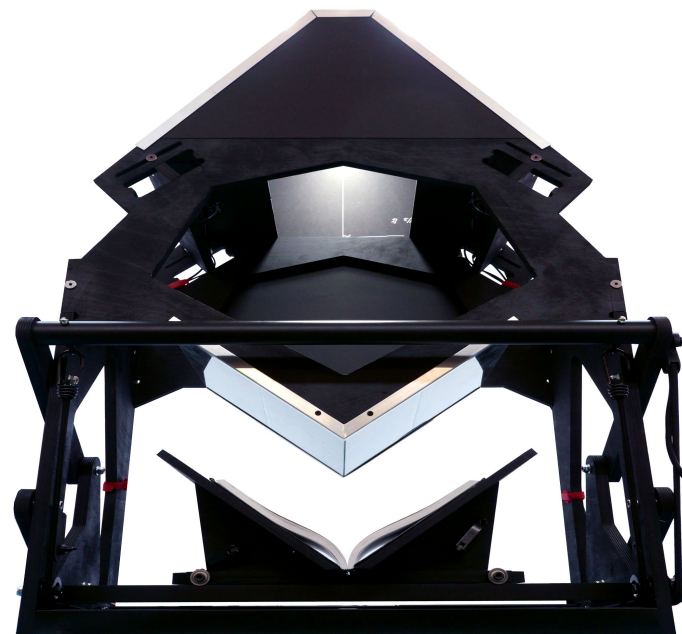
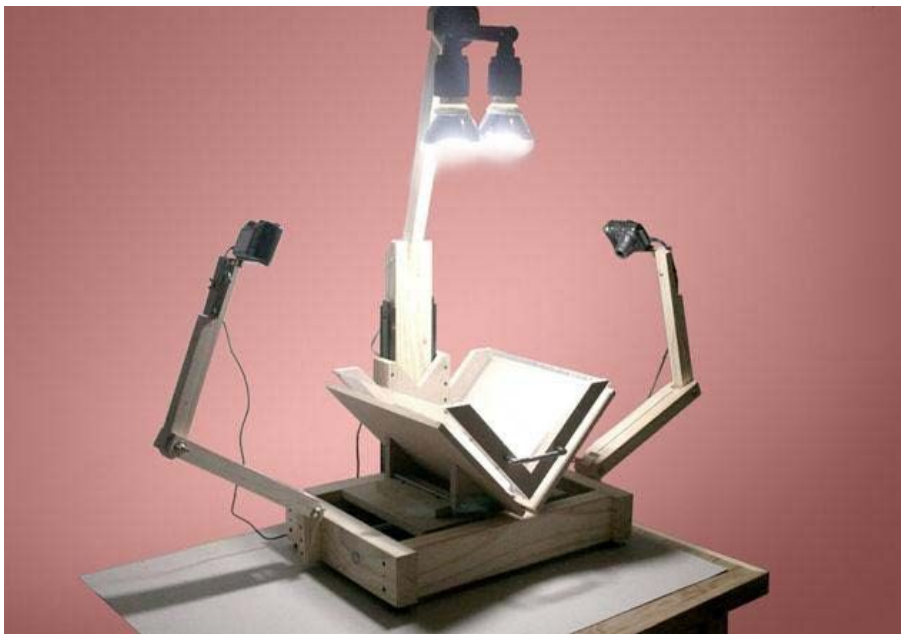
# DAL

Derecho a Leer (DaL) (Facultad de Filosofía y Letras - UBA). Es un espacio de difusión de las políticas a favor de la lectura. Inspirados en la experiencia de *Do It Yourself* desarrollaron un escáner para libros de forma artesanal y muy económica, utilizando madera y cámaras digitales estándar. El objetivo es tomar una foto con la mayor precisión posible de cada hoja del libro.

Ejemplo de digitalización realizada en SEDICI:  
Academia Nacional de Agronomía y Veterinaria (ANAV).  
Tomo I 1932- 1934



# Escáneres DAL



# Captura de imágenes

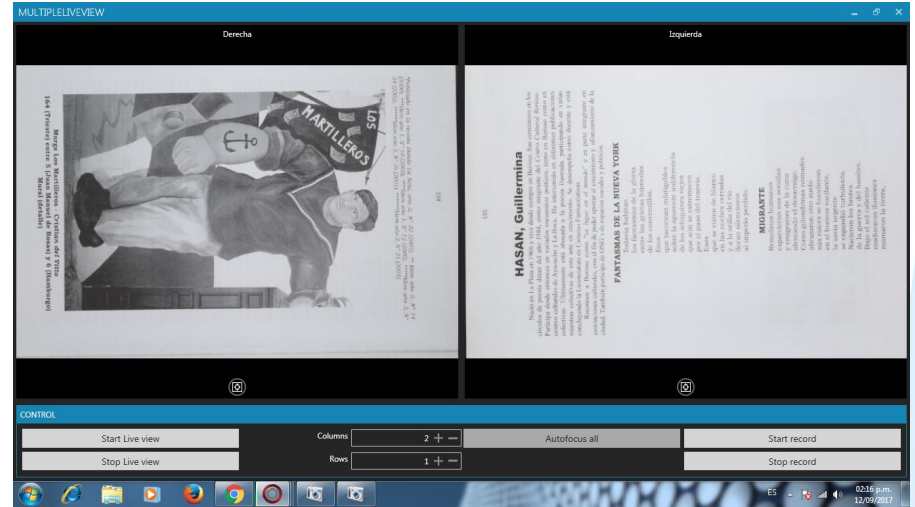
Maquinas Nikon D5300

Software *digiCamControl*

Mecanismo de disparador manual

Configuración a través de software de las cámaras

Previsualización en tiempo real



# Pre-procesamiento de imágenes

Luego de la descarga de las fotos al PC, se corre un script que rota, renombra y combina las imágenes sacadas de las cámaras.

```
prebi@Digitalización ~ $ ./editar  
MiNuevoProyecto
```

Procesamiento:

Se crea un proyecto en Scan Tailor donde a cada imagen se le hace un proceso de:

- alineación
- selección de contenido
- centralización e igualar tamaño
- blanquear fondo y borrar manchas.

ScanTailor crea imágenes en formato .TIFF Luego de generados los TIFF se crea un nuevo proyecto de AbbyyFineReader, donde se selecciona el texto e imágenes para crear un PDF/A-1a.

El PDF es verificado y optimizado para la vista rápida en web con Acrobat DC.



# HP 7500/ Fujitsu fi-7160

Se necesitan las hojas del documento sueltas, sin bordes salientes

Se pueden escanear a una cara o dos

Escaneos en blanco y negro: 600ppp

Escaneos en color/ escala de grises:

400ppp

TIFF con compresión sin pérdida (LZW).

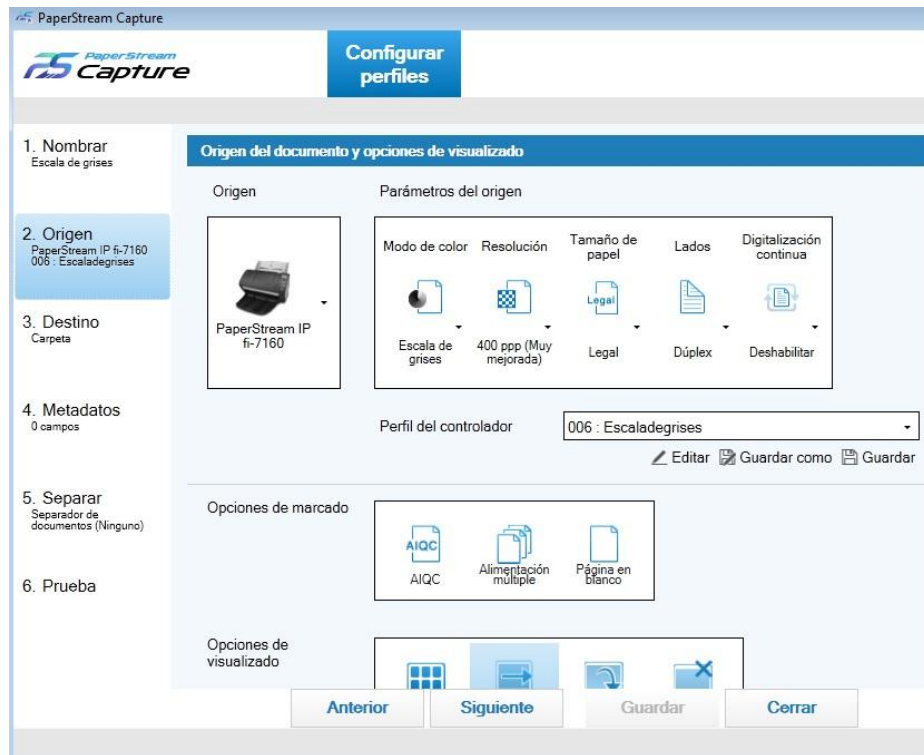
Máximo de hojas por paquete: 100

Software de escaneo: HP Smart

Document Scan Software 3 /

PaperStream

Formato de salida del escaneo: .TIFF



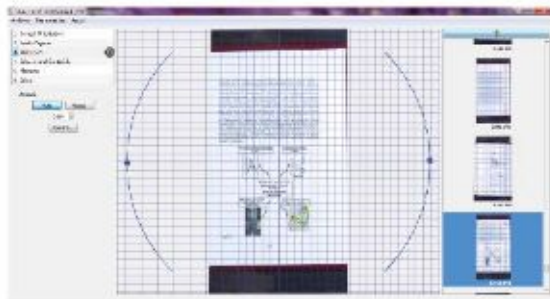
# Clasificación del material según relevancia

Dependiendo de la utilidad y el interés del material, los procesos de edición de imagen tienen mayor o menor automatización y revisión posterior. El material de alta relevancia (copias únicas por ejemplo) requieren un proceso de revisión de la edición de imagen y del OCR página por página. Materiales de relevancia media requieren una revisión detallada de la tapa y los índices y una revisión general del resto. En cambio, los materiales de digitalización rápida requieren una revisión general y un proceso casi totalmente automatizado.

# Programas utilizados

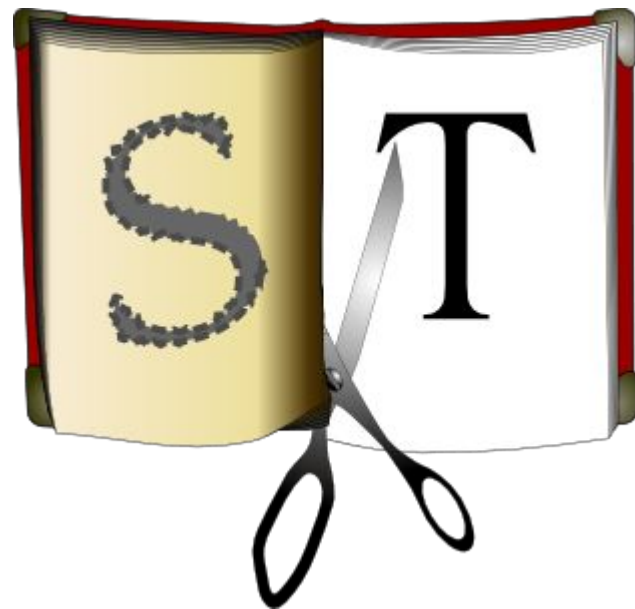


Scantailor



## Edición de imagen


Luego de obtener las imágenes escaneadas, se editan en Photoshop o ScanTailor. De esta manera se puede corregir la orientación, dividir y alinear las páginas, seleccionar el contenido, limpiar los márgenes, eliminar las manchas y modificar el contraste.



Archivo Herramientas Ayuda

- 1 Corregir orientación
- 2 Dividir páginas
- 3 Alineación
- 4 Seleccione contenido
- 5 Márgenes


**Rotar**



Restablecer

**Extensión**


Aplicar a...



Archivo Herramientas Ayuda

- 1 Corregir orientación
- 2 Dividir páginas
- 3 Alineación
- 4 Seleccione contenido
- 5 Márgenes


**Rotar**



Restablecer

**Extensión**

Aplicar a...




Archivo Herramientas Ayuda

- 1 Corregir orientación
- 2 **Dividir páginas**
- 3 Alineación
- 4 Seleccione contenido
- 5 Márgenes
- 6 Salida

### Diseño de página



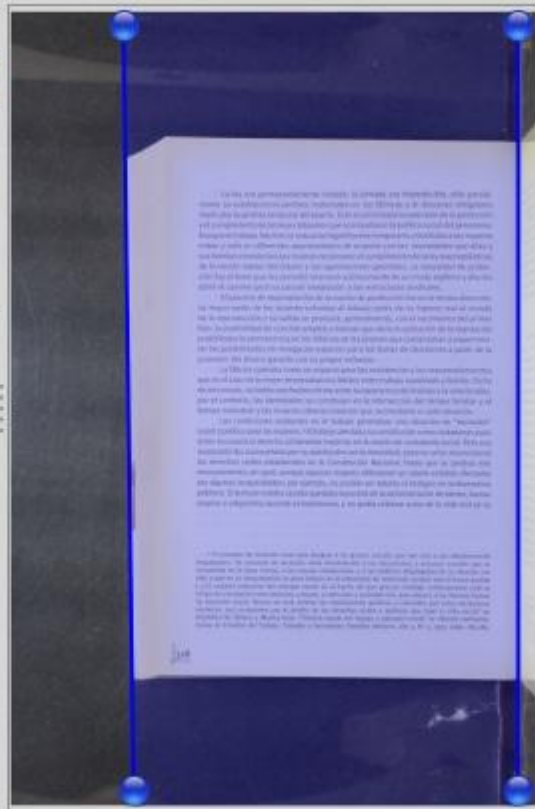
Configurar manualmente

Cambiar...

### Línea divisoria

Auto

Manual



316.jpg

317.jpg

Orden natural



Archivo Herramientas Ayuda

1 Corregir orientación

2 Dividir páginas

3 Alineación

4 Seleccione contenido

5 Márgenes

6 Salida

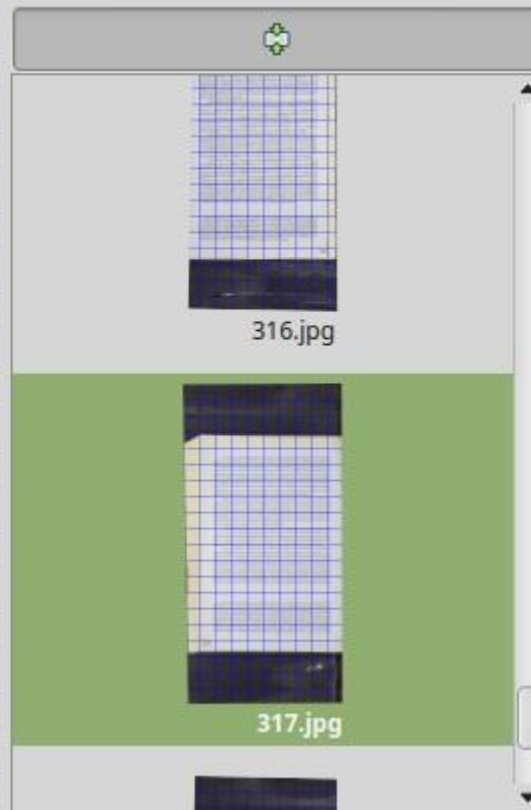
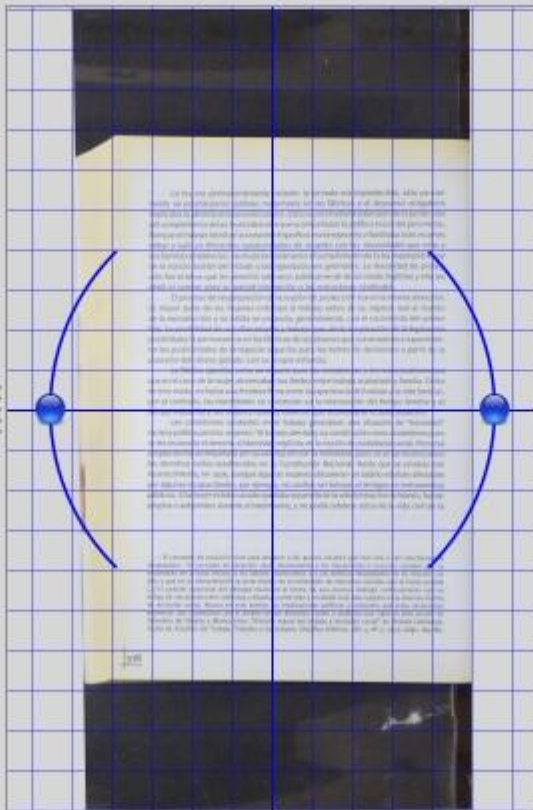
## Alineación

Auto

Manual

0,19°

Aplicar a...



Archivo Herramientas Ayuda


- 1 Corregir orientación
- 2 Dividir páginas
- 3 Alineación
- 4 **Seleccione contenido**
- 5 Márgenes
- 6 Salida

**Caja de contenido**

Auto Manual

**Extensión**

Aplicar a...



315.jpg

316.jpg

317.jpg

Orden natural

The image shows a software interface for PDF editing. On the left, a sidebar contains a list of tools, with 'Seleccione contenido' (Select content) highlighted. Below this are buttons for 'Auto' and 'Manual' modes, and an 'Aplicar a...' button. The main area is split into two panes. The left pane shows a document page with a blue selection box around a paragraph of text. The right pane shows a vertical list of three image thumbnails labeled '315.jpg', '316.jpg', and '317.jpg'. At the bottom of the right pane, there is a dropdown menu set to 'Orden natural'.



Archivo Herramientas Ayuda

3 Alineación  
4 Seleccione contenido  
5 Márgenes

**Márgenes**

Milímetros (mm)

Arriba 10,0  
Abajo 5,0  
Izquierda 10,0  
Derecha 10,0

Aplicar a...

**Alineación**

Igualar en tamaño con las demás páginas

Aplicar a...

315.jpg  
316.jpg  
317.jpg

Orden natural



Archivo Herramientas Ayuda

4 Seleccione contenido  
5 Márgenes  
6 Salida

**Resolución de salida (DPI)**  
400  
Cambiar...

**Modo**  
Combinado

0  
Más fino Más grueso

Aplicar a...

**Antideformación**  
Apagado  
Cambiar...

**Eliminar manchas**  
Aplicar a...

La ley era permanentemente violada: la jornada era impredecible, sólo parcialmente se establecieron jardines maternales en las fábricas y el descanso obligatorio implicaba la pérdida temporal del salario. Esto ocurrió hasta la extensión de la protección y el cumplimiento de las leyes laborales que acompañaron la política social del peronismo. Aunque el trabajo fabril en la industria frigorífica era temporario y facilitaba a las mujeres entrar y salir en diferentes oportunidades de acuerdo con las *necesidades* que ellas y sus familias establecían, las mujeres reclamaron el cumplimiento de la ley reapropiándose de la noción tutelar del Estado y las organizaciones gremiales. La necesidad de protección fue el tema que les permitió intervenir públicamente de un modo legítimo y ello les abrió el camino para su parcial integración a las estructuras sindicales.

El proceso de reapropiación de la noción de *protección* fue en la misma dirección. La mayor parte de las mujeres entraban al trabajo antes de su ingreso real al mundo de la reproducción y su salida se producía, generalmente, con el nacimiento del primer hijo. La posibilidad de conciliar empleo y trabajo que abrió la aplicación de la legislación posibilitaba la permanencia en las fábricas de las jóvenes que comenzaban a experimentar las posibilidades de renegociar espacios para las tomas de decisiones a partir de la posesión del dinero ganado con su propio esfuerzo.

La fábrica operaba como un espacio para las resistencias y los reacomodamientos que en el caso de la mujer atavesaban los límites entre trabajo asalariado y familia. Dicho de otro modo, no había una frontera firme entre la experiencia del trabajo y la vida familiar, por el contrario, las identidades se construían en la intersección del tiempo familiar y el tiempo industrial y las mujeres obreras tuvieron que acomodarse a cada situación.

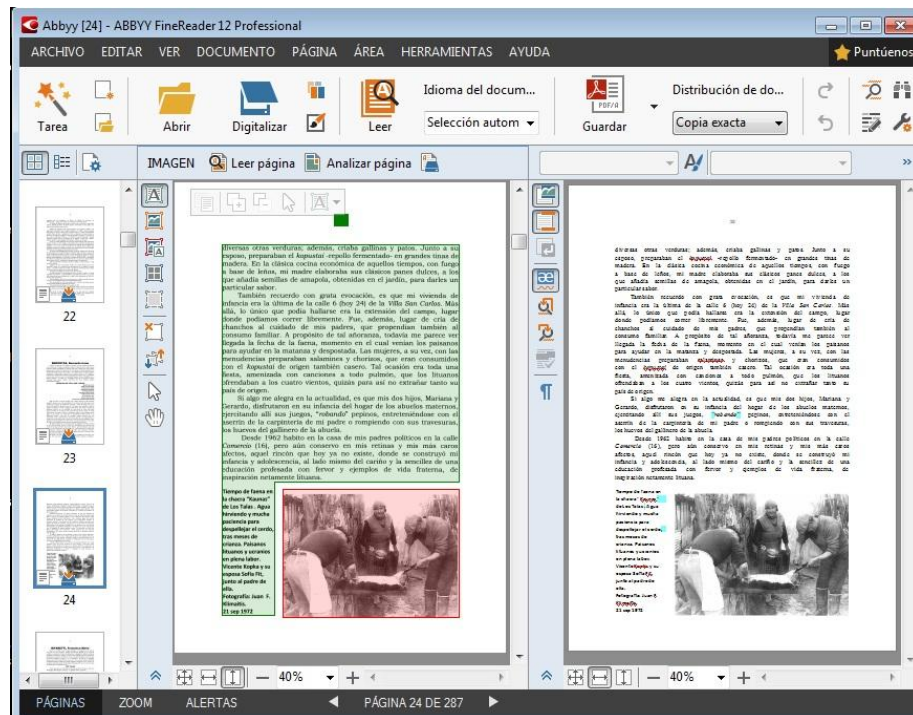
Las condiciones existentes en el trabajo generaban una situación de "exclusión" social y política para las mujeres.<sup>4</sup> El trabajo afectaba su constitución como ciudadanas pues se les reconocía el derecho al bienestar implícito en la noción de ciudadanía social. Pero esa aceptación iba acompañada por su asimilación con la minoridad, pues no se les reconocieron los derechos civiles establecidos en la Constitución Nacional. Hasta que se produjo ese reconocimiento, en 1926, aunque algunas mujeres obtuvieran un salario estaban afectadas por algunas incapacidades, por ejemplo, no podían ser tutoras ni testigos en instrumentos públicos. Si la mujer estaba casada quedaba separada de la administración de bienes, fueran propios o adquiridos durante el matrimonio, y no podía celebrar actos de la vida civil sin la

<sup>4</sup> El concepto de exclusión sirve para designar a los grupos sociales que han sido o son selectivamente desplazados. "El concepto de exclusión alude directamente a los mecanismos o procesos sociales que se encuentran en la base misma, a los actores involucrados y a las políticas desplegadas en su relación con ella; y que en su interpretación se pone énfasis en el entramado de relaciones sociales que la hacen posible [...] El carácter relacional del enfoque reside en el hecho de que procura inteligir continuamente cuál es el tipo de vinculación entre individuo y estado, y entre éste y sociedad civil, que subyace a las diversas formas de exclusión social. Abarca en este sentido las implicaciones políticas y culturales que estas exclusiones encierran, sus conexiones con el ámbito de los derechos civiles y políticos que rigen la vida social" en Oriandina de Oliveira y Marina Ariza: "División sexual del trabajo y exclusión social" en *Revista Latinoamericana de Estudios del Trabajo, Trabajo e Sociedades: Desafíos teóricos*, año 3, N° 5, 1997, págs. 184-185.

318

# Reconocimiento óptico de caracteres

Luego de editar las imágenes se realiza el OCR con el Abbyy FineReader. En esta etapa del proceso se selecciona el contenido según sea texto, imagen o cuadro. Luego se revisa la corrección del OCR y se generan los archivos PDF/A.



# Compresión de PDF

The image shows two overlapping dialog boxes from Adobe Acrobat. The background dialog is 'Opciones' (Options) with the 'Avanzadas' (Advanced) tab selected. The foreground dialog is 'Configuración personalizada' (Customized Configuration) for PDF/A export.

**Opciones (Background Dialog):**

- Documentos: Digitalizar/Abrir, Leer, Guardar, Ver, Avanzadas
- Formatos: DOCX/ODT/RTF, XLSX, PDF, PDF/A, HTML, PPTX, TXT, CSV, EPUB/FB2, DjVu
- Tamaño de papel predeterminado: Automático
- Modo de guardado: Texto debajo de imagen de página
- Conservar texto y colores de fondo (checked)
- Conservar encabezados y pies de página (checked)
- Comprimir las imágenes utilizando MRC (checked)
- Aplicar ABBYY Precise Scan para suavizar los caracteres de las imágenes (checked)
- Configuración de imágenes: Personalizado... (checked) Conservar imágenes

**Configuración personalizada (Foreground Dialog):**

- Resolución:  Reducir la resolución de imagen en función de la salida de destino
  - Impresión, pantallas de alta calidad (300 ppp)
  - Lectura en pantalla, web (150 ppp)
  - Compartir (96 ppp)
- Control del color: Automático
- Calidad: Pérdida de calidad permitida
- Imágenes en escala de grises y en color: JPEG, ZIP, LZW, J2K
- Imágenes en blanco y negro: CCITT, JBIG2
- Calidad (45%): [Slider]
- Buttons: Restablecer, Aceptar, Cancelar



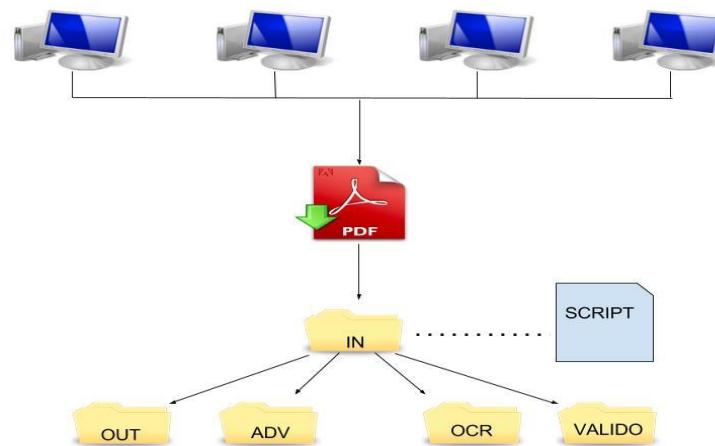
# Conversión a PDF/A

2) Conversión de formatos de texto (.doc, odt, etc).

Se guarda el documento como “PDF/A” desde el mismo word/libreOffice, etc.

3) Documentos PDF. Se utiliza un programa estilo servidor/cliente, encargado de:

- Detección de archivos
- Análisis
- Conversión
- Verificación

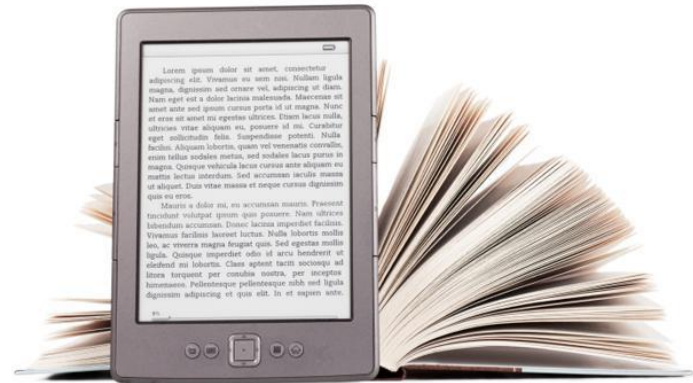


# Digitalización para dispositivos móviles

## MOBI

Es un formato para archivos de libro electrónico (e-book) creado por Mobipocket S. A. En el formato de libro digital MOBI se marca el contenido, pero no se delimita su formato, que se adapta a los diferentes tamaños de las pantallas de los múltiples lectores de libros electrónicos del mercado.

El formato MOBI está basado en las especificaciones Open eBook.

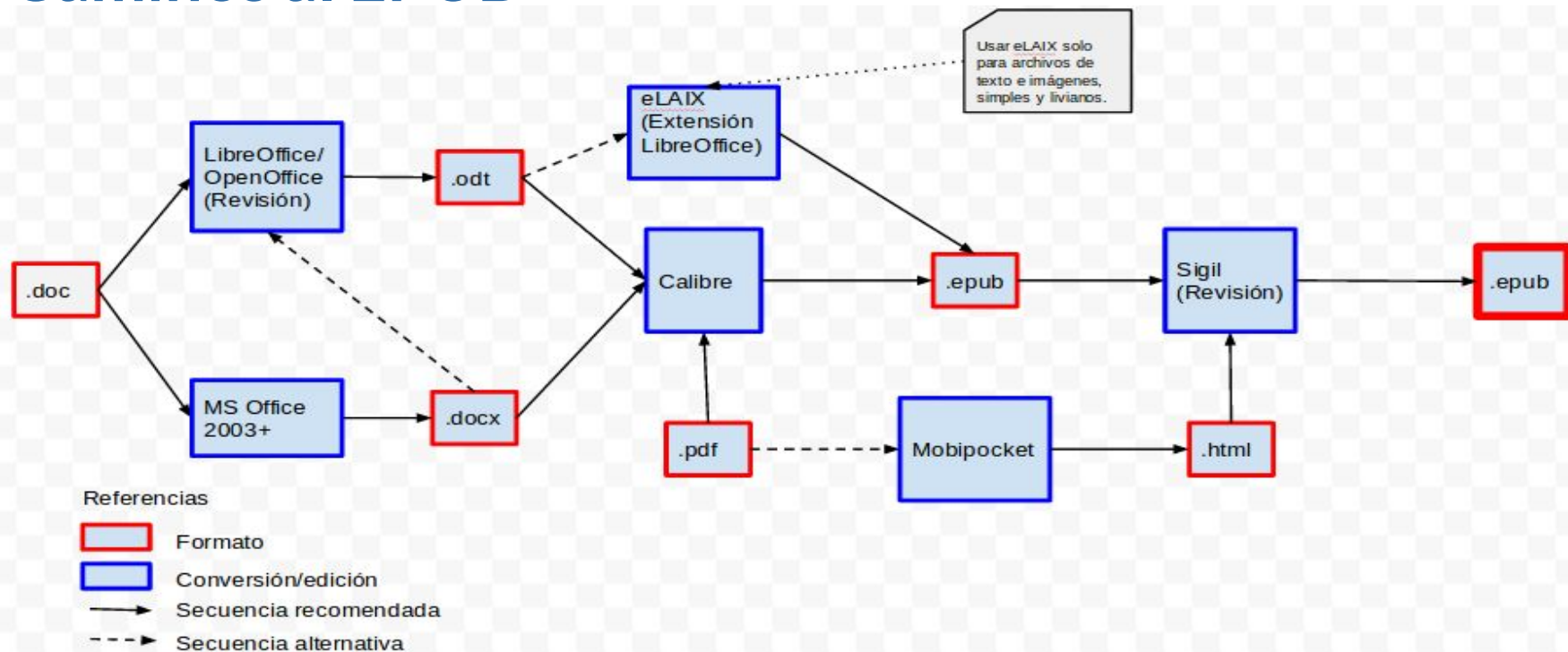


# Formatos EPUB

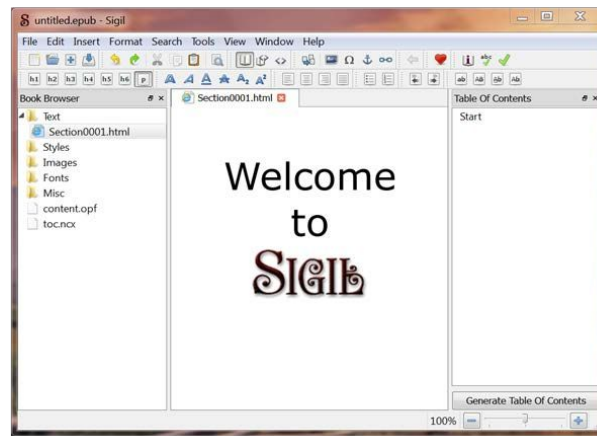
**EPUB:** El formato se diseñó para adaptarse a distintos tamaños de letra y pantalla. También se puede cambiar el tipo de letra. En el formato de texto para libro digital ePub, se marca el contenido, pero no se delimita su formato, que se adapta a las diferentes pantallas y tamaños de pantalla de los múltiples dispositivos. El formato fue creado por International Digital Publishing Forum (IDPF), como un formato específico para visualizar libros.

**EPUB3:** El formato, además de las funcionalidades habituales del EPUB ofrece salida de audio en la funcionalidad de texto-a-audio, características de voz, normas de pronunciación, especificación de fallbacks (conversión de un vídeo en texto, para discapacitados).

# Caminos al EPUB



# Programas utilizados en el proceso



# Audiolibros

Se trata de grabaciones habladas (o leídas automáticamente por programas informáticos) de libros.

Software usado para convertir texto a voz o archivos de audio: Natural reader

Formato de salida: .MP3 o .WAV



NaturalReader is an easy-to-use software that converts written text into natural sounding spoken words giving educators, professionals and students a powerful aid and learning tool. NaturalReader can assist ESL students (English as a Second Language) and people with learning difficulties such as dyslexia.



¡Muchas gracias!

Por consultas: [marisa.degiusti@sedici.unlp.edu.ar](mailto:marisa.degiusti@sedici.unlp.edu.ar)

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

<http://digital.cic.gba.gob.ar/>

<http://cesgi.cic.gba.gob.ar/>

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

<http://www.istec.org/liblink/>

<http://revistas.unlp.edu.ar/cientificas/>

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

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

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

Este material está disponible en la colección de **SEDICI** <http://sedici.unlp.edu.ar/handle/10915/25295>

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

