



FACULTAD DE INFORMATICA



UNIVERSIDAD
NACIONAL
DE LA PLATA

Bibliotecas y Repositorios Digitales. Tecnología y aplicaciones 2023

Prof. a cargo: Dra. Marisa R. De Giusti, Dr. Gonzalo L. Villarreal



Repositorio Institucional
Comisión de Investigaciones Científicas



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

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



Clase 2

Aspectos tecnológicos

Índice - Aspectos tecnológicos

- Aspectos generales sobre software de repositorios
 - Gestión y requerimientos
- Características y personalización del software
 - Arquitectura, librerías, lenguajes
- Modelo de contenidos
 - ¿Cómo se representan internamente los metadatos y los contenidos?
- Vías de ingesta y circuitos de carga
 - ¿Cómo ingresa el contenido al repositorio?
- Localización de contenidos
 - Id persistentes y exploración.
- Interoperabilidad
- Preservación digital
- Estadísticas

Software del repositorio - Gestión

¿Servicio tercerizado o gestión propia?

Aspectos a tener en cuenta

- Personal disponible
 - roles y formación
- Infraestructura
- Financiamiento
- Soberanía
 - ¿Cuánto control tenemos sobre los datos?
 - ¿Podemos implementar servicios alrededor del repositorio? ¿Qué tan fácil es?
 - ¿Quién es el responsable?
 - Licencias
 - Personalización

Gestión - Servicio tercerizado

Ejemplos

Duraspace:

- DspaceDirect → Repositorio listo para usarse en la nube
- DuraCloud → Almacenamiento de contenido en la nube
- Archives Direct → Creador de paquetes de contenido de preservación digital



Digital Commons:

- Plataforma comercial que ofrece un repositorio institucional totalmente alojado, listo para usarse.
- Pertenece a la editorial Elsevier



Gestión - Propia

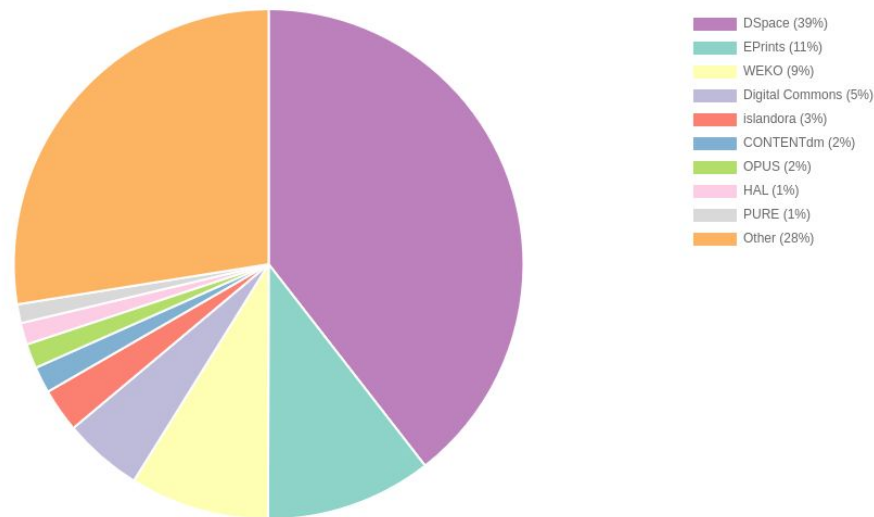
Tener en cuenta

- ¿Quién es responsable del desarrollo?
 - Desarrollo propio
 - Empresa ⇒ Ej. DuraSpace
 - Universidad ⇒ Ej. Universität Stuttgart
 - ONG ⇒ Ej. Public Knowledge Project (PKP)
- Software a utilizar - Dspace, eprints
- Comunidad
- Licencias - Abierta o cerrada
- Infraestructura
- Personal (roles y capacitación)



Gestión - Software de repositorios de Acceso Abierto

DSpace	2348
EPrints	662
WEKO	542
Bepress	301
Otros	2169 (27%)



Total = 6025 repositorios
[OpenDOAR](#) - Abril 2023

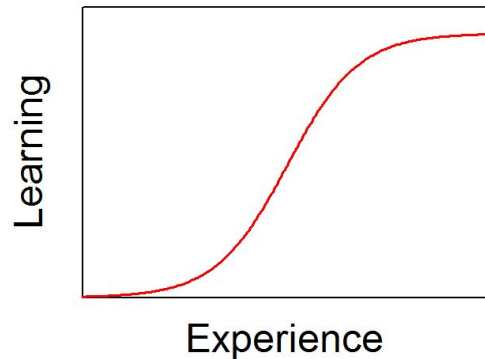
Gestión - Curva de aprendizaje

- para usuarios y administradores
 - Uso del software
 - Ingesta y exploración del contenido
- para desarrolladores y sysadmins
 - Arquitectura del software
 - Experiencia de los desarrolladores

¿Cuánta documentación existe?

Tutoriales, cursos, webinars, talleres, congresos

S-Curve (Sigmoid)

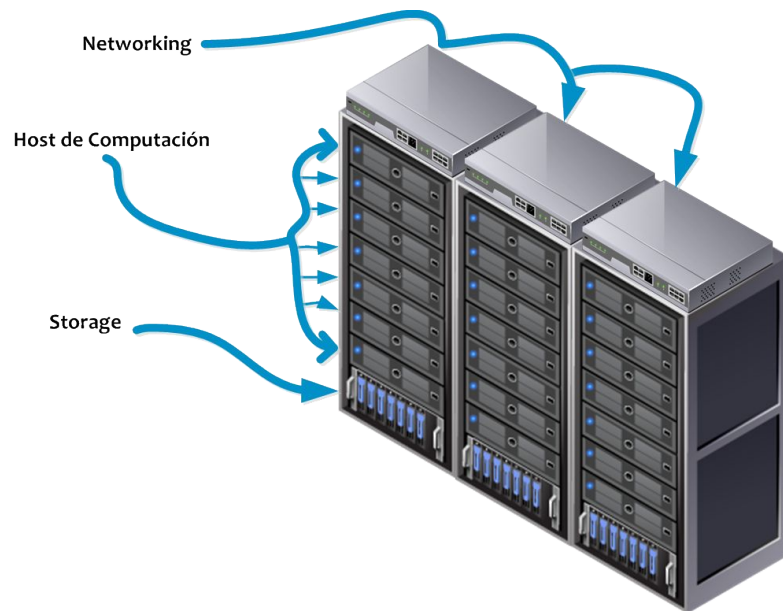


© Alan Fletcher 2013 This file is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported

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
 - backups
- **Procesamiento** ⇒ memoria y procesador
 - base de datos
 - aplicación



Requerimientos físicos

- 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



Contenido del repositorio

¿Que se almacena en un repositorio?

Recurso + Esquema de metadatos que lo describe

- Recurso
 - Son los archivos, los **objetos digitales** que representan las obras que se publican en el repositorio (PDFs de artículos, audios, imágenes, documentos de texto, etc).
- Metadatos
 - Conjunto de datos que describen a un recurso en particular (título de la obra, autores, resumen, etc).

Contenido del repositorio - Representación

¿Cómo afecta la forma en que representamos los recursos de un repositorio?

- catalogación,
- exposición web
- interoperabilidad

- Formatos permitidos
 - Ej: PDF, Word, mp4, jpg.

- Esquema de metadatos
 - Fijo ⇒ Ejemplo en sistemas como OJS, OCS, OMP
 - Esquema configurable, ampliable

- Uso de Vocabularios controlados

Almacenamiento de datos y metadatos

¿Qué tecnología de base de datos utiliza?

- BD relacional, XML, NO SQL, RDF

Considerar el volumen de 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 (OD)

- ¿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?



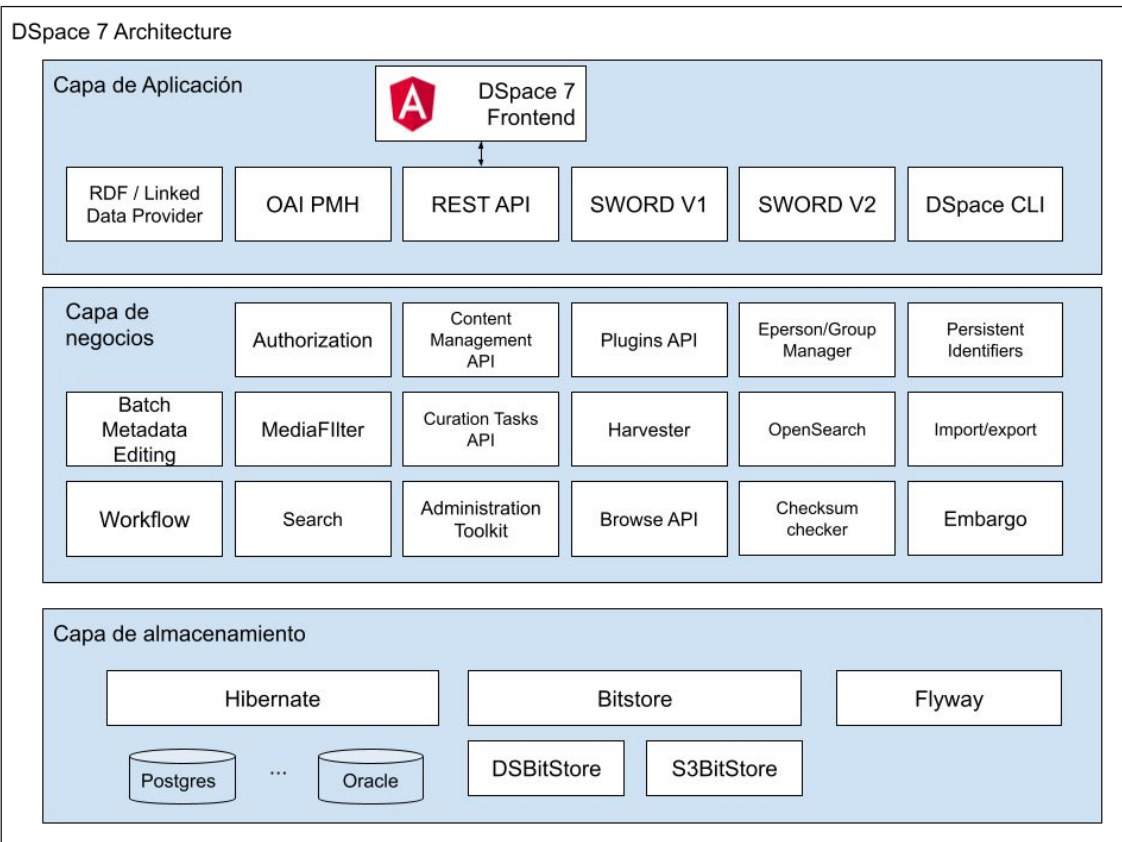
Características y Personalización del software

Arquitectura, librerías, lenguajes

Características generales de DSpace

- Software libre y open-source
- Proyecto escrito mayoritariamente en lenguaje **JAVA**.
- Desarrollado y mantenido por una amplia **comunidad**
- 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 - Arquitectura v7



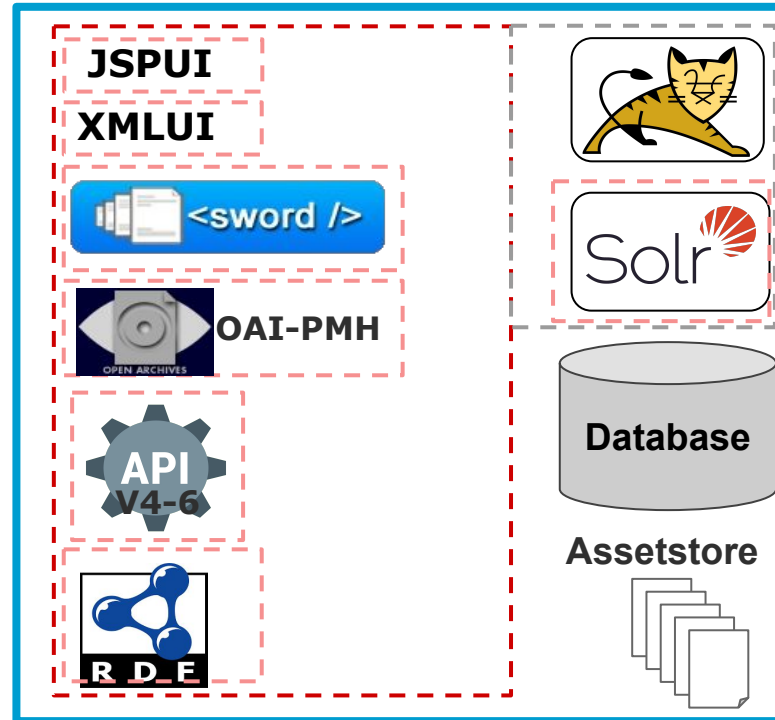
Organizado en 3 capas

- [Aplicación](#)
- [Lógica de negocios](#)
- [Almacenamiento](#)

Características generales de DSpace - Arquitectura v1-6

7 webapps separadas:
2 UIs, 4 servicios web &
Solr

Front & Back End



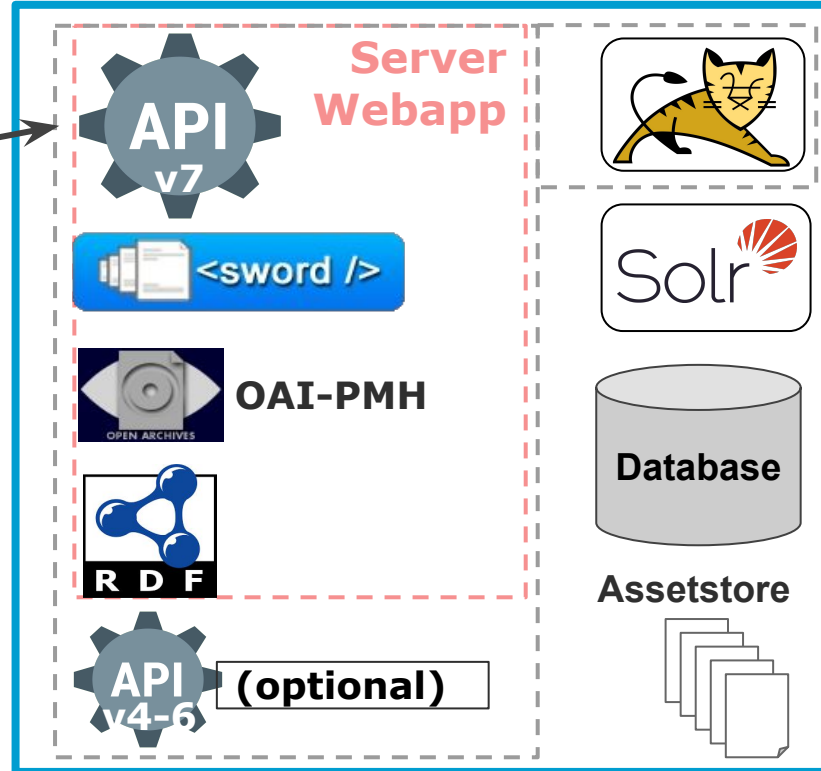
Características generales de DSpace - Arquitectura v7

Back End

Front End



1 webapp backend
(Solr ahora se instala por separado)



Características generales de DSpace - Stack

Proyecto Maven escrito mayoritariamente JAVA. Compuesto por varios subproyectos. Se requiere:

- Java + Spring
 - jre7 o jre8
- Web container o web server
 - Jetty, Tomcat, entre otros
- Índice de búsqueda
 - Solr, Elasticsearch
- Base de datos relacional para persistir el modelo.
 - PostgreSQL, Oracle, etc
- Angular
 - para la interfaz de usuario, solo DSpace 7



Interfaz web

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

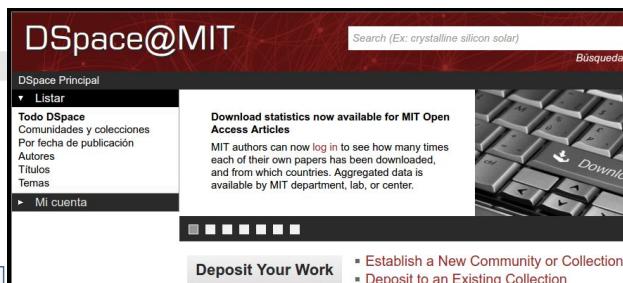
- Características deseables
 - simple para el usuario final
 - intuitiva (nadie debería necesitar un tutorial)
 - Internacionalizable - aunque sea en inglés
 - accesibilidad - usable por todas las personas
 - *responsiveness* - usable desde distintos dispositivos

Interfaz web

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



CIC-Digital es el Repositorio Institucional de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, creado con el fin de preservar y dejar accesible en abierto toda la producción científico-tecnológica generada en el ámbito de las instituciones CIC.



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



Repositorio Universidad de La Sabana

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

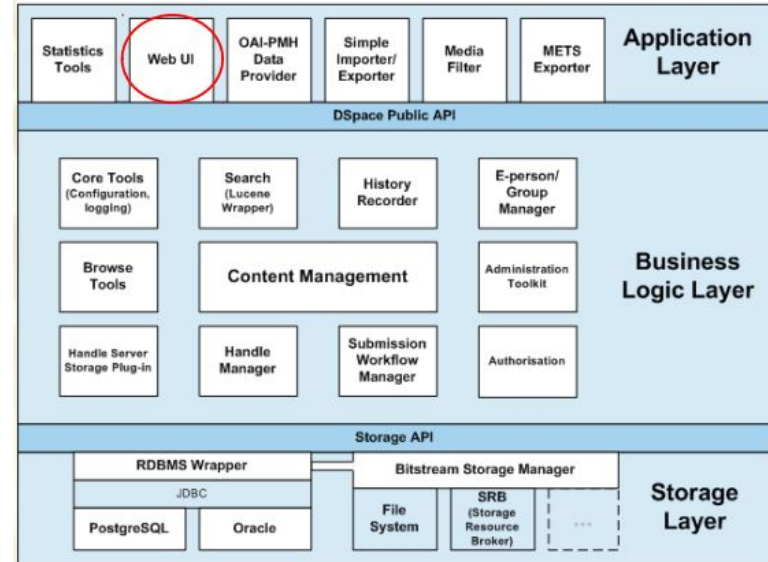


Emergencia hídrica

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

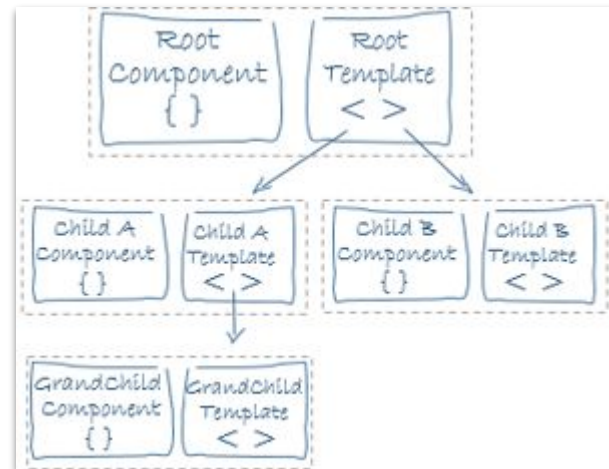


Arquitectura, capas y componentes de DSpace. Fuente <https://wiki.duraspace.org/>

Web UI en DSpace7 - Angular

Conceptos básicos de Angular.

- **component** → controla una parte de la pantalla (vista)
- **template** → plantilla HTML que define cómo renderizar un componente.
- **style** → estilos CSS/SCSS
- **theme** → conjunto de templates y estilos que componen el aspecto del sitio



Web UI en DSpace7 - Temas

DSpace7 provee tres capas/directorios de temas por defecto:

- Base Theme
- DSpace Theme
- Custom Theme
 - DSpace permite crear un **tema personalizado** extendiendo o copiando este tema.

Mejoras y nuevas funciones

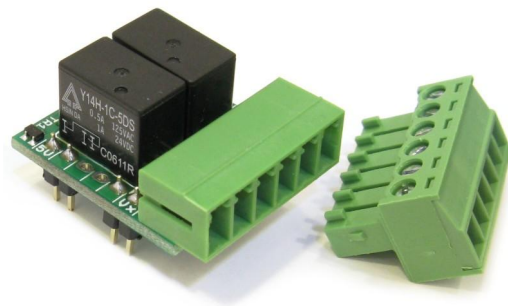
¿Para qué se extiende el software?

- cambios estéticos
- optimización de procesos / circuitos de trabajo
- detección y corrección 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)

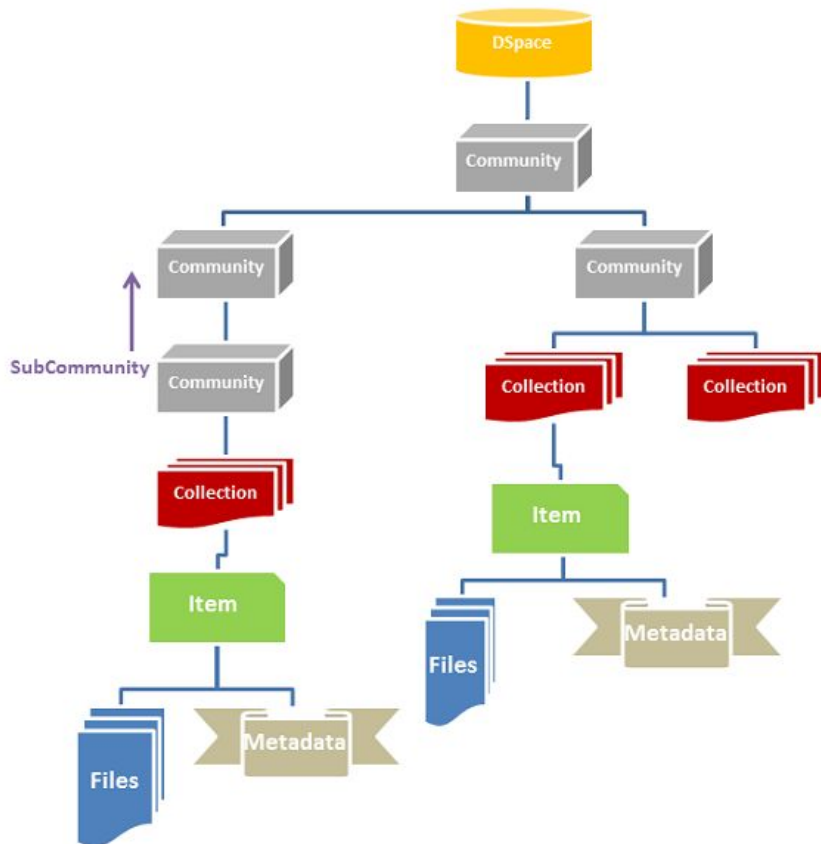


La arquitectura, los lenguajes, las librerías, las herramientas y las licencias que componen el software determinarán la facilidad (y viabilidad) de personalizar o mejorar el software (del repositorio, y cualquier otro software)

Modelo de contenidos

Representación, almacenamiento y gestión de datos y metadatos

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



Repositorio 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]

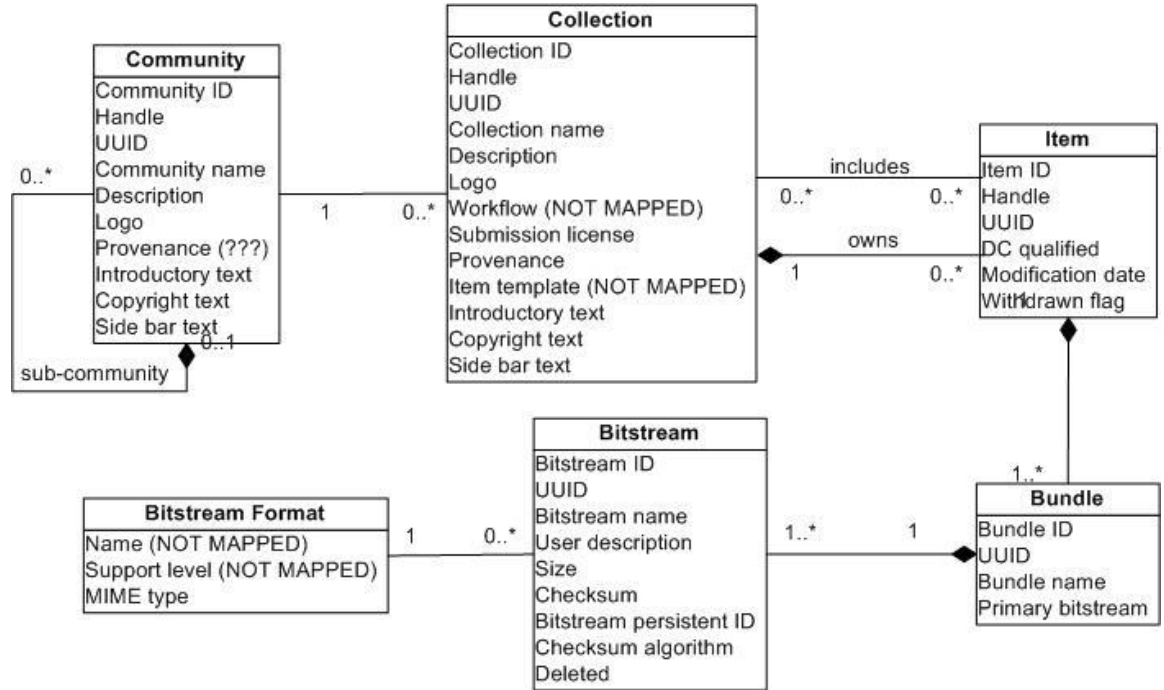
▶ [Anales LEMIT. Serie III](#) [6]

▶ [CIOp](#) [198]

Centro de Investigaciones Ópticas

Modelo de datos de DSpace

- 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
- PDF
- 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 displays 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 shows the breadcrumb 'Centros / LINTA / III Jornadas "Turismo y desarrollo". Turismo cultural: perspectivas y desafíos' and the document title 'El Falcon Verde como lugar simbólico y la posibilidad de su activación para un uso turístico responsable'. A 'Resumen:' section follows, containing text about the Falkland Islands and tourism. A 'Descargas' section on the right offers a 'Documento completo' as a 'Archivo PDF (5.525Mb)'. A red box highlights the download area.

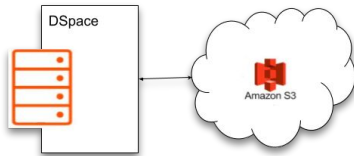
Modelo de contenidos - BitStore

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:** los bitstreams se alojan en un directorio local en el servidor:

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

- **S3Store:** los bitstreams se alojarán en la nube (en el Cloud Storage de [Amazon S3](#)).



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	http://creativecommons.org/licenses/by/4.0/	
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 extendido 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:
 - o 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



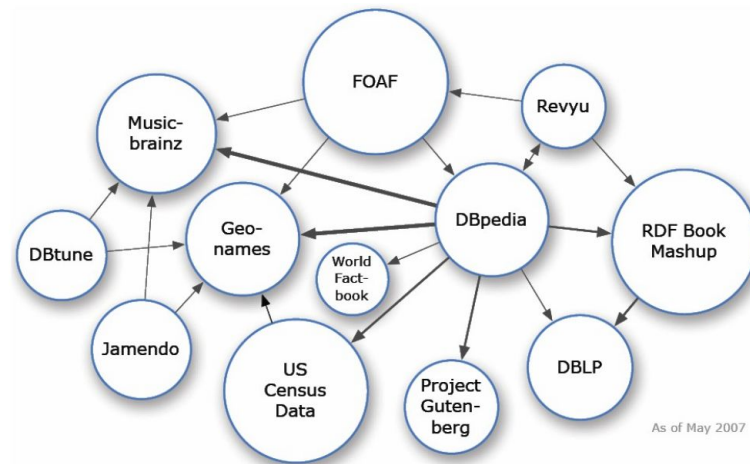
Gestión de autoridades

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

En **SEDICI** y **CIC-DIGITAL** la gestión de autoridades se realiza en un 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.
- Expone su contenido mediante **REST**

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

- 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

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)	http://digital.cic.gba.gob.ar/auth/node/86555	600
Perez, Matías F.	http://digital.cic.gba.gob.ar/auth/node/204702	600
Attribution 4.0 International (BY 4.0)	http://creativecommons.org/licenses/by/4.0/	600
UNLP	http://digital.cic.gba.gob.ar/auth/node/86555	500
Perez, Matías G.		-1

Vías de ingesta

Circuitos de carga y gestión de formularios

Vías de ingesta

Cómo ingresan los recursos al repositorio?

- **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)

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

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 precargados

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

Web UI en DSpace - Formularios

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

- Un **formulario de DSpace** se compone de **múltiples formularios HTML**
- Cada Formulario se descompone en una serie de secciones que agrupan los campos de entrada para los metadatos

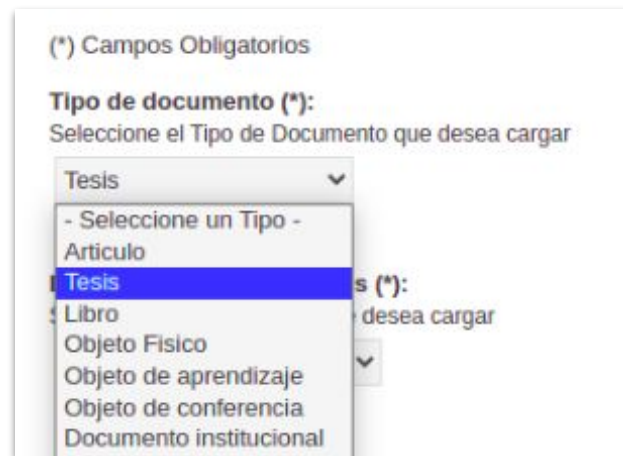
Web UI en DSpace - Formularios

Distintos conjuntos de metadatos para distintos **tipos de contenidos**

- **Items 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**.*



(*) Campos Obligatorios

Tipo de documento (*):
Seleccione el Tipo de Documento que desea cargar

Tesis ▼

- Seleccione un Tipo -
- Artículo
- Tesis**
- Libro
- Objeto Físico
- Objeto de aprendizaje
- Objeto de conferencia
- Documento institucional

s (*):
desea cargar ▼

Web UI en DSpace - Formularios

Restricciones y validaciones en el formulario

- Input fields (campos de entrada)
 - texto
 - selección
 - párrafo
- Dominios (tipos de dato)
 - números
 - fechas
 - si/no (boolean)
- Validaciones
 - campos obligatorios
 - selección de al menos x ítems

Idioma (*):
Idioma principal de la obra

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

Día Mes Año

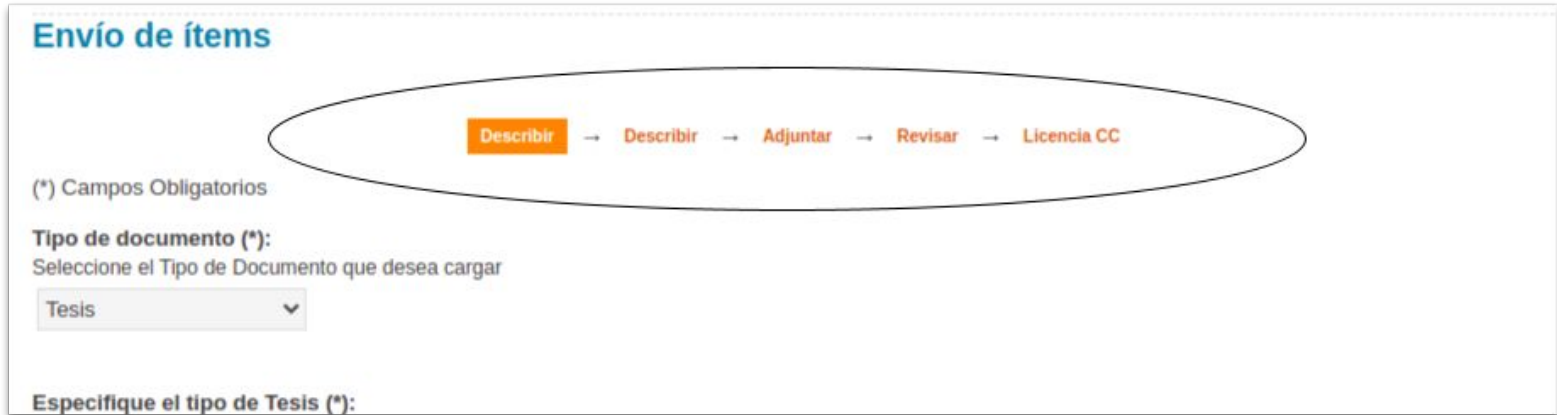
Localización Física:
URL de acceso al catálogo de la biblioteca o información suficiente para ubicar físicamente al recurso (ej.: Museo de Física, vitrina 12, estante 1). Es obligatorio cuando el recurso no es FULLTEXT
 [+ Agregar Otro](#)

Texto completo (*):
Indica si se cuenta con el texto completo de la obra

Entidad de origen (*):
Institución donde el documento fue creado
 [+ Eliminar](#)
 [+ Agregar Otro](#)

Web UI en DSpace - Formularios

- Cada formulario tiene varias etapas (steps)
 - El step “describe” define los metadatos de carga del ítem
- Para cada etapa se presentan varios campos



Envío de ítems

Describir → Describir → Adjuntar → Revisar → Licencia CC

(*) Campos Obligatorios

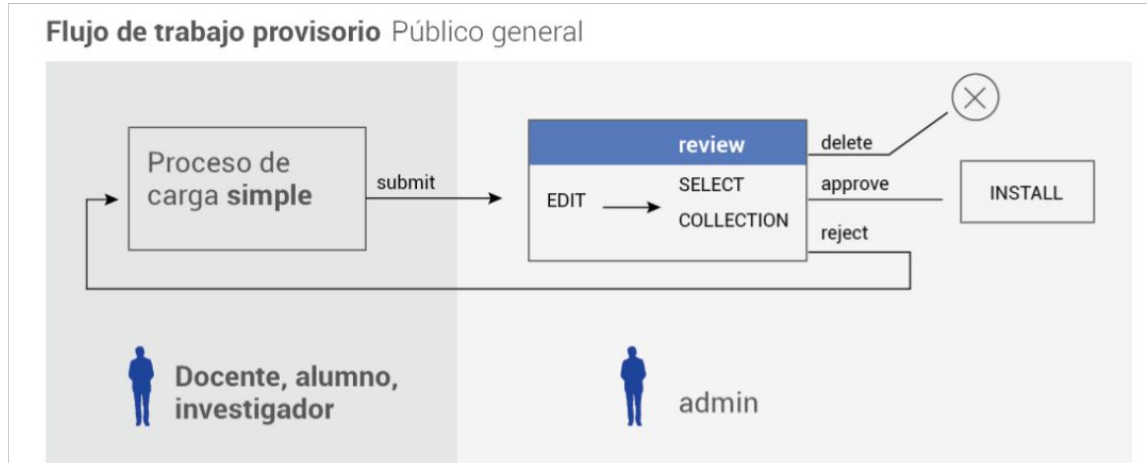
Tipo de documento (*):
Seleccione el Tipo de Documento que desea cargar

Tesis ▼

Especifique el tipo de Tesis (*):

Web UI en DSpace - Workflows

- Distintos workflows (o flujos de trabajo) con sus formularios
- Cada workflow se compone de etapas o steps
 - Carga, Revisión, Aceptación o Rechazo, etc.
- Una colección puede tener su propio workflow.



Localización de recursos

Identificadores persistentes. Exploración y búsqueda

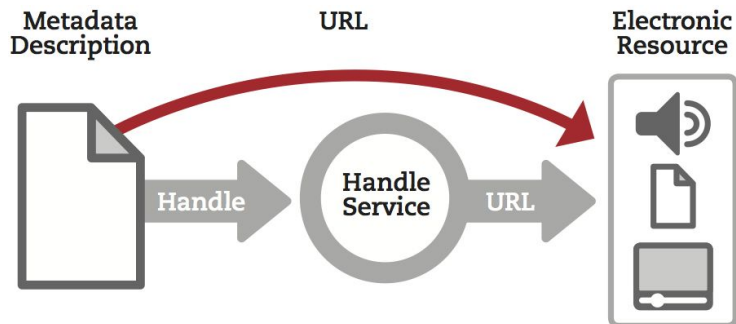
Identificadores persistentes

- Asignan URLs **fijas** a recursos en línea
- Garantizan el acceso a los recursos (incluso si estos cambian su URL de acceso)

sedici.unlp.edu.ar/handle/10915/54681 ⇒ hdl.handle.net/10915/54681

revistas.ucr.ac.cr/index.php/eciencias/article/view/23690 ⇒ 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

¿Quién realiza la redirección?

Identificadores persistentes

Los recursos pueden tener **más de un Identificador Persistente:**

- 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

Descriptor: 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/)

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

Created on: 3 de junio de 2016
Available in SEDICI since: 3 de junio de 2016

Identificadores persistentes

Los identificadores persistentes pueden aplicarse a:

- **documentos de texto** (artículos, tesis, libros)
- **datos de investigación** bases de datos, imágenes, audios
- **personas (Ej. ORCID)**
- **instituciones (Ej. ROR)**

¿Dónde se usan?

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

The image shows a Scopus profile for Gonzalo L. Villarreal. The profile includes the name, ORCID ID (0000-0002-3602-8211), and a link to the public version. Below this, the Scopus search results page is visible, showing the author's name, affiliation (Universidad Nacional de La Plata), and the ORCID ID link.

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
<http://orcid.org/0000-0002-3602-8211>

Servicios para localizar contenidos

Mecanismos para ayudar a los usuarios a acceder a los contenidos que aloja



- Exploración
- Búsqueda
 - ◆ por metadatos
 - ◆ por texto completo
- Faceting

Servicios para localizar contenidos - 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 posgrado 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](#)

Servicios para localizar contenidos - 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

Servicios para localizar contenidos - 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

The screenshot displays a search interface with a 'Refine su búsqueda' (Refine your search) sidebar on the left and search results on the right. The sidebar is divided into three sections: 'Tipo de documento' (Document type), 'Fecha de publicación' (Publication date), and 'Materia' (Subject). The 'Fecha de publicación' section is highlighted with a blue circle around the '2010 - 2016 (100)' option. The search results on the right show a search bar, the text 'PREBI-SEDICI Resultados de su búsqueda...', and three search results, each starting with 'Objeto de conferencia' (Conference object).

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 A
Discurso de inauguración del BIREADIAL 2013

Búsqueda por texto completo

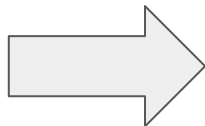
Permite localizar objetos digitales más allá de sus metadatos

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

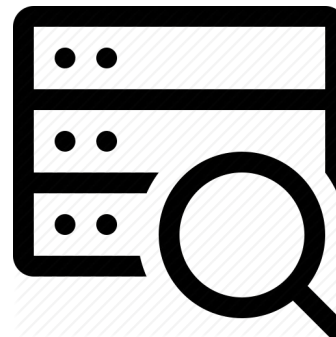
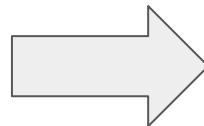
Requiere extraer el texto de los OD e integrarlo en un sistema de búsquedas



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)



Interoperabilidad

Servicios, protocolos y directrices

Servicios de Interoperabilidad

Los repositorios digitales deben pensarse como sistemas interoperables 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

Tener en cuenta...

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

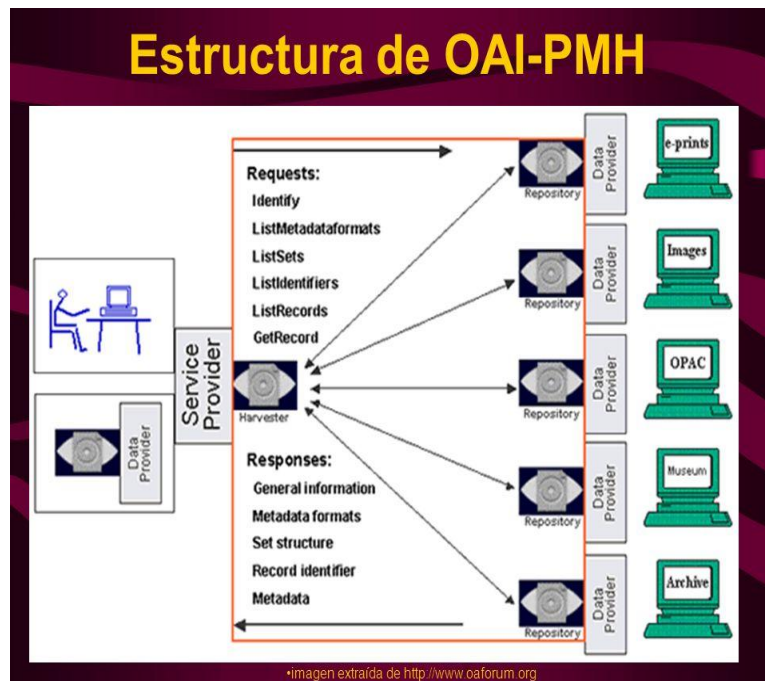
Interoperabilidad - OAI PMH

¿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/interaccionan los Service y Data Providers?

Mediante:

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

Interoperabilidad - OAI PMH - Directrices

¿Cuál es su objetivo?

- La **normalización** de la representación de los metadatos
- 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>)
- BASE (<https://www.base-search.net/>)
- OpenAire

Interoperabilidad - OAI PMH - DSpace

DSpace como data provider (OAI 2.0)

Configuración flexible y adaptable basada en 4 conceptos:

- **Context** → Define el conjunto de ítems a exponer dependiendo el service provider (ej: snrd, openaire)
- **Filter** → Define qué condiciones debe cumplir un ítem para ser expuesto en un contexto
- **Transformer** → Realiza cambios en los metadatos antes de exponerlos en OAI, para que cumplan con las directrices
- **Format** → mapea los metadatos al perfil expuesto
 - De dcterms.creator.author a dc.creator

Interoperabilidad - RSS y OpenSearch



RSS

- Sindicación de noticias y contenidos en línea

OpenSearch

- Conjunto de tecnologías que permiten publicar los **resultados de una búsqueda** en un formato adecuado para la sindicación y agregación
- 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;

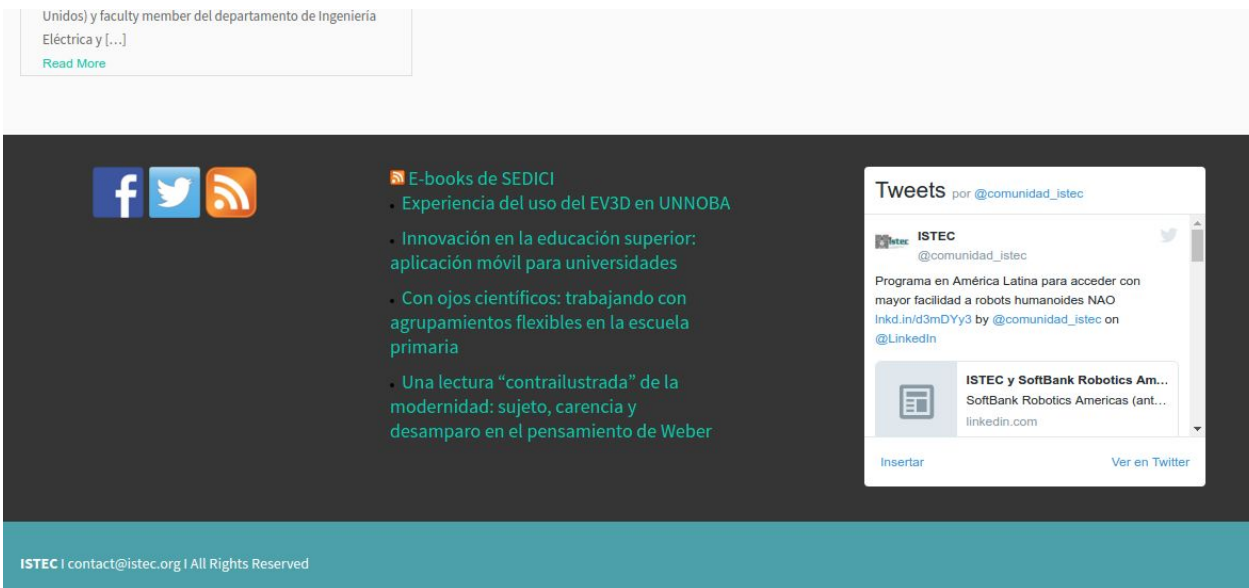
Aislamiento, caracterización genética y análisis filogenético de *Rickettsia massiliae*, Ciudad Autónoma de Buenos Aires
Cicuttin, Gabriel L.;

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

Interoperabilidad - REST

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

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
 - ◆ Incompleto hasta la versión 6
 - ◆ Nueva API completa desde la versión 7
- Provee varios *endpoints* para la manipulación de los datos

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

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://philosophy-science-humanities-controversies.com/XML/sru.php?version=1.1&operation=searchRetrieve&query=dc.title=Darwinism>

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
- Fácil de integrar con otras aplicaciones
 - ◆ Existen librerías para desarrolladores en PHP, Java, Ruby, Python (<http://swordapp.org/>)
- Recibe un XML, que es procesado y transformado a la representación interna del repositorio
- Permite la transferencia de metadatos y archivos.
- 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)



The screenshot displays the 'Conector de depósito de importación/exportación SWORD' interface. It includes a form for deposit details (Punto de depósito, Nombre usuario/a, Contraseña) and a table of items. A green arrow points from the 'Tareas en cola' table to the 'Tareas del flujo de trabajo' section.

NÚMERO	TÍTULO	AUTORES/AS
<input checked="" type="checkbox"/> VOL. 1 (2008)	Aplicación de proceso oxidativo avanzado basado en fotocatalizadores heterogéneos (TiO ₂ /UV-vis) para el pre-tratamiento de efluente lácteo	
<input checked="" type="checkbox"/> VOL. 1 (2011)	Avaliação comparativa de iscas atrativas a partir de registros 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 (2011)	Avaliação da Redução da Poluição do Chorume Tratado por Processo Fotoquímico	

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 iscas 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 (PREMIS)
- 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, Matomo (Piwik), awstats) y estadísticas propias de un repositorio



Lo que se viene...

- DSpace 7
 - Modelo de datos flexible
 - Entidades relacionadas
 - FrontEnd basado en Angular
 - API REST completa basada en HATEOAS
 - Submission y workflow renovados
 - Demo <https://dspace7-demo.atmire.com/>
- Next Generation Repositories:
 - El OD como objeto principal del repositorio
 - Interoperabilidad
 - El conjunto de repositorios como una red interconectada
 - Sostenibilidad a largo plazo de los recursos.
 - <https://ngr.coar-repositories.org/>

