



UNIVERSIDAD
NACIONAL
DE LA PLATA

SEDICI

REPOSITORIO INSTITUCIONAL DE LA UNLP



Capacitación IUPA

Clase 3

Pablo de Albuquerque, Santiago Tettamanti, Ariel Lira
{pablo, santit, alira}@sedici.unlp.edu.ar

PREBI
prebi.unlp.edu.ar



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

04 de Mayo de 2022

Índice

- Vías de ingesta
- WF y Submission
- Autoridades
- Identificadores persistentes
- Autorización y Autenticación



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



Submission y workflow



Submission

Proceso que permite enviar un nuevo **item** al repositorio.

- Se pueden definir varios procesos de submission
- Consisten en una serie de "pasos"
- Se definen en [item-submission.xml](#)

Item puede ser cualquier entidad
(Ejemplo Person, Organization, etc)

```
<submission-map>
  <name-map collection-handle="default" submission-name="traditional"/>
  <name-map collection-handle="12345.6789/42" submission-name="custom" />
  ...
</submission-map>

<submission-definitions>
  <submission-process name="traditional">
    ...
  </submission-process>
  <submission-process name="custom">
    ...
  </submission-process>
</submission-definitions>
```

Submission - Mapeos

- Se definen también las colecciones a las que apunta cada **submission process**:
 - En este ejemplo existen dos submission process
 - “**autoarchive**” relaciona los ítems enviados a la colección “**11746/2**”
 - “**traditional**” relaciona los ítems enviados al resto de las colecciones.

```
<submission-map>
  <name-map collection-handle="default" submission-name="traditional" />
  <name-map collection-handle="11746/2" submission-name="autoarchive" />
</submission-map>
```



Submission - Submission process

Cada submission process define los steps o "pasos" que lo componen.

Cada "paso" se corresponde a una o varias "secciones" en la interfaz

Por default

- Select Collection
- Describe (uno o varios)
- Upload
- License

```
<submission-process name="traditional">
  <step id="collection"/>
  <!-- Describe step 1 -->
  <step id="traditionalpageone"/>
  <!-- Describe step 2 -->
  <step id="traditionalpagetwo"/>
  <step id="upload"/>
  <step id="license"/>
</submission-process>
```

DSpace Communities & Collections All of DSpace

Home > Edit Submission

Drop files to attach them to the item, or browse

Collection 3-step Workflow collection

Describe

Author

Author

Enter the author's name (Family name, Given names).

Title *

Title

You must enter a main title for this item.

Describe

Upload files

Deposit license

Submission - Steps

Cada **step** se define en la sección “step-definitions”.

- Cada step se corresponde con una clase java de la API de DSpace.
- Se puede definir la visibilidad de cada step (workflow o submission) a través de un Scope

```
<step-definitions>

  <step-definition id="collection">
    <heading></heading>
    <processing-class>org.dspace.app.rest.submit.step.CollectionStep</processing-class>
    <type>collection</type>
    <scope visibility="hidden" visibilityOutside="hidden">submission</scope>
  </step-definition>

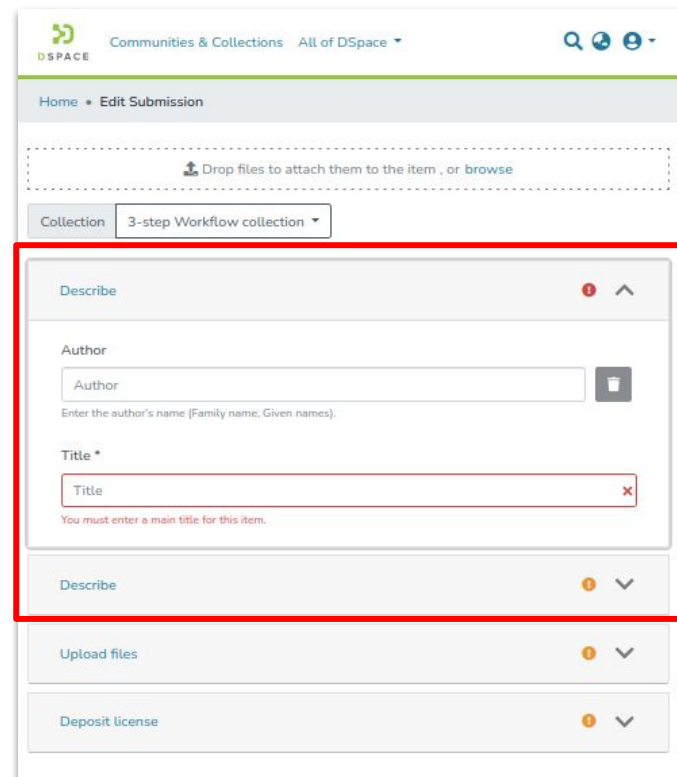
  <step-definition id="traditionalpageone" mandatory="true">
    <heading>submit.progressbar.describe.stepone</heading>
    <processing-class>org.dspace.app.rest.submit.step.DescribeStep</processing-class>
    <type>submission-form</type>
  </step-definition>
  ...
</step-definitions>
```



Submission - Describe Steps

A diferencia del resto de los steps, requieren una configuración adicional.

- Define los formularios de carga para cada colección, comunidad y default.
- En estos formularios se definen los metadatos que luego describirán al ítem en cuestión.
- Cada formulario se divide en páginas con campos



The screenshot shows the 'Edit Submission' page in DSpace. The page header includes the DSPACE logo, 'Communities & Collections', and 'All of DSpace'. The breadcrumb trail is 'Home > Edit Submission'. A dashed box contains the instruction 'Drop files to attach them to the item, or browse'. Below this is a 'Collection' dropdown menu set to '3-step Workflow collection'. The main content area is a list of steps: 'Describe', 'Upload files', and 'Deposit license'. The 'Describe' step is highlighted with a red border and contains the following fields:

- Author**: A text input field with the value 'Author' and a trash icon. Below it, the text reads 'Enter the author's name (Family name, Given names)'.
- Title ***: A text input field with the value 'Title' and a close icon. Below it, the text reads 'You must enter a main title for this item'.

Submission - Describe Steps

- Se configuran desde el archivo [submission-forms.xml](#)
- El nombre del formulario definido debe coincidir con el nombre del step.

```
<form-definitions>
  <form name="traditionalpageone">
    <row>
      <field>
        <dc-schema>dc</dc-schema>
        <dc-element>title</dc-element>
        <dc-qualifier></dc-qualifier>
        <repeatable>>false</repeatable>
        <label>Title</label>
        <input-type>onebox</input-type>
        <hint>Enter the name of the file.</hint>
        <required>You must enter a name for this file</required>
      </field>
    </row>
```



Submission - Describe Steps - Formulario

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 -

Articulo

Tesis

Libro

Objeto Físico

Objeto de aprendizaje

Objeto de conferencia

Documento institucional

s (*)

desea cargar



Submission - Describe Steps - 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](#)

Demo

Personalización del submission



Workflow

Workflow es el proceso que permite la revisión de los ítems que fueron cargados al repositorio

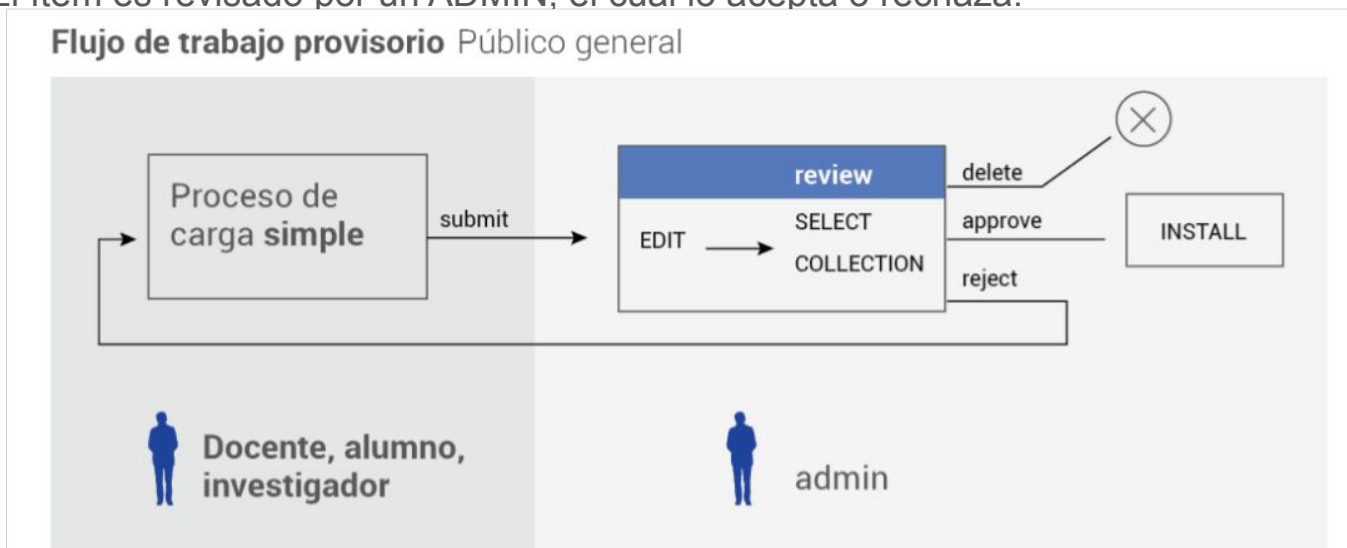
- Se aplica para ítems que provienen del submission o que fueron importados
- Se compone a partir de pasos y transiciones entre ellos
- Los pasos más comunes son:
 - Edición: revisión y mejora de metadatos previamente cargados
 - Decisión: aprobación o rechazo en función si está apto o no para su publicación
- De acuerdo al circuito, el circuito termina cuando se publica o rechaza el ítem en revisión.



Workflow - Casos de uso

Workflow Autoarchivo (para público en gral)

1. Un usuario registrado, sin pertenencia a los grupos *-ADMIN, envía un ítem al repositorio
2. El ítem es revisado por un ADMIN, el cual lo acepta o rechaza.



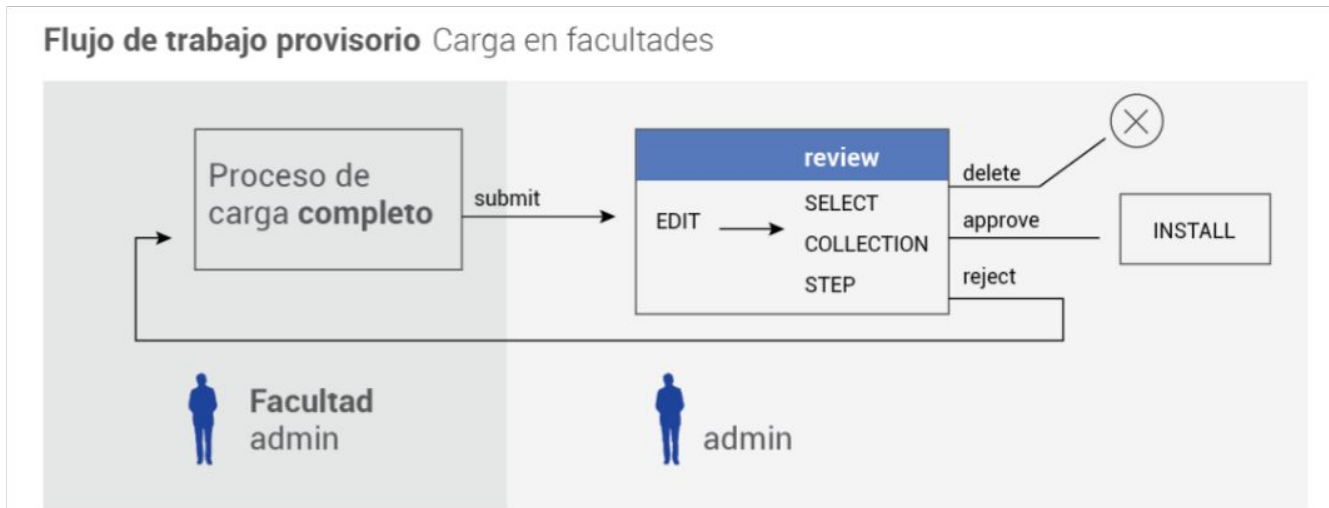
- Por lo general en este flujo sólo se permite el depósito en la colección **"Autoarchivo"**



Workflow - Casos de uso

Workflow Dependiente

1. El ADMIN de una COMUNIDAD o COLECCION carga material es una de sus colecciones
2. Un ADMIN corrige y aprueba o rechaza el ítem



Este flujo de trabajo apunta al control central de todo lo que se publica.

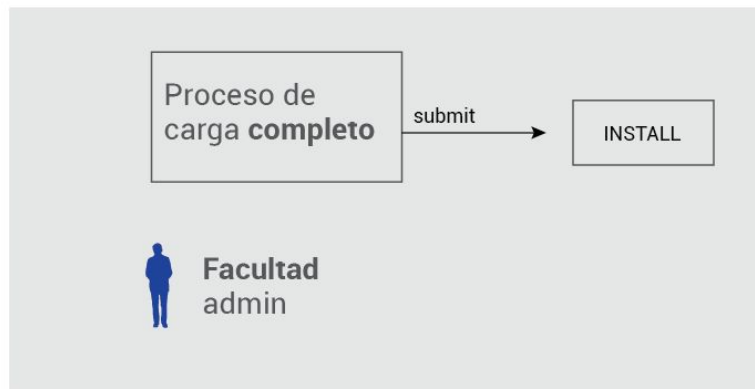


Workflow - Casos de uso

Comunidad Independiente

1. Un ADMIN de una COLECCIÓN/COMUNIDAD carga y publica un material en cualquiera de sus colecciones

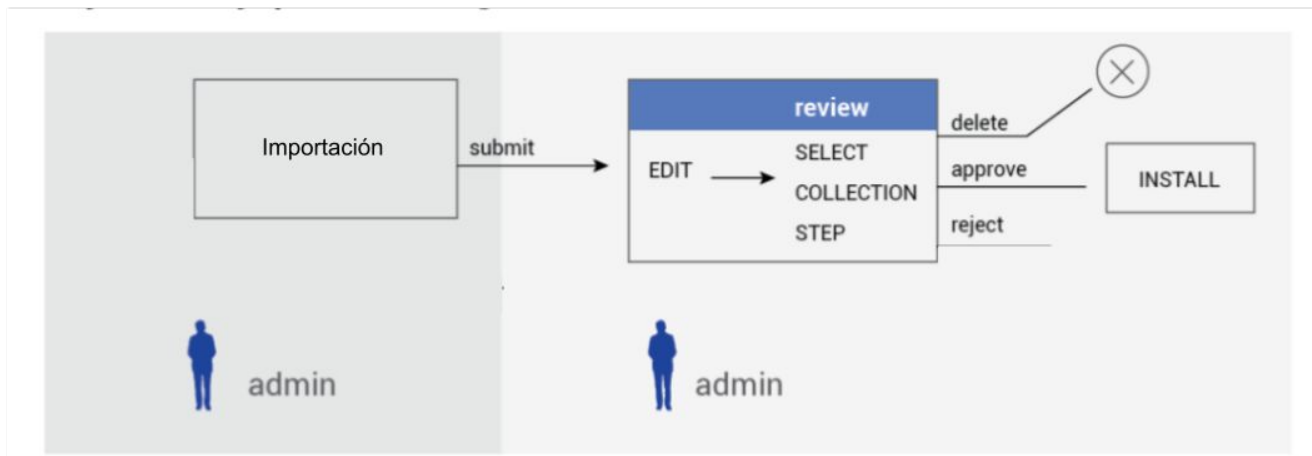
Flujo de trabajo definitivo Carga en facultades



Workflow - Casos de uso

Workflow a partir de una importación

1. Un ADMIN realiza una importación de nuevos ítems a partir de un SAF o CSV.
2. Se configura la importación de manera tal que estos ítems pasen por un proceso de revisión.
3. Un ADMIN corrige y aprueba o rechaza los ítems



Configuración de Workflow

En el archivo [workflow.xml](#) se encuentra la configuración sobre los workflows creados.

```
<bean class="org.dspace.xmlworkflow.XmlWorkflowFactoryImpl">
  <property name="workflowMapping">
    <util:map>
      <entry key="defaultWorkflow" value-ref="defaultWorkflow"/>
<!--      <entry key="123456789/5" value-ref="scoreReview"/>-->
    </util:map>
  </property>
</bean>
```



Configuración de workflow - Ejemplo

"Asignar el WF 'comunidad independiente' a la comunidad de Facultad de Cs. Económicas con handle 123456789/7"

Para esto, agregar la siguiente
<entry> en el archivo workflow.xml:

```
<bean class="org.dspace.xmlworkflow.XmlWorkflowFactoryImpl">
  <property name="workflowMapping">
    <util:map>
      <entry key="defaultWorkflow" value-ref="defaultWorkflow"/>
      <entry key="123456789/7" value-ref="comunidad independiente"/>
    </util:map>
  </property>
</bean>
```



Configuración de workflow - definición de workflows

Cada workflow está compuesto por una serie de pasos o Steps.

- El workflow por defecto:
 - Review step
 - Edit step
 - Final Edit Step

```
<bean name="defaultWorkflow" class="org.dspace.xmlworkflow.state.Workflow">
  <property name="firstStep" ref="reviewstep"/>
  <property name="steps">
    <util:list>
      <ref bean="reviewstep"/>
      <ref bean="editstep"/>
      <ref bean="finaleditstep"/>
    </util:list>
  </property>
</bean>
```



Configuración de workflow - Workflow Steps

Cada Steps tiene asociado.

- Un rol → define qué conjunto de usuarios podría ejecutar este paso.
- userSelectionMethod → define cómo los usuarios pueden interactuar con el paso actual.
- Una serie de acciones (actions)
- Una serie de pasos siguientes (outcomes)

```
<bean name="reviewstep" class="org.dspace.xmlworkflow.state.Step">
  <property name="userSelectionMethod" ref="claimaction"/>
  <property name="role" ref="reviewer"/>
  <property name="outcomes">
    <util:map>
      <entry key="#{ T(org.dspace.xmlworkflow.state.actions.ActionResult).OUTCOME_COMPLETE}" value-ref="editstep"/>
    </util:map>
  </property>
  <property name="actions">
    <util:list>
      <ref bean="reviewaction"/>
    </util:list>
  </property>
</bean>
```



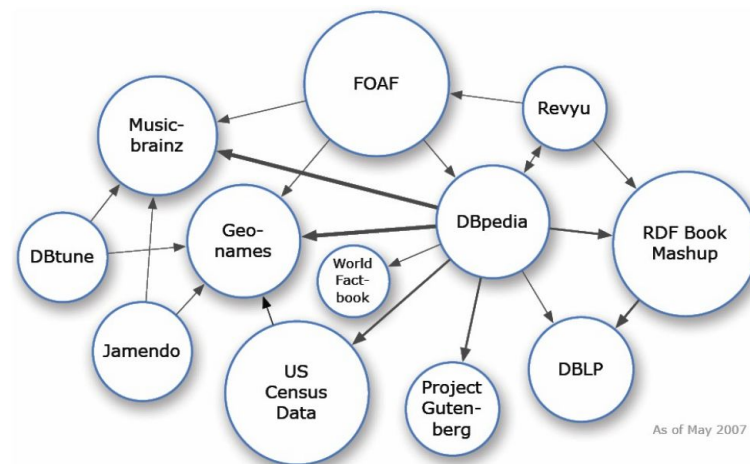
Autoridades



Autoridades

Uso de Vocabularios controlados gestionados en otros sistemas

- Tesoros
- Sistemas de clasificación / Taxonomías
- 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**.



Autoridades

- **Autoridad:** conjunto de entidades normalizadas
 - **Registro de autoridad:** información asociada con una entidad
 - **Clave de autoridad:** código que identifica unívocamente una entidad.



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



Autoridades - API

DSpace incluye una API para el uso de autoridades en los metadatos.

- la fuente puede ser externa o interna al repositorio
- para integrar nuevas autoridades deben usarse las 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.



Autoridades - Plugins

DSpace cuenta con un mecanismo para *manejo de autoridades*, que permite la **selección de valores autorizados** durante la carga de metadatos al repositorio.

- Implementan la interfaz [ChoiceAuthority](#).
- Autoridades **predefinidas en DSpace** (*DCInputAuthority*, *SHERPARoMEOPublisher*, etc.)
- Capacidad de agregar nuestras **propias autoridades**.

```
public abstract class SPARQLAuthorityProvider implements ChoiceAuthority {  
  
    public Choices getMatches(String field, String text, Collection collection,  
        int start, int limit, String locale) {  
        if (text == null)  
            text = "";  
        else  
            text = text.replace("\n", "");  
  
        ParameterizedSparqlString query = this.getSparqlSearchByTextQuery(  
            field, text, locale);  
        Choice[] choices = this.evalSparql(query, start, limit);  
        log.trace(choices.length + "matches found for text " + text);  
        return new Choices(choices, start, limit, Choices.CF_ACCEPTED, false);  
        // TODO hasMore??  
    }  
  
    public final Choices getBestMatch(String field, String text,  
        Collection collection, String locale) {  
        return this.getMatches(field, text, collection, 0, 1, locale);  
    }  
  
    public String getLabel(String field, String key, String locale) {  
  
        ParameterizedSparqlString query = this.getSparqlSearchByIdQuery(field,  
            key, locale);  
        Choice[] choices = this.evalSparql(query, 0, 1);  
        if (choices.length == 0)  
            return null;  
        else  
            return choices[0].label;  
    }  
}
```



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.



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



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



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
Cañueto, 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
Cañueto, Matías G.		-1



Autoridades

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

Drupal:

- Cuenta con un conjunto de **módulos** que facilitan el modelado.
- Comunicación entre esta plataforma y ambos repositorios es mediante una **API REST**



Drupal™



Identificadores persistentes



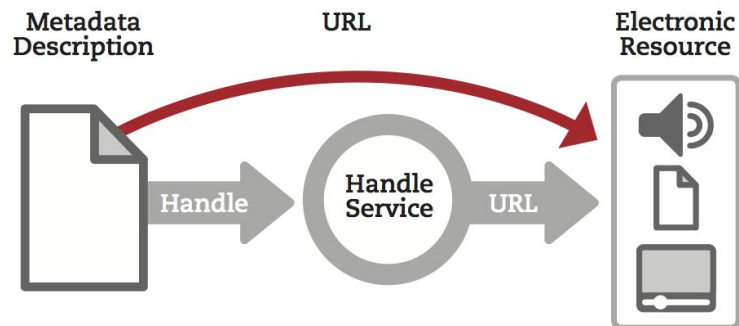
Identificadores persistentes

Permiten identificar un recurso de manera independiente a la dirección del repositorio o sitio donde dicho recurso se encuentra.

sedici.unlp.edu.ar/handle/10915/54681 ⇒ <http://hdl.handle.net/10915/54681>

revistas.ucr.ac.cr/index.php/eciencias/article/view/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
- por
 - 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 en DSpace

Dspace permite utilizar identificadores

- **internos:** uuid auto-generados (SIEMPRE)
- **handle:** servidor local
- **doi:** registro externo
- **custom:** según se implemente

Generalmente se utiliza un servidor handle.net local al servidor de DSpace con un prefijo único propio de la institución.



Autorización y Autenticación



Autenticación

A partir de diversos métodos configurables:

- por **password**
- por ip o rango de ips
- por **LDAP**
- por Shibboleth
- X509
- custom

Más info <https://wiki.duraspace.org/display/DSDOC7x/Authentication+Plugins>



Módulo de autorización - Modelo

El esquema de autorización de dspace funciona a partir de 3 elementos:

1. E-Person
 - representa un usuario dentro del sistema.
2. E-Group:
 - representa un grupo de trabajo.
3. Resource Policies
 - representación de los derechos o permisos



Módulo de autorización - E-Person

Los **personas**,

- representadas por la entidad *EPerson*
- se corresponden con los usuarios registrados en el sitio.
- Cada persona tiene al menos:
 - Nombre y Apellido
 - Email
 - Password (hasheada)



Módulo de autorización - E-Group

- Representan un grupo de trabajo.
- Se componen de :
 - múltiples E-Person, y/o
 - otros E-Group
- Existen 2 grupos de usuarios por defecto:
 - **Anonymous**
 - Representa a todos los usuarios, autenticados o no.
 - Cualquier usuario tiene pertenencia a este grupo
 - **Administrator**
 - Contiene los usuarios que son administradores del sistema.
 - representa el rol de SUPERUSUARIO, con autorización plena sobre DSpace
- Todos los usuarios pertenecen al grupo Anonymous

IMPORTANTE: Los permisos de un grupo son heredados por todos los grupos que contiene.



Módulo de autorización - E-Group - Grupos de trabajo

Los grupos también pueden ser usados para configurar los flujos de trabajo.

Por ejemplo:

1. INSTITUTION-ADMIN:

- bibliotecarios y responsables centrales del repositorio
- se encargarían de gestionar todo el contenido del repositorio

2. Facultad-ADMIN:

- bibliotecarios y responsables por cada Facultad que compone a la institución
- gestionarían todo el contenido de las comunidades asociadas a una Facultad

3. COMMUNITY_ADMINS

- Contendría a los grupos INSTITUTION-ADMIN y a todos los Facultad-ADMIN.
- Podría existir únicamente para hacer funcionar los workflows de la UNLP.



Módulo de autorización - Resource Policies - Permisos

El [Authorization system](#) regula **quién** puede hacer **qué** acciones sobre cada **objeto**.

Estos permisos se representan en las [Resource Policies](#) y se componen de:

- **persona o grupo autorizado:**
 - eperson_id/epersongroup_id
- **acción permitida:**
 - action_id
- **sobre qué objeto:**
 - resource_id+resource_type_id
- **desde cuando:**
 - start_date
- **hasta cuando:**
 - end_date



Autorización - Resource Policies - Modelos

Resource Policies: derechos explícitos de la forma

(Objeto, Usuario, Derecho, Fecha inicio, Fecha fin)

<p>Objeto es cualquier DSO</p> <ul style="list-style-type: none">● Comunidad● Colección● Ítem● Bundle● Bitstream	<p>Usuario es:</p> <ul style="list-style-type: none">● E-Person● E-Group	<p>Derecho puede ser:</p> <ul style="list-style-type: none">● READ: ver o descargar● WRITE: modificar datos● ADD: agregar a un contenedor● REMOVE: eliminar de un contenedor● ADMIN: administración plena de elementos
---	--	---

Por defecto, todas las comunidades, colecciones e ítems tienen permiso de READ para el grupo Anonymous



Módulo de autorización - Acciones posibles

READ	Action of reading, viewing or downloading something
WRITE	Action of modifying something
ADD	Action of adding something to a container.
REMOVE	Action of removing something from a container. Different from deletion.
DEFAULT_BITSTREAM_READ	Default Read policies for Bitstreams submitted to container.
DEFAULT_ITEM_READ	Default Read policies for Items submitted to container.
ADMIN	Administrative actions - System Admin, Community Admin, Collection Admin
WITHDRAWN_READ	Action of reading, viewing or downloading an embargoed item.

** un container es una comunidad o colección*



Actividad 3

- Agregar un nuevo metadato al submission-forms.xml, actualizando el formulario de submission.
- Controlar un metadato en el formulario, con los valores definidos como value-pairs
 - Por ejemplo, se podría agregar el metadato snrd-type que contenga los tipos de documentos definidos por el snrd y que esté controlado por un value-pairs que liste los tipos posibles.

