



Red Mexicana de  
Repositorios Institucionales

## Estándares de metadatos y directrices de interoperabilidad de ReMeRI


Dr. Alfredo Sánchez  
Mtro. Antonio Razo




7 de diciembre de 2012

## El Reto

- Cada repositorio institucional tiene características propias
  - modelo de datos (esquema, vistas, índices,...)
  - software de base de datos
  - Interfaces de actualización y consulta
- Pero necesitamos interfaces de consulta y visualización uniformes, para todos


04/12/2012 REMERI 2 

## UNA SOLUCIÓN POSIBLE: ESTÁNDARES DE METADATOS

04/12/2012 REMERI 3 


## ¿Metadatos?—¿para qué?

- Metadatos: Datos (estructurados) acerca de los datos (ejemplos: tipos de datos, formatos, estilos,...)
- A nivel documento u objeto de información, los metadatos indican estructura, organización y facilitan la interpretación
- A nivel colección, pueden facilitar la interoperabilidad
- ... y la construcción de servicios

© 2008-2012, Alfredo Sánchez 


## Estándares de Metadatos

- Existen cientos, o miles... por tanto podría cuestionarse si son estándares
- Ejemplos:
  - Definiciones de tipos de documentos (DTDs) para dominios específicos
    - Química (<http://www.chemistrycentral.com/xml/article.dtd>)
    - Matemáticas (MathML, <http://www.w3.org/TR/MathML3/>)
    - Música (MusicXML, <http://en.wikipedia.org/wiki/MusicXML>)
    - GIS
    - Derecho, etc.
- Ejemplo en bibliotecas tradicionales: MARC
- Promovidos por organizaciones como OASIS, WWW Consortium ([www.w3c.org](http://www.w3c.org)), ANSI (American National Standards Institute), ISO, etc.

© 2008-2012, Alfredo Sánchez 

## Dublin Core (DC)

- Propuesta de estándar "general" para facilitar la localización de recursos en Internet
- Promovido por DCMI, Dublin Core Metadata Initiative (<http://dublincore.org>), integrada por organizaciones en bibliotecas digitales y tradicionales, museos, gobiernos
  - Algunos miembros: NSF, CNRI, NCSA, OAI, OCLC, bibliotecas nacionales de EEUU, Canadá, Australia, Finlandia, ...
- Meta: un "lenguaje" simple de intercambio
- Conjunto de 15 descriptores de recursos

© 2008-2012, Alfredo Sánchez 

## Dublin Core: Descriptores

Contenido	Propiedad Intelectual	Instanciación
<a href="#">Title</a>	<a href="#">Creator</a>	<a href="#">Date</a>
<a href="#">Subject</a>	<a href="#">Publisher</a>	<a href="#">Type</a>
<a href="#">Description</a>	<a href="#">Contributor</a>	<a href="#">Format</a>
<a href="#">Source</a>	<a href="#">Rights</a>	<a href="#">Identifier</a>
<a href="#">Language</a>		
<a href="#">Relation</a>		
<a href="#">Coverage</a>		

© 2008-2012, Alfredo Sánchez



## Dublin Core: Descriptores

- **Title** – Nombre dado al recurso
- **Subject** – Tema del recurso, tomado normalmente de un tesauro
- **Description** – Descripción textual, resumen
- **Source** – Identificador de un recurso origen (ej. URL)
- **Language** – Idioma identificado por 2 caracteres
- **Relation** – Identificador de un segundo recurso asociado al actual
- **Coverage** – Cobertura (espacial o temporal) del recurso
- **Creator** – Persona autor del recurso
- **Publisher** – Organización responsable de publicar el recurso
- **Contributor** – Co-autor
- **Rights** – Referencia (ej. URL) a notas sobre derechos de autor
- **Date** – Fecha de publicación
- **Type** – Tipo del recurso, tomado de un tesauro
- **Format** – Formato, usado para identificar el software necesario para desplegar o reproducir el recurso
- **Identifier** – Identificador del recurso (URL, ISBN, etc.)

© 2008-2012, Alfredo Sánchez

Ejemplo: DC para tesis digitales



## INTEROPERABILIDAD Y OAI-PMH

04/12/2012

REMER

9



## Iniciativa de Archivos Abiertos

- "Open Archives Initiative" (OAI)
- Comunidad que desarrolla y promueve estándares de interoperabilidad para facilitar la difusión de contenidos
- "Archives" identifica colecciones o repositorios de documentos *abiertos* pero típicamente solo accesibles a través de consultas a bases de datos
- Proyectos principales:
  - OAI-PMH – Protocolo para Cosecha de Metadatos
  - OAI-ORE – Especificaciones para reuso e intercambio de objetos compuestos o "agregados"

© 2008-2012, Alfredo Sánchez



## Interoperabilidad en OAI

- Meta: Compartir recursos heterogéneos (distintos formatos, plataformas de hardware y software)
- Mecanismo básico: protocolo compartido (OAI-PMH)
- Participantes:
  - Proveedores de datos – ofrecen colecciones vía OAI-PMH
    - Actualmente, 1854 proveedores registrados, otros sin registrar ([www.openarchives.org/Register/BrowseSites](http://www.openarchives.org/Register/BrowseSites))
  - Proveedores de servicios – construyen servicios (ej. IR) usando metadatos
    - Decenas registrados, muchos más sin registrar
    - <http://www.openarchives.org/service/listproviders.html>

© 2008-2012, Alfredo Sánchez



## OAI-PMH


- Protocolo para cosecha de metadatos (Protocol for Metadata Harvesting)
- Supone que los proveedores de metadatos usarán al menos el estándar Dublin Core, pero pueden ofrecer otros
- Conceptos básicos:
  - Identificador de colecciones
  - Identificador de documentos
  - Formato o estándar de metadatos
  - Conjunto (set): grupo de documentos con un identificador asignado por el proveedor de datos
  - Registro (record) : conjunto de metadatos asignado a cada documento en la colección

© 2008-2012, Alfredo Sánchez




## Peticiónes vía OAI-PMH

- Un servidor debe atender **6 peticiones o verbos**:
  - Identify** – identifica la colección
  - ListMetadataFormats** – obtiene formatos de metadatos disponibles en la colección
  - ListIdentifiers** – obtiene encabezados de objetos (id, fecha, conjunto)
  - ListSets** – Obtiene la organización de conjuntos de la colección
  - GetRecord** – obtiene metadatos de un objeto
  - ListRecords** – obtiene metadatos de objetos filtrando por conjunto o fecha
- Referencia: <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- Un **cosechador** es una aplicación "cliente" que recolecta documentos de un repositorio OAI enviando verbos a un servidor
- Ejemplo: OAI 2.0 Tales (Tesis Digitales UDLAP)

© 2008-2012, Alfredo Sánchez 


## Ejemplos con Tales

- Identify:
  - [http://catarina.udlap.mx/u\\_dl\\_a/tales/oai/requestETD.jsp?verb=Identify](http://catarina.udlap.mx/u_dl_a/tales/oai/requestETD.jsp?verb=Identify)
- ListMetadataFormats
  - [http://catarina.udlap.mx/u\\_dl\\_a/tales/oai/requestETD.jsp?verb=ListMetadataFormats](http://catarina.udlap.mx/u_dl_a/tales/oai/requestETD.jsp?verb=ListMetadataFormats)
- ListIdentifiers
  - [http://catarina.udlap.mx/u\\_dl\\_a/tales/oai/requestETD.jsp?verb=ListIdentifiers&metadataPrefix=oai\\_dc](http://catarina.udlap.mx/u_dl_a/tales/oai/requestETD.jsp?verb=ListIdentifiers&metadataPrefix=oai_dc)

© 2008-2012, Alfredo Sánchez 

## Ejemplos con Tales

- GetRecord
  - [http://catarina.udlap.mx/u\\_dl\\_a/tales/oai/requestETD.jsp?verb=GetRecord&identifier=oai:ciria.udlap.mx:u-dl-a/tesis/2011017614981&metadataPrefix=oai\\_dc](http://catarina.udlap.mx/u_dl_a/tales/oai/requestETD.jsp?verb=GetRecord&identifier=oai:ciria.udlap.mx:u-dl-a/tesis/2011017614981&metadataPrefix=oai_dc)
- ListSets
  - [http://catarina.udlap.mx/u\\_dl\\_a/tales/oai/requestETD.jsp?verb=ListSets](http://catarina.udlap.mx/u_dl_a/tales/oai/requestETD.jsp?verb=ListSets)
- ListRecords
  - [http://catarina.udlap.mx/u\\_dl\\_a/tales/oai/requestETD.jsp?verb=ListRecords&from=2008-01-15&set=103202&metadataPrefix=oai\\_dc](http://catarina.udlap.mx/u_dl_a/tales/oai/requestETD.jsp?verb=ListRecords&from=2008-01-15&set=103202&metadataPrefix=oai_dc)

© 2008-2012, Alfredo Sánchez 

## METADATOS E INTEROPERABILIDAD EN REMERI

04/12/2012 REMERI 16 

Metadatos (OAI_DC Type)	Metadatos REQUISITO para cumplir con REMERI	Metadatos REQUISITO para cumplir con directrices de DRIVER 2.0 (2008) y con proyecto intranacional*
Title	Title	Title
Identifier	Identifier	Identifier
Description		(No obligatorio; y se recomienda en español e inglés)
Subject		
Creator	Creator	Creator
Type	Type	Type
Date	Date	Date
Rights		Rights
Publisher		
Language		Language
Format		
Coverage		
Contributor		
Source		
Relation		

Dic-05-12 Día virtual sobre capacitación en Repositorios Institucionales 17 

### Documento de requisitos técnicos y directrices de interoperabilidad para REMERI

**Acerca del metadato dc:type**  
 En el caso especial de dc:type se debe especificar con mayor prioridad los siguientes tipos documentales de acuerdo al tipo de conocimiento al que se hace referencia:

- Article (Artículo, Artigo, Journal Article, Artículo de Revista)
- DoctoralThesis (Tesis de Doctorado, Tesis Doctoral)
- MasterThesis (Tesis de maestría)
- Report (Informe técnico, referido a resultados de investigación)

**Acerca del metadato dc:rights**  
 Se desearán aquellos registros que posean en la primer ocurrencia alguno de los siguientes valores:

- Acceso Abierto
- Licencia específica de Creative Commons:
  - Atribución (CC BY)
  - Atribución-CompartirIgual (CC BY-SA)
  - Atribución-SinDerivadas (CC BY-ND)
  - Atribución-NoComercial (CC BY-NC)
  - Atribución-NoComercial-CompartirIgual (CC BY-NC-SA)
  - Atribución-NoComercial-SinDerivadas (CC BY-NC-ND)

Se recomienda el uso de la segunda ocurrencia de dc:rights la cual especificará el tipo de licencia y uso del objeto digital (tipo de licencia Creative Commons: <http://creativecommons.org/licenses>).

Dic-05-12 Día virtual sobre capacitación en Repositorios Institucionales 18 

## Documento de requisitos técnicos y directrices de interoperabilidad para REMERI

**Acerca del metadato dc:language**  
Se va a aceptar ISO con dos o tres caracteres y siempre resultará convertido a ISO-639-3 caracteres como las directrices recomiendan ([www.sil.org/iso639-3/codes.asp](http://www.sil.org/iso639-3/codes.asp)).

Español (spa)  
Inglés (eng)

**Acerca del metadato dc:description**  
En el caso de dc:description, para las instituciones miembro de REMERI el mismo no será controlado. Sin embargo para la red nacional será responsabilidad del componente cosechador/agregador de enviar los resúmenes en el idioma original e inglés, esto en el orden especificado por las directrices DRIVER 2.0.

**Acerca del metadato dc:date**  
Para todas las fechas se verificará el cumplimiento de la norma ISO-8601. Las fechas que no superen dicha verificación serán eliminadas del registro a incluir en el cosechador.  
yyyy-MM-dd Time HH:mm:ss  
aaaa-MM-dd Hora HH:mm:ss

Dic-05-12 Día virtual sobre capacitación en Repositorios Institucionales 19

## CADA PARTICIPANTE DEBE CONTAR CON UN SERVIDOR OAI-PMH

04/12/2012 REMERIC 20

### Servidor de metadatos de Redalyc

<http://ojs.redalyc.uaemex.mx/ServidorOAI/PagOaiVer.html>

**Servidor Redalyc OAI**  
Generado por Voai

The Open Archives Initiative Protocol for Metadata Harvesting  
(Protocol Version 2.0)

**La petición que usted intentó no es válida!**  
A continuación se muestra una lista de las posibles peticiones para OAI-PMH:

**GetRecord**  
[http://ojs.redalyc.uaemex.mx/redalyc/oai?verb=GetRecord&identifier=oai:redalyc:uaemex.mx:10201406&metadataPrefix=oai\\_dc](http://ojs.redalyc.uaemex.mx/redalyc/oai?verb=GetRecord&identifier=oai:redalyc:uaemex.mx:10201406&metadataPrefix=oai_dc)

**Identify**  
<http://ojs.redalyc.uaemex.mx/redalyc/oai?verb=Identify>

**ListMetadataFormats**  
<http://ojs.redalyc.uaemex.mx/redalyc/oai?verb=ListMetadataFormats>

**ListRecords**  
<http://ojs.redalyc.uaemex.mx/redalyc/oai?verb=ListRecords>

**Peticiones OAI-PMH**

### Servidor de metadatos de Tesis Digitales UDLAP

[http://catarina.udlap.mx/u\\_dl\\_a/tales/oai/](http://catarina.udlap.mx/u_dl_a/tales/oai/)

**UDLAP** **CIRIA**  
UNIVERSIDAD DE LA AMÉRICA PÚBLICA

Tesis Digitales    Acerca de    Búsquedas    Catálogo    Construcción

**The Open Archives Initiative Protocol for Metadata Harvesting**  
(Protocol Version 2.0 of 2002-06-14)  
Registered Data Provider

**Identify**  
This verb is used to retrieve information about a repository, including administrative, identity, and community-specific information.

- `?verb=Identify`

**Error and exceptions conditions**

- `?verb=Identify&metadataPrefix=oai_dc (badArgument)`

**GetRecord**  
This verb is used to retrieve an individual record (metadata) from an item in a repository. Required arguments specify the **Identifier**, or key, of the requested record and the **format of the metadata** that should be included in the record.

### Generadores de Proveedores OAI-PMH

<http://ict.udlap.mx/oai/>

**XOAI | VOAI**

GENERADOR DE PROVEEDORES DE DATOS OAI-PMH PARA COLECCIONES EN RED XML

GENERADOR DE PROVEEDORES DE DATOS OAI-PMH PARA COLECCIONES EN RED RELACIONALES

1CT & Agosto 2007. Raúl Parrón y Antonio Ruiz. INGLES VERSION

Dic-05-12 Día virtual sobre capacitación en Repositorios Institucionales 23

### DSpace OAI-PMH

<https://wiki.duraspace.org/display/DSPACE/Use+the+OAI-PMH+interface>

Dashboard - DSpace    Browse    Log In    Sign Up    Search

Use the OAI-PMH interface

Added by [wiki\\_translator](#), last edited by [Ivan Masjón](#) on Jun 08, 2012 ([view change](#) | [show comment](#))

**Files:**

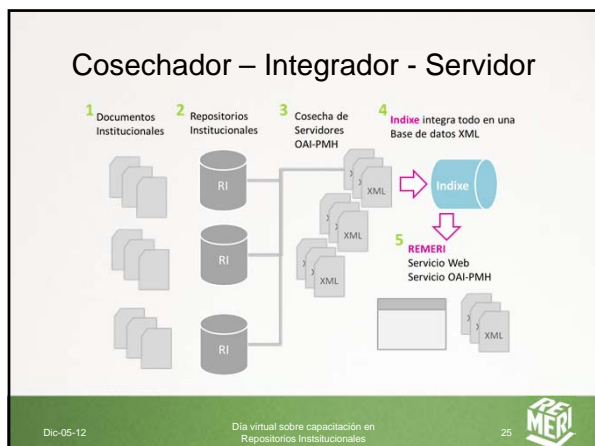
- `dspace-source/buildspace-oai-war`
- `http://_web-address-to-my-dspace/_dspace-oai`

(OAI-PMH base URL)

**Instructions:**

- Whenever you rebuild DSpace (see [Rebuild DSpace](#)), the OAI-PMH (<http://www.openarchives.org/oai/>) interface is automatically rebuilt into the `dspace-oai-war`. In order to enable the OAI-PMH interface, you just need to copy this `dspace-oai-war` file into your `tomcat/webapps/` directory. (It likely will already be there if you followed the installation directions completely).
- If you've never used OAI-PMH before, there is a [Beginner's Tutorial](#) from the Open Archives Forum available at: <http://www.openarchives.org/tutorial/>
- Here's a few basic OAI-PMH "verbs" (i.e. commands) you may wish to familiarize yourself with (all example DSpace URLs should be on a single line):
  - `Identify` - used to retrieve information about a repository.

Dic-05-12 Día virtual sobre capacitación en Repositorios Institucionales 23



### Servidor de Metadatos de REMERI

<http://www.remeri.org.mx/remeri/pservicios.html>

**OAI PARA REMERI**

El servicio de integración Indíxate cuenta con un servidor de metadatos en el estándar OAI-PMH para hacer interoperable la colección:

- URL del Servidor [http://www.remeri.org.mx/evost/rest/db/remeri/oa/oa\\_server.sq](http://www.remeri.org.mx/evost/rest/db/remeri/oa/oa_server.sq)
- Verbo ListSets
- Verbo ListMetadataFormats
- Verbo ListRecords
- Verbo ListIdentifiers
- Validación del servidor usando nuestra herramienta Valida OAI REMERI

De la misma manera Indíxate utiliza los servidores de metadatos de los repositorio institucionales para recopilar la información de cada uno de ellos

Servidor OAI especializado para LA Referencia

- URL del Servidor [http://www.remeri.org.mx/evost/rest/db/remeri/driver/driver\\_server.sq](http://www.remeri.org.mx/evost/rest/db/remeri/driver/driver_server.sq)
- Verbo ListSets
- Verbo ListMetadataFormats
- Verbo ListRecords
- Verbo ListIdentifiers
- Validación del servidor usando nuestra herramienta Valida OAI REMERI Driver

Dic-05-12      Día virtual sobre capacitación en Repositorios Institucionales      26

### Más información...

Dr. Alfredo Sánchez  
 Mtro. Antonio Razo  
 Universidad de las Américas Puebla  
<http://ict.udlap.mx>

Dic-05-12      Día virtual sobre capacitación en Repositorios Institucionales      27