

# Taller de Bibliotecas Digitales

"Construcción automática de servidores de metadatos para Archivos Abiertos (OAI)"

Mtro. Antonio Razo, UDLA-Puebla

# Agenda

- Interoperabilidad
- Open Archives Initiative OAI
- Construcción de servidores de metadatos

# Iniciativa de Archivos Abiertos OAI

**<http://www.openarchives.org>**

- Organización de instituciones interesadas en el acceso libre a la información
- Se origina en 1999 para promover el acceso abierto a “e-prints”—artículos y otros productos de investigación

# OAI

- Desarrolla y promueve estándares de interoperabilidad para la difusión eficiente de contenido
- Define un protocolo para compartir metadatos: “Protocol for Metadata Harvesting” (OAI-PMH)
- Mantiene un centro de información sobre proveedores de datos y de servicios

# Ilustración para conceptos básicos



← resource

item =  
identifier

all available metadata  
about *David*

← item

Dublin Core  
metadata

MARC  
metadata

SPECTRUM  
metadata

← records

Fuente: <http://www.oaforum.org/tutorial>

# Especificaciones para proveedores de datos

- Las peticiones deben hacerse a través de los métodos **HTTP GET** y **POST** (los repositorios deben soportar ambos)

[http://ict.udlap.mx/Tales/Oai\\_tesis?verb=Identify](http://ict.udlap.mx/Tales/Oai_tesis?verb=Identify)

- Las respuestas deben generarse en **XML** bien formado:

```

- <OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2005-04-21T03:45:40Z</responseDate>
  <request verb="Identify">http://ict.udlap.mx:9090/Tales/Oai_tesis</request>
- <Identify>
  - <repositoryName>
    Universidad de las Americas, Puebla: Digital Theses
  </repositoryName>
  <baseURL>http://ict.udlap.mx:9090/Tales/Oai_tesis</baseURL>
  <protocolVersion>2.0</protocolVersion>
  <adminEmail>tesis@mail.udlap.mx</adminEmail>
  <earliestDatestamp>1989-05-04</earliestDatestamp>
  <deletedRecord>no</deletedRecord>
  <granularity>YYYY-MM-DD</granularity>
  <compression>identity</compression>
- <description>
  - <oai-identifier xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai-identifier
http://www.openarchives.org/OAI/2.0/oai-identifier.xsd">
    <scheme>oai</scheme>
    <repositoryIdentifier>thesisUDLAP</repositoryIdentifier>
    <delimiter>:</delimiter>
    <sampleIdentifier>oai:thesisUDLAP:24</sampleIdentifier>
  </oai-identifier>
  </description>
</Identify>
</OAI-PMH>

```

# GetRecord

**[http://ict.udlap.mx/Tales/Oai\\_tesis?](http://ict.udlap.mx/Tales/Oai_tesis?verb=GetRecord&identifier=oai:thesisUDLAP:116&metadataPrefix=oai_dc)**  
**verb=GetRecord&**  
**identifier=oai:thesisUDLAP:116&**  
**metadataPrefix=oai\_dc**

[http://ict.udlap.mx/Tales/Oai\\_tesis?verb=GetRecord&identifier=oai:thesisUDLAP:116&metadataPrefix=oai\\_dc](http://ict.udlap.mx/Tales/Oai_tesis?verb=GetRecord&identifier=oai:thesisUDLAP:116&metadataPrefix=oai_dc)

- <OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  - <responseDate>2005-04-21T04:33:48Z</responseDate>
  - <request verb="GetRecord" identifier="oai:thesisUDLAP:116" metadataPrefix="oai\_dc">http://ict.udlap.mx:9090/Tales/Oai\_tesis</request>
- <GetRecord>
  - <record>
    - <header>
      - <identifier>oai:thesisUDLAP:116</identifier>
      - <datestamp>2002-05-16</datestamp>
      - <setSpec>dis</setSpec>
    - </header>
    - <metadata>
      - <oai\_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai\_dc/http://www.openarchives.org/OAI/2.0/oai\_dc.xsd">
        - <dc:title>
          - Búsquedas por contenido y anotaciones en Bibliotecas Digitales multimediales
        - </dc:title>
        - <dc:creator>José Anibal Arias Aguilar</dc:creator>
        - <dc:subject>
          - Digital libraries, Information storage and retrieval systems, Virtual storage (Computer science), Neural networks (Computer science), Speech processing systems, Computer storage devices
        - </dc:subject>

• • •

• • •

<dc:publisher>Universidad de las Américas, Puebla</dc:publisher>

<dc:contributor>José Alfredo Sánchez Huitrón</dc:contributor>

<dc:date>2002-05-16</dc:date>

- <dc:type>

Maestria, Ciencias con Especialidad en Ingeniería en Sistemas Computacionales

</dc:type>

<dc:format>text/html</dc:format>

<dc:format>application/pdf</dc:format>

<dc:identifier>http://www.udlap.mx/~tesis/msp/arias\_a\_ja</dc:identifier>

<dc:source>>null</dc:source>

<dc:language>ES</dc:language>

<dc:relation>http://www.udlap.mx/~tesis/msp/arias\_a\_ja</dc:relation>

<dc:rights>http://www.udlap.mx/reglamte/capitulo2/t2c4.html</dc:rights>

</oai\_dc:dc>

</metadata>

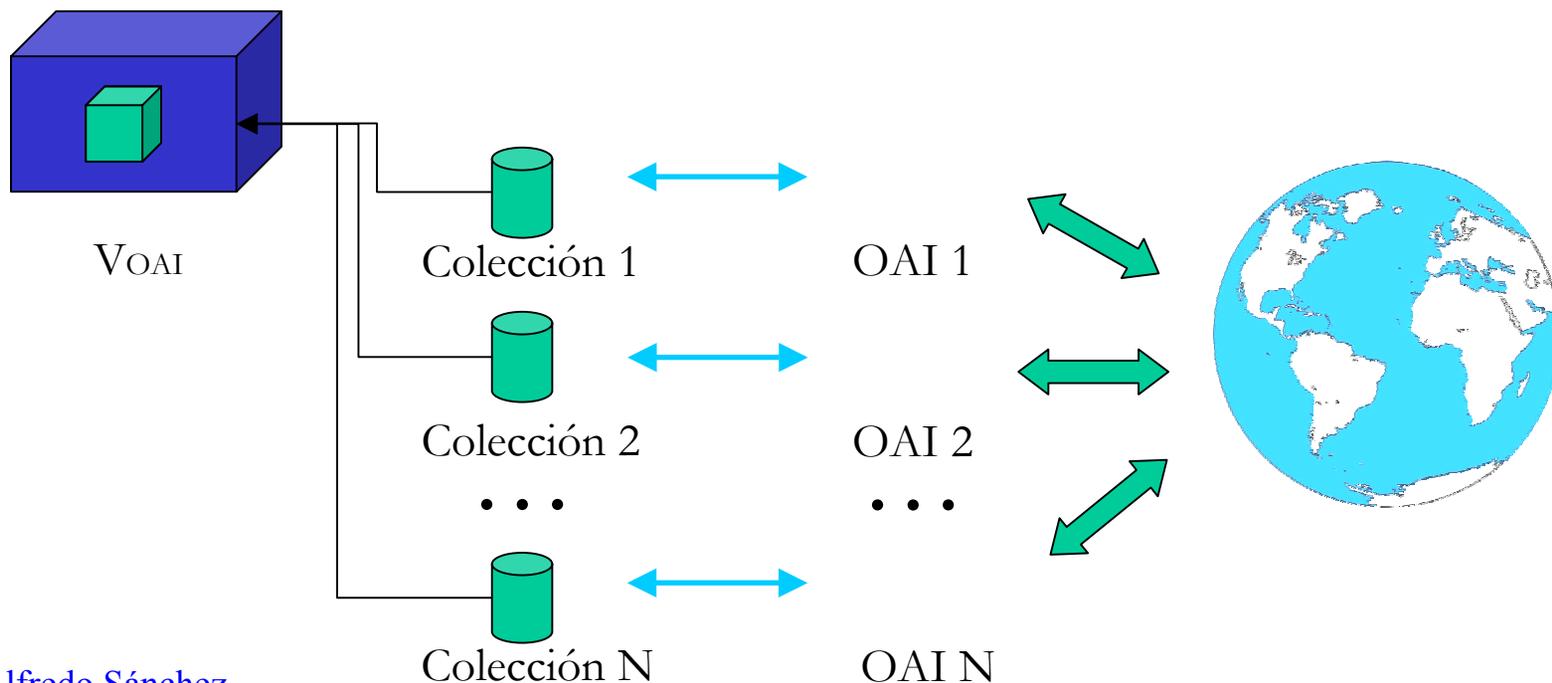
</record>

</GetRecord>

</OAI-PMH>

# ¿Qué es Voai?

- Es una herramienta que permite la construcción automática de un servidor OAI.
  - Es necesario especificar la forma de recuperar los metadatos de la colección.



## ¿Qué necesito para usarlo?

- Conocer la estructura de la base de datos.
- Lenguaje de consulta SQL.
- Tener nociones del protocolo OAI-PMH.
  
- Servidor de aplicaciones Tomcat
- Manejador de Base de Datos
- Driver JDBC

## Voai: los 3 pasos

- Para crear un servidor OAI es necesario seguir los siguientes pasos mostrados en la interfaz principal.
  - 1) Proporcionar información de la colección a diferentes módulos de especificación de Voai
  - 2) Seleccionar la ubicación donde se desea instalar el servidor generado
  - 3) Instalar servidor generado.

# Interfaz Paso 1

PASO 1 - Mozilla Firefox

Archivo Editar Ver Ir Marcadores Herramientas Ayuda

Customize Links Free Hotmail TOSHIBA Access Windows Media Windows

**voai**  
generador de servidores

**GetRecord:** Este verbo es usado para recuperar un registro específico de un repositorio.  
[Registrar GetRecord](#)

**Identify:** Este verbo es usado para recuperar información específica de un repositorio, por ejemplo, información administrativa y de identificación.  
[Modificar registro de Identify](#)

**ListIdentifiers:** Este verbo es usado para recuperar los identificadores de registros que fueron cosechados de un repositorio. Argumentos opcionales permiten hacer una recuperación selectiva.

**ListRecords:** Este verbo es usado para recuperar registros completos de un repositorio. Argumentos opcionales permiten hacer una recuperación selectiva.  
[Modificar registro de ListRecords y ListIdentifiers](#)

**ListMetadataFormats:** Este verbo es usado para recuperar los formatos de metadatos que soporta el repositorio.  
[Modificar registro de ListMetadataFormats](#)

**ListSets:** Este verbo es usado para recuperar la estructura de conjuntos en un que se organizan los registros de un repositorio.  
[Modificar registro de ListSets](#)

[Regresar al menú principal](#)

Listo 0

# ListMetadataFormats

- En este modulo se especifican las consultas para recuperar los metadatos para los estándares de metadatos.
  - Especificación para Dublin Core (formato requerido)
    - Se especifica el prefijo del formato de metadatos.
    - Se especifica la ubicación de esquema de validación de este formato de metadatos.
    - Se especifica la ubicación del nameSpace de esta formato de metadatos.
    - Para cada metadato de Dublin Core que la base de datos soporte, se especifica una consulta que me recupere el metadato de la base de datos al que corresponde el de Dublin Core.

# ListMetadataFormats (2)

Por ejemplo, si tenemos la siguiente tabla:

tesis		
idTesis	autorTesis	departamento
“is111”	“autor 111”	“sistemas”
“ad087”	“autor 087”	“administracion”

Para recuperar el metadato “Creator” de Dublin core se hace la siguiente consulta:

```
correspondencia: autorTesis  —————> Campo que permite  
select: *                    mapear metadato local a  
from: tesis                  metadato DC u otros  
where: idTesis=identifier
```

Gracias al campo correspondencia, el generador sabrá que el metadato “autorTesis” corresponde a “creator”.

Lo mismo se hace para los 15 metadatos de Dublín Core.

## ¿Qué es Xoai?

- Es una herramienta que permite la construcción automática de un servidor OAI a partir de metadatos almacenados en XML.

## ¿Qué necesito para usarlo?

- Manejo de documentos XML
- Consulta de documentos usando XSL.
- Tener nociones del protocolo OAI-PMH.
  
- Servidor de aplicaciones Tomcat
- Base de datos XML
- Manejador de JDBC

## Xoai: los 3 pasos

- Para crear un servidor OAI es necesario seguir los siguientes pasos mostrados en la interfaz principal.
  - 1) Registrar los metadatos en la BDXML
  - 2) Determinar hojas de estilo
  - 3) Determinar características del servicio

# Diseño e implementación de servidores OAI-XML

## Configuración – Generación de Headers

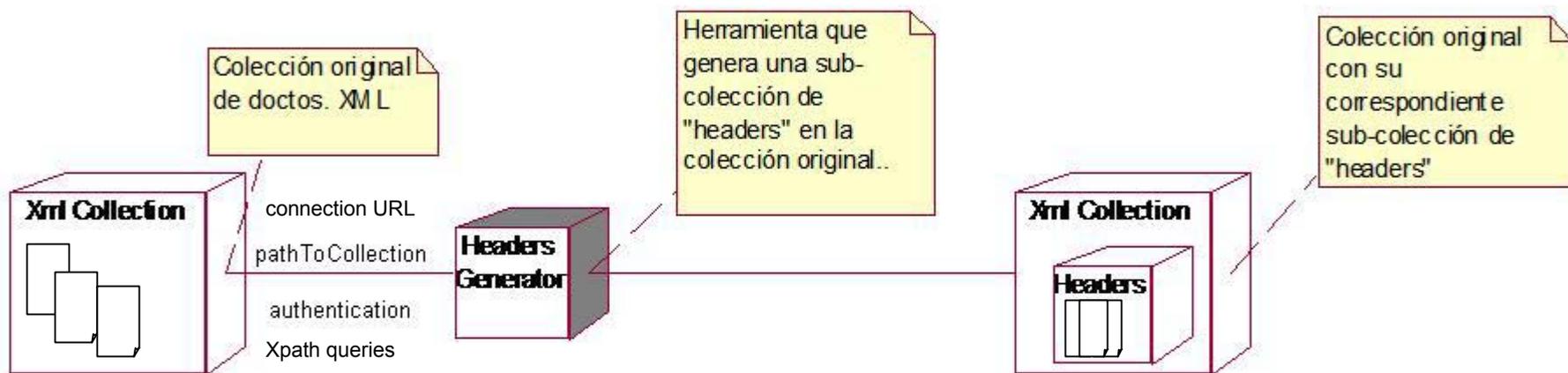


Diagrama de interacción de entidades durante el proceso de Generación de Headers

- Header document:**
- documento XML con una estructura predefinida
  - utiliza (principalmente) "tags" de Dublin Core
  - permite establecer una sección "extra" para propósitos específicos de cada colección.
  - a cada docto. en la col. Original, le corresponde un header

# Diseño e implementación de servidores OAI-XML

## Configuración

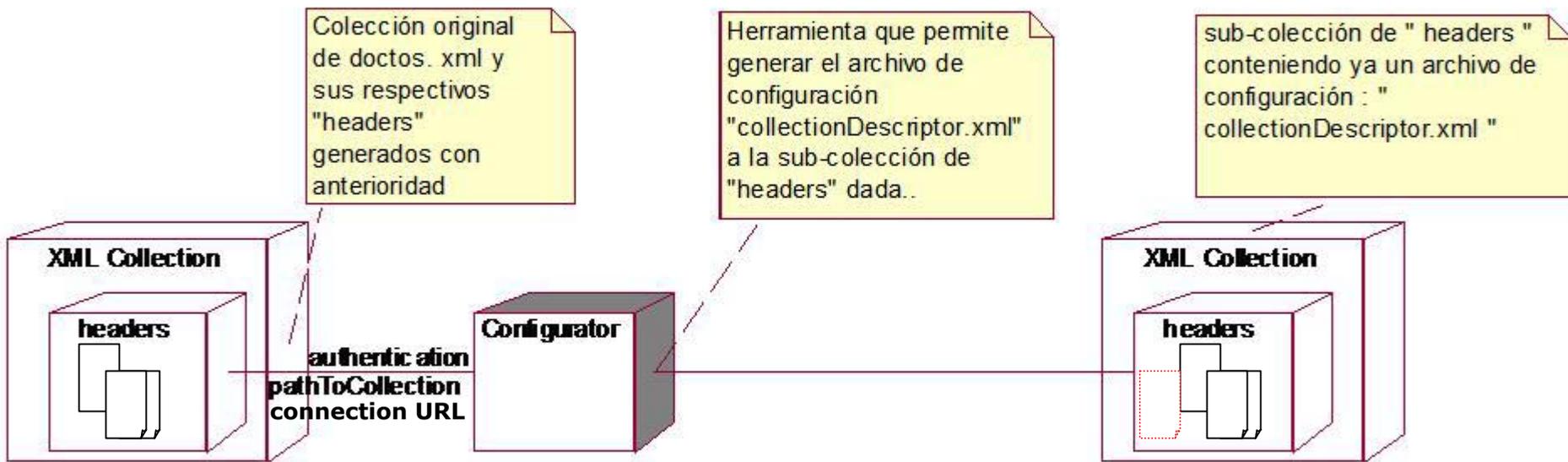


Diagrama de interacción de entidades durante el proceso de configuración del servidor OAI\_XML

Simbología:

 → " collectionDescriptor.xml "

# "Configurator"



The screenshot shows a Mozilla Firefox browser window titled "Configurator - Mozilla Firefox". The address bar displays the file path: `file:///C:/Archivos%20de%20programa/Apache`. The browser's toolbar includes navigation buttons (back, forward, home, stop, refresh) and utility icons (print, copy, paste, delete). Below the toolbar, there are three tabs: "Firefox Help", "Firefox Support", and "Plug-in FAQ".

A yellow sticky note is placed over the top-left portion of the page content, containing the text: "Check README.txt before configuration".

## OAI-XML Server Configuration

*(Información general acerca del Repositorio)*

**General Information**

Repository Name:

Protocol Version:

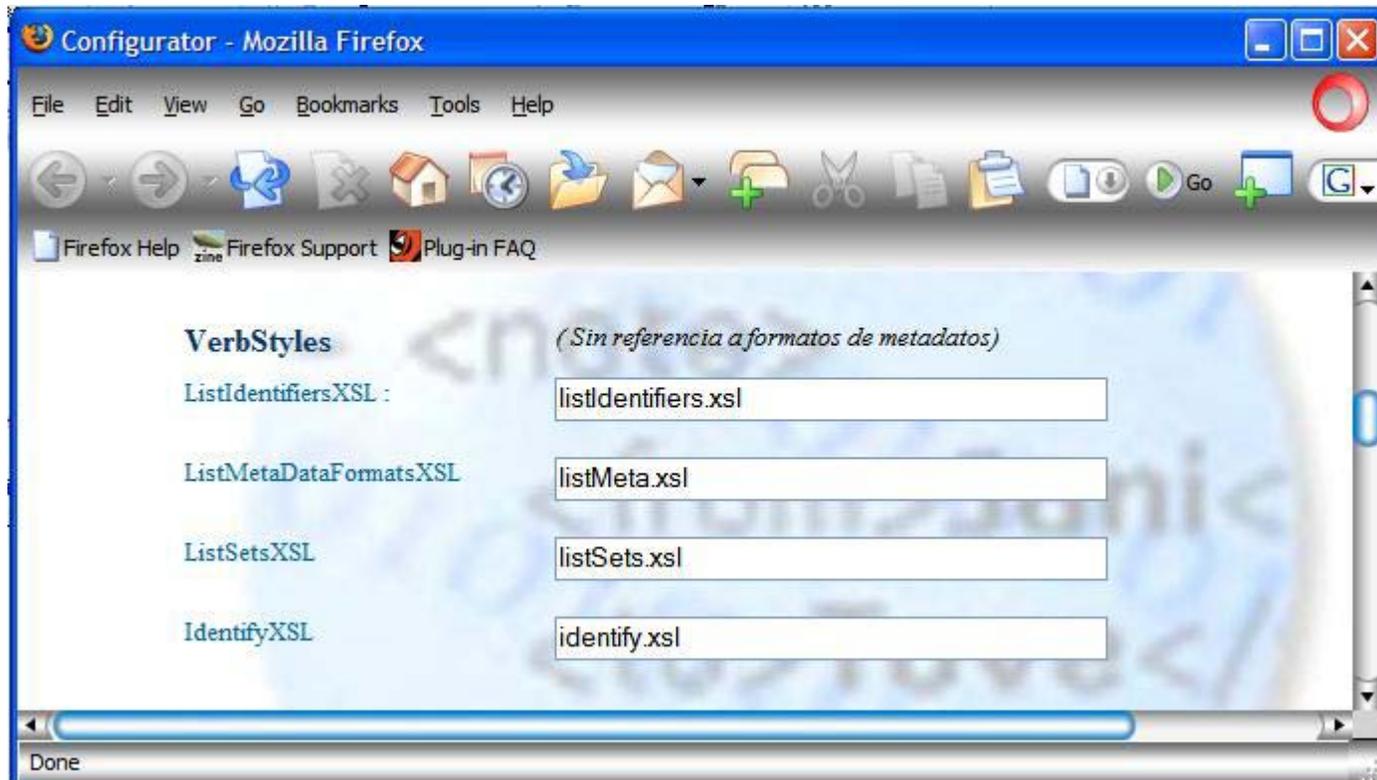
Admin-email:

Granularity:

Identifier Prefix:

Sample Identifier:

# "Configurator"



# Resumen

- Un enfoque común para facilitar la interoperabilidad es la definición de protocolos
- OAI desarrolla y promueve estándares de interoperabilidad para la difusión eficiente de contenido
- OAI-PMH define roles de proveedor de datos y proveedor de servicios
- Los proveedores de datos construyen servidores que atienden peticiones vía HTTP y producen respuestas en XML

# Referencias

- Corporación Universitaria para el desarrollo de Internet A.C. (CUDI)

<http://www.cudi.edu.mx>

Universidad de las Américas, Puebla

<http://www.udlap.mx>

Laboratorio de Tecnologías Interactivas y Cooperativas ICT

<http://ict.udlap.mx>

# Referencias

- Open Archives Initiative

<http://www.openarchives.org>

- Dublin Core

<http://www.dublincore.org>

- Servidor de Metadatos Tesis-UDLA

[http://ict.udlap.mx/Tales/Oai\\_tesis?verb=Identify](http://ict.udlap.mx/Tales/Oai_tesis?verb=Identify)

# Referencias

- Proyecto CUDI 2004 : Agentes y Movilidad en Colecciones Multimediales Heterogéneas

<http://ict.udlap.mx/projects/cudi/udlatec/>

- VOAI
- META-Explorador
- Valida-OAI