



Universidad Autónoma
del Estado de México

MANUAL TÉCNICO

SIEA

OCO oficina de
conocimiento
abierto



Contenido

INTRODUCCIÓN	3
REPOSITORIO INSTITUCIONAL	4
INTEROPERABILIDAD	8
OAI-PMH	9
DUBLIN CORE	11
CONFIGURACIÓN DEL OAI	12
DIRECTORIOS DE REPOSITORIOS INSTITUCIONALES Y TEMÁTICOS	13
GLOSARIO	14
REFERENCIAS	15



INTRODUCCIÓN

De acuerdo con lo establecido en el Mandato Institucional de Open Access, la UAEM implementa estrategias que dan cuenta de la congruencia entre su decir y su actuar, incorpora indicadores que promueven el uso de recursos tecnológicos para ampliar la visibilidad de la producción académica, científica, tecnológica, de innovación y cultural generada por la Institución.

Como todo proceso de innovación, requiere la inclusión de mecanismos para su instrumentación, resulta necesario precisar a la comunidad académica cual será el mecanismo implementación y validación de este ejercicio.

La Oficina de Conocimiento Abierto (OCA) con base en la normatividad vigente —Mandato de Acceso Abierto y el Acuerdo de Creación de la misma— propone el proceso para dar viabilidad a este ejercicio.

De esta manera se promueve el uso del Repositorio Institucional para dar cumplimiento a esta filosofía y pone a disposición de la comunidad universitaria el manual técnico como herramienta para comprender su estructura e interoperabilidad.



REPOSITORIO INSTITUCIONAL

La ciencia más útil es aquella cuyo fruto es el más comunicable. Leonardo Da Vinci.

El Repositorio Institucional es el sistema de información que reúne, preserva, divulga y da acceso a la producción académica, científica, tecnológica, de innovación y cultural que genera nuestra Institución.

Representado por el “Árbol de la Mora” el RI se nutre a través de las contribuciones de toda la comunidad universitaria. En 2011 se crea el RI, en tan solo cuatro años se ha posicionado en el sexto lugar a nivel nacional.

En el afán de promover su crecimiento, la Rectoría a través de la Oficina de Conocimiento Abierto, lo fortalece con la incorporación de indicadores gráficos de uso, una plataforma multilingüe, vía verde y vía dorada.

Con este escalamiento tecnológico la producción universitaria adquiere mayor visibilidad, puede comunicarse y viajar de manera más eficiente al mundo y sumarse al discurso global.

RIUAEMex de manera tecnológica cuenta con una estructura organizada por comunidades que reflejan la estructura de la Universidad —Rectoría, Secretarías, Institutos y Centros de Investigación, Facultades, Escuelas, Centros Universitarios, Unidades Académicas y Profesionales, Plateles de la Escuela Preparatoria, Centros Académicos y Culturales—.

Las Comunidades contienen colecciones, que son agrupaciones de contenidos relacionados, una colección puede aparecer en más de una comunidad. Cada colección se compone de un conjunto de ítems, que son uno o más archivos. Cada ítem es propiedad de una colección. Además, un ítem puede aparecer en colecciones adicionales; Sin embargo cada ítem pertenece a una y sólo una colección.

Los ítems que se subdividen en paquetes (bundles) con nombre de flujos de bits (bitstreams), que son conjuntos o arreglos de bits, que por lo general son archivos ordinarios, estos están de alguna manera estrechamente relacionados, por

ejemplo, archivos HTML e imágenes que componen un único documento HTML y se organizan en paquetes.



En la práctica, la mayoría de los ítems tienden a tener estos paquetes con nombre:

- *ORIGINAL*, paquetes con bitstreams, depositados originalmente.
- *THUMBNAILS*, miniaturas de bitstreams de imagen
- *TEXT*, Texto completo del bitstreams del original, para la indexación
- *LICENSE* – contiene la licencia de depósito que el remitente haya concedido la organización de acogida; en otras palabras, especifica los derechos que la Universidad.
- *CC_LICENSE* Contiene la licencia de distribución, en su caso (una licencia de Creative Commons) asociado con el ítem. Esta licencia específica a los usuarios finales la descarga del contenido y lo que pueden hacer con él.

Cada bitstream se asocia con un formato Bitstream, es importante capturar los formatos de archivo específicos que los usuarios presenten. Un formato bitstream es una manera única y consistente para referirse a un formato de archivo en particular. Una parte integral de un formato de bitstream es una noción ya sea implícita o explícita de cómo el material en ese formato se puede interpretar. Por ejemplo, la interpretación de los bitstream codificados en el estándar JPEG para la compresión de imágenes fijas se define explícitamente en la Norma ISO/IEC 10918-1.

La interpretación de los bitstream en formato Microsoft Word 2000 se define implícitamente, a través de la referencia a la aplicación de Microsoft Word 2000. Los formatos Bitstream pueden ser más específicos que los tipos MIME o sufijos de archivo. Por ejemplo, application/ms-word y .doc abarcan múltiples versiones de la aplicación Microsoft Word, cada uno de los cuales produce bitstream con características presumiblemente diferentes.

Cada formato de flujo de bits, además, tiene un nivel de soporte, lo que indica que tan probable es capaz de preservar el contenido en el formato, en el futuro dentro de la Universidad. Hay tres posibles niveles de soporte de formatos Bitstream, estos pueden ser asignados por la Universidad, ya que ésta debe determinar el significado exacto de cada nivel de soporte, después de una consideración de los costos y requisitos. Recomendaciones:

Soportado	El formato es reconocido, y la institución anfitriona confía en que puede hacer que los Bitstream de este formato sean utilizables en el futuro, usando cualquier combinación de técnicas (como la migración, la emulación, etc.) es apropiada dado el contexto de necesidad.
Desconocido	El formato es reconocido, y la institución anfitriona prometerá para preservar el Bitstream tal y como son, y que pueda ser recuperada. La institución anfitriona intentará obtener suficiente información para que el formato a ser actualizado a nivel 'apoyado'.
No soportado	El formato es desconocido, pero la institución anfitriona se compromete a preservar el Bitstream como está y que pueda ser recuperada.

Tabla 1. Recomendaciones para la preservación de formatos.

Cada ítem tiene un cualificado registro de metadatos Dublin Core. Otros metadatos pueden ser almacenados en un ítem como un bitstreams en serie, pero almacenan Dublin Core para cada ítem para la interoperabilidad y facilidad de búsqueda. Dublin Core puede ser introducido por los usuarios finales, ya que presentan contenido, o puede derivarse de otros metadatos como parte de un proceso.

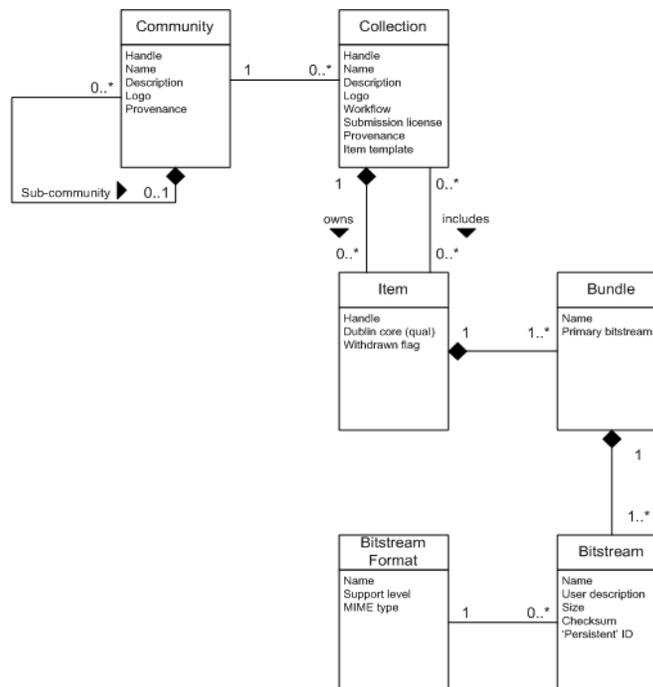




Figura 1. Diagrama de la estructura utilizada por el RIUAEMex.

Los ítems que se pueden eliminar de RIUAEMex son de dos formas: Pueden ser 'retirados', lo que significa que permanece el archivo, pero están completamente ocultos a la vista. En este caso, si un usuario final intenta acceder al elemento retirado, se presenta una etiqueta, que indica que el ítem ha sido eliminado. Por alguna razón, un ítem también puede ser 'borrado' de ser necesario, en cuyo caso se eliminan todo rastro del archivo.

Objeto	Ejemplo
Comunidades	Secretarías; Centros Académicos
Colecciones	Dominio Público; Tesis
Ítem	Artículo Científico; un conjunto de datos con el acompañamiento de la descripción;
Bundle	Un grupo de HTML y de imagen flujos de bits que componen un documento HTML
Bitstream	Un único archivo HTML; un solo archivo de imagen; un archivo de código fuente
Bitstream Format	Microsoft Word versión 6.0; JPEG Formato de imagen codificada

Tabla 2. Objetos que componen la estructura del RIUAEMex



INTEROPERABILIDAD

La normalización ha jugado un papel primordial en la creación de repositorios en acceso abierto. Se hacía necesario un estándar para poder intercambiar la información digital contenida en los diferentes archivos abiertos. El objetivo era partir de un mismo esquema de descripción de la información para poder intercambiar los datos. Esto se consiguió con el protocolo OAI-PMH, desarrollado por la Open Archives Initiative y cumplido por las aplicaciones informáticas diseñadas para la puesta en marcha de repositorios y colecciones digitales. Los repositorios que emplean OAI-PMH posibilitan que sus archivos abiertos sean recuperados a partir de recolectores (harvesters), los cuales actúan a modo de meta-buscadores.

Estos metabuscadores juegan un papel esencial en la visibilidad de presentación/publicación de nuestro conocimiento hacia el mundo. En este sentido es esencial contar con un software que garantice esta funcionalidad. Algunos de los meta-buscadores más usados son: Google Académico, Google Scholar, HISPANA, DIALNET, RECOLECTA, DRIVER, OAISTER, SCIENTIFIC COMMONS. 3.3.1.

Uno de los elementos primordiales para llevar a cabo la interoperabilidad es un correcto uso y definición de metadatos, esto nos ayudara a la mejor localización de datos dentro o fuera de la organización, la correcta transferencia e interpretación de los datos, conocer información que es clave y una mejor comprensión de los datos.

Los estándares con los que el repositorio institucional puede operar son los siguientes Dublin Core, Darwin Core e ID3.

Las plataformas con las que Dspace es interoperable son Dspace, Eprints, Fedora, CDSware, Bepress, OPUS, HTML, Wildfire, ETD-db todas estas plataformas tienen en común que cumplen con el protocolo OAI-PMH lo que permite la interoperabilidad entre todos los sistemas que utilicen alguna de las plataformas listadas anteriormente.

Características de interoperabilidad del RI

- El RI maneja el esquema de codificación de caracteres UNICODE UTF-8, compatible con otros sistemas.
- Estándares de metadatos, Dublin Core(DC).
- Licenciamiento de contenidos
- Tipos de documentos



OAI-PMH

OAI-PMH es un protocolo desarrollado por Open Archives Initiative. Es usado para recolectar (harvest) descripción de metadatos en un archivo usando metadatos de otros archivos. OAI-PMH utiliza XML sobre HTTP.

Esquema de recolección de metadatos de OAI-PMH. OAI-PMH ofrece a los proveedores de datos una solución técnica sencilla que posibilita mantener disponible sus metadatos para los servicios que los recolectan, basados en el estándar HTTP y XML. De esta manera los metadatos de muchas fuentes pueden ser reunidos en una base de datos y ofrecidos a los proveedores de servicios para ser consumidos desde ese repositorio centralizado. En la figura 2 muestra el esquema de recolección de metadatos.

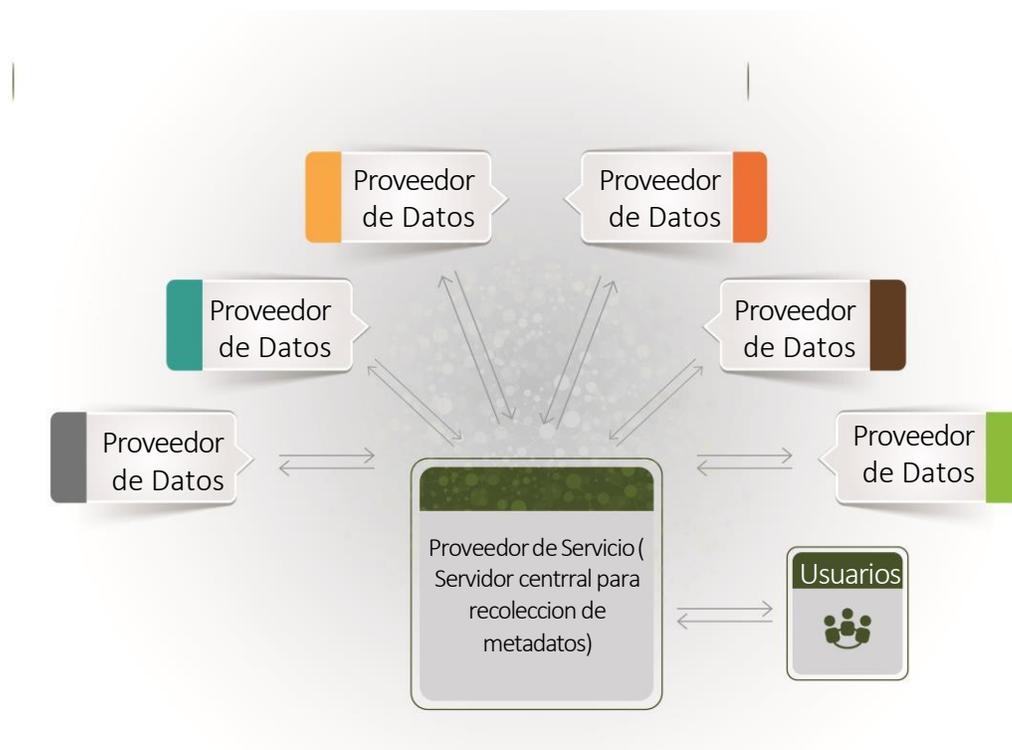


Figura 2 Esquema de recolección de metadatos.

OAI-PMH es un mecanismo de bajo nivel para interoperabilidad de repositorios. Los proveedores de datos son repositorios que exponen su estructura de metadatos vía OAI-PMH. Entonces los proveedores de servicios hacen llamadas (request) al servicio OAI-PMH para recolectar los metadatos. OAI-PMH utiliza un conjunto de seis verbos (o servicios) que son invocados desde HTTP mediante operaciones de (request/replies). En la figura 3 se muestra dicho proceso.

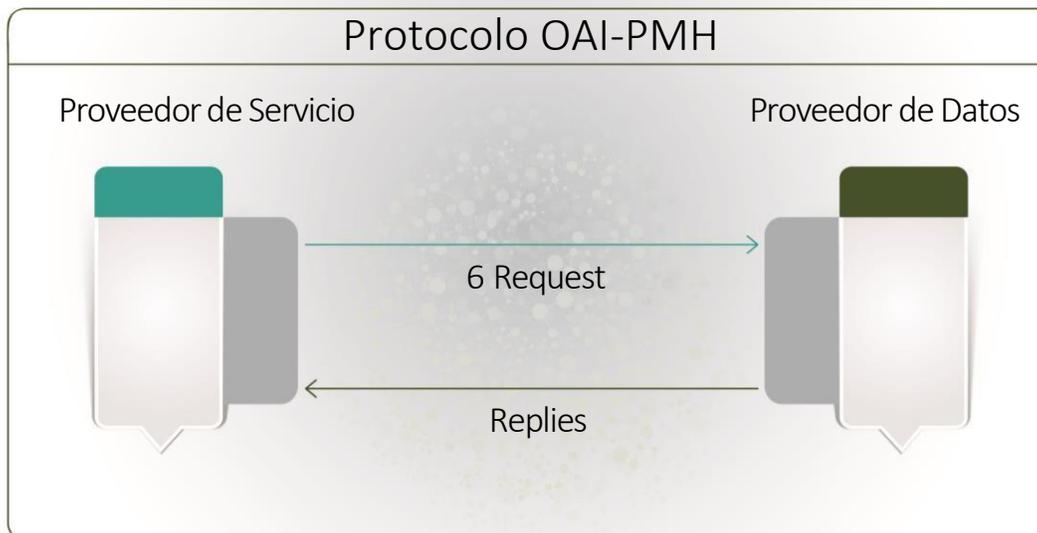


Figura 3 Proceso de recolección de metadatos

OAI-PMH está basado en los estándares comunes de XML y DC, y también es compatible con MARC. Sin embargo, DC ha sido adoptado como el formato de metadatos común que debería ser soportado por todos los proveedores de datos. Una de las concepciones erróneas es considerar que la recolección se realiza a nivel de contenido (documentos, ficheros, imágenes). Realmente, OIAPMH es un protocolo de intercambio de metadatos, por lo tanto, los recolectores incorporan referencias a la localización original de los contenidos (de ahí la importancia de crear/mantener identificadores persistentes).

Esquema de comunicación (request/reply) de OAI-PMH. Actores clave de OAI-PMH Se distinguen dos clases de participantes en el protocolo OAI-PMH:

- Proveedores de datos (Data providers): mantienen uno o más repositorios (webservers) que soportan OAI-PMH como forma de exponer los metadatos. Estos proveedores procesan y responden las solicitudes recibidas desde los proveedores de servicios (service providers) con mensajes de respuesta OAI-PMH (OAI-PMH response).
- Proveedores de servicio (Service providers): Un proveedor de servicios envía solicitudes OAI-PMH (OAI-PMH request) a los proveedores de datos y utiliza los metadatos obtenidos como base para la creación de servicios de valor añadido. Son los recolectores de metadatos que los utilizan con el propósito de ofrecer un servicio de interfaz de búsqueda, sistemas de corrección, etc.. La figura 3.3 muestra la interacción de entre los diferentes tipos de proveedores.



DUBLIN CORE

Dublin Core (DC) es el formato de metadatos por defecto utilizado por Dspace y por consecuencia por el RI. DC es un modelo de metadatos que ha sido creado por Dublin Core Metadata Initiative (DCMI). Las implementaciones de DC usan generalmente XML y se basa en el Resource Description Framework. Se define por: ISO en su norma 15836 de febrero del 2009, estándar ANSI/NISO Z39.85-2007 de mayo del 2007 e IETF RFC 5013 de agosto del 2007. DC define 15 definiciones semánticas básicas, las cuales pretenden proporcionar información descriptiva básica sobre cualquier recurso. Dichas definiciones son opcionales, pueden repetirse y aparecer en cualquier orden (2).

Etiqueta	Descripción
Contributor	Entidad responsable de realizar contribuciones al recurso
Coverage	Espacio o tema temporal del recurso, la aplicabilidad espacial del recurso, o la jurisdicción en la cual el recurso es relevante
Creator	La entidad responsable en primera instancia de crear el recurso
Date	Un punto o período de tiempo asociado con un evento en el ciclo de la vida del recurso
Description	Una descripción del recurso
Format	Formato de archivo, medio físico, o dimensiones del recurso
Identifier	Una referencia no ambigua del recurso en un contexto dado
Language	El lenguaje del recurso
Publisher	La entidad responsable de hacer disponible el recurso
Relation	Un recurso relacionado
Rights	Información sobre los derechos del recurso
Source	Un recurso relacionado del cual se deriva el recurso descripto
Subject	El tópico del recurso
Title	El nombre dado al recurso
Type	La naturaleza o género del recurso

DSpace cumple con muchos protocolos estándar para el acceso, la ingesta, y la exportación. Los soportes estándares DSpace incluyen: OAI-PMH, OAI-ORE, ESPADA, WebDAV, OpenSearch, OpenURL, RSS, ATOM.



CONFIGURACIÓN DEL OAI

Para acceder a servidor OAI se puede hacer desde la página principal o bien desde la liga <http://ri.uaemex.mx/oai>

Para poder utilizar el protocolo OAI siga las siguientes Instrucciones:

Cada vez que se vuelve a migrar DSpace el protocolo OAI-PMH, la interfaz se reconstruye automáticamente. Existen "verbos" (es decir, comandos) que se pueden utilizar de forma en línea a través del navegador.

- Identify, se utiliza para recuperar información acerca de un repositorio:
<http://ri.uaemex.mx/oai/request?verb=Identify>
- ListMetadataFormats, listas de formatos de metadatos soportadas por OAI-PMH: <http://ri.uaemex.mx/oai/request?verb=ListMetadataFormats>
- ListRecords, permite metadatos recolección de la totalidad o una gran variedad de productos:
http://ri.uaemex.mx/oai/request?verb=ListRecords&metadataPrefix=oai_dc
- ListSets - enumera los "sets" (es decir DSpace Collections) disponibles en un repositorio: <http://ri.uaemex.mx/oai/request?verb=ListSets>

Este ejemplo muestra la respuesta a una solicitud ListIdentifiers que especifica un rango de fechas y un proveedor de datos.

http://ri.uaemex.mx/oai/request?verb=ListIdentifiers&from=2015-11-06&until=2015-11-12&metadataPrefix=oai_dc



```
<OAI-
PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.
w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
<responseDate>2015-11-13T00:49:19Z</responseDate>
<request verb="ListIdentifiers" from="2015-11-06" until="2015-11- 12"
metadataPrefix="oai_dc">http://ri.uaemex.mx/oai/request</request>
<ListIdentifiers>
<header>
<identifier>oai:ri.uaemex.mx:123456789/21932</identifier>
<timestamp>2015-11-09T22:01:53Z</timestamp>
<setSpec>hdl_123456789_21380</setSpec>
</header>
<header>
<identifier>oai:ri.uaemex.mx:123456789/21929</identifier>
<timestamp>2015-11-09T21:19:55Z</timestamp>
<setSpec>hdl_123456789_21380</setSpec>
</header>
<header>
<identifier>oai:ri.uaemex.mx:123456789/21930</identifier>
<timestamp>2015-11-09T21:38:22Z</timestamp>
<setSpec>hdl_123456789_21380</setSpec>
</header>
<header>
<identifier>oai:ri.uaemex.mx:123456789/21933</identifier>
<timestamp>2015-11-09T22:21:36Z</timestamp>
<setSpec>hdl_123456789_21380</setSpec>
</header>
</ListIdentifiers>
</OAI-PMH>
```

DIRECTORIOS DE REPOSITORIOS INSTITUCIONALES Y TEMÁTICOS

Existen directorios internacionales de repositorios y se recomienda, una vez creado el repositorio institucional, registrarse en uno de ellos para que se constate su existencia y se difunda. Los más relevantes y/o utilizados son:

- ROAR
- OpenDOAR
- Buscarepositorios



GLOSARIO

Repositorio Institucional Es una Base de datos de resultados de investigación digitales. Pueden tener base institucional o ser un servicio a una comunidad en particular disciplinaria geográfica o de otro tipo.

Interoperatividad: Capacidad que tienen las organizaciones dispares y diversas para intercambiar, transferir y utilizar, de manera uniforme y eficiente datos, información y documentos por medios electrónicos, entre sus sistemas de información.

OAI-PMH: Protocolo de archivos abiertos que da soporte a la difusión de múltiples formatos de metadatos de un repositorio.

Dublín Core: Es un estándar de metadatos que ubica, dentro de internet, los datos para describir, identificar, procesar, encontrar y recuperar un documento introducido en la red.

Dspace: Es un software de código abierto que promueve herramientas para la administración de repositorios institucionales.

Metadato: Datos que describen otros datos.

Meta-buscadores: Websites(páginas web) especializados en consultar varios buscadores al mismo tiempo y presentar la información proporcionada de forma entendible y ordenada.

Recolectores (harvesters): Es un servidor externo a los repositorios que permite que el usuario busque, localice y acceda a través de un único interface y de forma simultánea a la información y los materiales depositados en varios de ellos.

OpenDOAR: Es un directorio de repositorios académicos y de investigación que fue desarrollado por la Universidad de Nottingham, del Reino Unido.

ROAR: Es un directorio de repositorios en acceso abierto de todo el mundo que, además de facilitar su localización y acceso, proporciona información sobre el crecimiento y la actividad de cada uno de ellos.

REFERENCIAS

Federico Zubiri y Andrés Mereles. (2012). Repositorios Institucionales. Consultado el 13 de Noviembre de 2015. En:

<http://www.fing.edu.uy/inco/grupos/pln/prygrado/InformePGRepositorios.pdf#section.3.2>

Gobierno Bolivariano de Venezuela, Ministerio el poder popular para la educación Universitaria, Ciencia y Tecnología.

Directrices para políticas y desarrollo del acceso abierto, unesco 2013 alma swan, parís Francia

SEDIC (2006). Acceso Abierto y Repositorios de Documentos. Consultado el 13 de Noviembre de 2015. En: http://www.sedic.es/autoformacion/acceso_abierto/5-repositorios-OAI.html

Open Archive Forum. (2003). Principales conceptos técnicos sobre OAI-PMH. Consultado el 12 de Noviembre de 2015. En: <http://travesia.mcu.es/portaln/jspui/html/10421/1823/page3.htm>

UCLM Biblioteca Universitaria.(). Servicio de Apoyo a la Investigación. Consultado el 12 de Noviembre de 2015. En:

<http://biblioteca.uclm.es/Archivos/Investigacion/Recolectores.pdf>