

INTRODUCCIÓN A OAI-PMH Y SU IMPLANTACIÓN EN EL PORTAL E-REVISTAS

JULIO PEMAU ALONSO (jpmau@cti.csic.es)
JOSE ANGEL BARROSO CORROTO (jbarroso@cti.csic.es)

INDICE:

1. PRESENTACIÓN
 - 1.1. INTRODUCCIÓN
 - 1.2. OBJETIVOS DEL PROYECTO.
2. OPEN ACCESS
3. ARQUITECTURA
 - 3.1. REQUISITOS DEL SISTEMA
 - 3.2. ARQUITECTURA LÓGICA GENERAL
 - 3.3. DATA PROVIDER
 - 3.4. SERVICE PROVIDER
4. PROTOCOLO OAI-PMH
5. DUBLIN CORE
6. SOFTWARE SERVICE PROVIDER
7. SOFTWARE DATA PROVIDER
8. BIBLIOGRAFÍA Y ENLACES.

1. PRESENTACIÓN

1.1 INTRODUCCIÓN

Tecnociencia: <http://www.tecnociencia.es>

Es un **Punto de Encuentro** entre el mundo de la investigación y el de la empresa que facilita y promueve el intercambio de conocimientos científicos y tecnológicos de forma rápida y eficiente.

Su objetivo es favorecer la cooperación entre los diversos agentes del sistema de ciencia y tecnología, ayudando a que los resultados de los proyectos de I+D lleguen al sector empresarial y contribuyan a la mejora de sus procesos productivos.

Tecnociencia es un **servicio de utilidad pública** y carácter global, con un ámbito de actuación nacional y multisectorial que complementa y apoya los intereses específicos de iniciativas particulares, impulsando el éxito de las mismas como parte de sus objetivos.

E-revistas: <http://www.tecnociencia.es/e-revistas>

Plataforma digital dentro del Portal Tecnociencia donde se recogen, seleccionan y alojan las revistas electrónicas españolas o latinoamericanas existentes o de nueva creación que cumplan una serie de requisitos de calidad rigurosos.

El acceso a las revistas será abierto sin que ello menoscabe los principios de calidad que deben regir los textos de las publicaciones integradas en el proyecto.

1.2 OBJETIVOS DEL PROYECTO.

Las revistas científicas constituyen el principal canal de comunicación entre los científicos y tecnólogos. Este tipo de información es uno de los más solicitados por los usuarios en el Portal Tecnociencia, y por tanto, supondría un incremento considerable en cuanto al número de visitas, aumentaría la fidelización de los usuarios y en definitiva se optimizaría la comunicación y difusión científica y tecnológica a todos los sectores de la sociedad, objetivo principal de dicho portal.

De forma explícita los objetivos se resumirían en:

- Establecer un sistema de calidad normalizado y validado de las revistas electrónicas. Aquellas revistas que estén interesadas en seguir estos criterios de calidad, certificarán que sus publicaciones cumplen con las normas anteriores, convirtiéndose en un marchamo de calidad para las publicaciones.
- Crear una plataforma que aglutine a todas las revistas electrónicas que quieran formar parte de la misma y cumplan los requisitos de calidad establecidos. Esta plataforma residirá dentro del portal **Tecnociencia**, y permitirá el acceso abierto a todos los usuarios a las revistas. El acceso a las revistas se realizará a través de bases de datos y por directorios. Tal y como está diseñada esta plataforma, permitirá además de la consulta a las revistas, constituir un repositorio de todas las revistas que quieran formar parte del portal.
- Crear una interfaz al protocolo **OAI-PMH** (*Open Archives Initiative – Protocol for Metadata Harvesting*) que permita que los contenidos del portal de revistas estén disponibles en múltiples de bases de datos y catálogos internacionales.
- Aportar servicios de valor añadido sobre los documentos indicados. Por ejemplo: sistema de alertas, acceso personalizado a las colecciones, estadística de uso de los documentos y estudio de medidas de impacto de los mismos.
- Difundir el movimiento de acceso abierto entre la comunidad científica nacional.
- Asesorar técnicamente para la creación de nuevas revistas electrónicas de acceso abierto.
- Proporcionar servicio opcional de hospedaje de revistas electrónicas de libre acceso.
- Difundir ampliamente del proyecto entre instituciones académicas para su conocimiento y uso de los recursos generados.

2. OPEN ACCESS

La investigación científica, principalmente financiada por entidades públicas, tiene como principal vía de difusión la publicación de artículos científicos en revistas especializadas. La aparición y crecimiento simultáneo de la red de redes, Internet, ha provocado un cambio revolucionario en las publicaciones y en general en la difusión del conocimiento. Ya no hay una única vía casi lineal entre producción de información y difusión-distribución, propia de la era impresa.

Las relaciones a través de Internet pueden ser lineales, radiales, neuronales etc., tantas como se deseen a partir de un nodo en la red. Sin embargo, a pesar de que las facilidades tecnológicas han aumentado inmensurablemente, las restricciones dictadas por las editoriales de acceso a las publicaciones científicas cada vez son mayores.

Se producen situaciones como el hecho de que los científicos produzcan y ofrezcan gratuitamente los artículos que nutren a las revistas para que después no tengan acceso libre a sus propios contenidos.

O por ejemplo, la imposibilidad de crear bases de datos sin violar leyes de copyright que ensamblen información de distinta procedencia para ser utilizadas en aplicaciones de búsqueda.

Todas estas circunstancias han provocado en los últimos años una reacción en cadena de científicos e instituciones denunciando la falta de accesibilidad a sus propias publicaciones debido a las barreras económicas impuestas por los monopolios editoriales. Como consecuencia, la idea o movimiento del libre acceso (***open access***) a las publicaciones electrónicas cada vez genera más núcleos en torno al apoyo tanto institucional como económico de las publicaciones de libre acceso a través de Internet.

La Budapest Open Access Initiative (BOAI, 2002; versión oficial en español) define open access como "disponibilidad gratuita en la Internet pública, para que cualquier usuario la pueda leer, descargar, copiar, distribuir, imprimir, con la posibilidad de buscar o enlazar todos los textos de estos artículos, recorrerlos para indexación exhaustiva, usarlos como datos para software, o utilizarlos para cualquiera otro propósito legal, sin barreras financieras, legales o técnicas, distintas de la fundamental de ganar acceso a la propia Internet".

Como podemos ver el concepto de *open access* no sólo tiene que ver con la accesibilidad a la documentación científica sino con la idea de eliminar la obligatoriedad de cesión del copyright de los artículos

publicados, lo que facilita a los autores su inserción posterior, sin barreras, en archivos o repositorios temáticos de difusión internacional. A partir de este concepto cada vez está más extendido, han surgido numerosas iniciativas con la premisa de facilitar el acceso a las publicaciones científicas por medios electrónicos.

Caben destacar:

- *BioMed Central* (BMC, 2001) editorial a través de internet con publicaciones que se ajustan a los supuestos de *open access*.
- *Budapest Open Access Initiative* (BOAI, 2002) facilita ayuda y promueve proyectos encaminados a liberar el acceso electrónico a artículos de investigación, de todas las áreas, bien a través de su archivo en un repositorio o mediante su publicación en revistas de libre acceso, es una actividad de la Fundación Soros.
- *The Public Library of Science* (PLOS, 2000) es una organización para el apoyo y la creación de revistas de libre acceso.
- *The Scholarly Publishing and Academic Resources Coalition* (SPARC, 1998), coalición de universidades y bibliotecas para difundir y contribuir a la iniciativa de libre acceso a las revistas académicas a través de Internet.
- *Scientific Electronic Library Online* (SciELO, 2002) es un proyecto de países latinoamericanos para facilitar la consulta y visibilidad de sus publicaciones en Internet, sirve de repositorio de sus artículos para los que se han desarrollado protocolos y software propios para la evaluación y difusión de sus publicaciones.
- *Directory of Open Access Journals* (DOAJ, 2003) proyecto internacional liderado por la Universidad de Lund en el que se han recogido más de 700 publicaciones de libre acceso de diferentes procedencias.

En octubre de 2003, se hizo pública la Declaración de Berlín (*Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*) suscrita por diferentes representantes y científicos de países europeos. En ella, explícitamente se manifiesta las grandes posibilidades que brinda Internet y su papel en la difusión del conocimiento.

La declaración asume el espíritu de la declaración del *Budapest Open Access Institute* (BOAI), de la *European Cultural Heritage Online* (ECHO, 2003) y de la declaración de *Bethesda* (2003). Además de las

vías clásicas de difusión, avala el paradigma de *open access* a través de Internet.

Tal y como recoge la Declaración, las instituciones firmantes se comprometen a respaldar este paradigma y fomentar entre los investigadores el uso de revistas que contemplen este concepto. Se comprometen también a desarrollar nuevos métodos de evaluación y reconocimiento de los méritos académicos de los científicos que publiquen en ellas.

Recientemente (OCDE, 28-29 enero 2004) la OECD ha publicado también un documento en el que instan a los países firmantes a promover el libre acceso a la documentación científica generada de la investigación financiada con fondos públicos.

Todas estas iniciativas internacionales, entre otras, han surgido con un único objetivo común el de mejorar el sistema tradicional de comunicación científica y facilitar el acceso a las publicaciones científicas. Tal mejora puede concretarse en tres puntos:

- Aumentar la rapidez de distribución de los trabajos por medios electrónicos.
- Reducir los costes de las publicaciones científicas, y aumentar la visibilidad de los trabajos publicados.
- Reducir el fenómeno de "ciencia perdida", es decir aquella que no llega a tener un impacto relevante por su ausencia en las grandes bases de datos y repertorios científicos.

Estos objetivos pueden lograrse a través de diferentes proyectos que podrían resumirse en tres categorías:

- 1.** Creación de repositorios de documentos.
- 2.** Creación de revistas electrónicas de libre acceso.
- 3.** Fomentar y propugnar el acceso libre a las publicaciones científicas mediante políticas de apoyo a este tipo de proyectos.

Destacar que el acceso a la información se extiende a toda la ciudadanía, legitimando así la viabilidad de la participación en ciencia de los llamados expert-citizen que emergen de entre los colectivos de afectados (como los enfermos de SIDA), los activistas de diversa procedencia (del tipo ecologistas) y los partidarios del software libre.

Existe una clara relación entre la filosofía open access comentada y los problemas de licencias que existen en el mundo del software libre (GPL, copyleft, creative commons...). La tendencia errónea a confundir gratuidad con open access origina debates de viabilidad cuando se plantea poner en práctica estas licencias.

Muchos científicos opinan que el open access multiplica la visibilidad de sus papers y es mucho más cómodo bajarse la información de Internet que acudir a una biblioteca. No obstante aún existen muchas barreras: intereses económicos e incluso opiniones de pérdida de prestigio a la hora de adoptar estas ideas.

Son muchas las posibilidades de implantación. Desde las que se lanzan por completo como la plataforma BioMed Central que opta por liberar todo su contenido, hasta otras que liberan sus artículos pasados seis meses de su publicación e incluso algunas que regalan sólo la versión electrónica y cobran por la de papel.

Por último comentar que una de las opciones más importantes es el **auto-archivo**. Los propios autores depositan sus documentos en un repositorio de archivos electrónicos abiertos (normalmente que cumplen los requisitos OAI) mediante un simple interfaz web donde especifican los metadatos del documento (autor, fecha, título, etc...) y adjuntan el texto completo del paper.

El propio autor es el que solicita a la editorial la inclusión del texto completo pudiendo corregir los datos necesarios en caso de no conseguir el permiso.

Se prevé que multitud de instituciones, y muy especialmente las asociaciones y academias de carácter profesional, se agolparán por tener una iniciativa que facilite y estimule la práctica del auto-archivo.

3. ARQUITECTURA

Se propone una arquitectura abierta, basada en tecnologías estandarizadas y de amplia aceptación en el ámbito de desarrollo de esta iniciativa. Queda fuera de nuestros objetivos elementos que tengan que ver con la explotación comercial de las publicaciones y por tanto esta arquitectura no incluirá elementos relacionados con estos aspectos.

La propuesta que se expone contempla la infraestructura tecnológica sobre la que se apoyarán las aplicaciones que componen este proyecto. Dentro de este ámbito se encuentran: plataformas hardware, sistemas operativos, bases de datos, servidores y subsistemas especializados.

3.1 REQUISITOS DE SISTEMA

La arquitectura que se presenta deberá dar respuesta a las necesidades tecnológicas derivadas de los siguientes requisitos a satisfacer por el sistema a desarrollar:

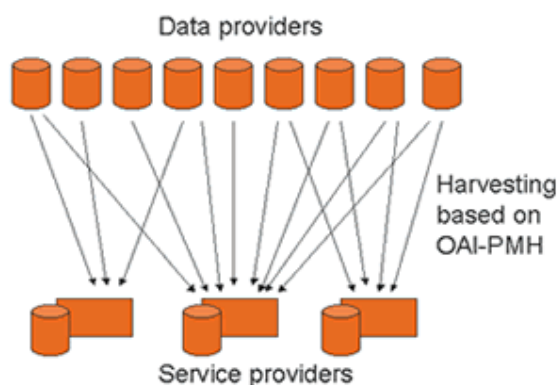
- Integración en los sistemas de archivos abiertos (*Open Archives*) existentes en el mundo y reconocidos por la comunidad científica internacional.
- Mantenimiento de un archivo distribuido de contenidos científicos incluidos en las revistas científicas en formato electrónico.
- Implementación del protocolo OAI-PMH (*Open Archive Initiative-Protocol Metadata Harvesting*) para el intercambio de metadatos entre servicios.
- Implementación de un "cosechador" (*harvester*) de metadatos que permita la captura automática de metadatos entre los distintos repositorios que componen las revistas.
- Creación de un proveedor de metadatos (*data provider*), destinado a mantener uno o más repositorios que soporten OAI-PMH para la exposición de los metadatos, de forma que cada revista pueda poner a disposición de cualquier usuario y aplicación los metadatos referentes a sus contenidos.
- Creación de un proveedor de servicios (*service provider*) que recoja las respuestas de los proveedores de datos (las revistas) y los utilice para crear servicios de valor añadido relacionados con las revistas que forman parte de la iniciativa. Este sería el

portal que agruparía los datos de todas las revistas incluidas en la iniciativa y de los contenidos que incluyen.

3.2 ARQUITECTURA LÓGICA GENERAL

La arquitectura lógica del sistema está basada en la arquitectura UPS (*Universal Preprint Service*) y en su derivación propuesta por la iniciativa OAI (*Open Archives Initiative*).

La arquitectura UPS identifica dos entidades lógicas: el proveedor de metadatos (*data provider*) y el proveedor de servicios (*service provider*). El proveedor de datos, que en nuestro caso sería la revista electrónica será el responsable de la publicación y almacenamiento de los recursos en un repositorio y del mantenimiento de metadatos de esos recursos para que puedan “recolectarse” desde los proveedores de servicios. El proveedor de servicios, por su parte, es el responsable de recopilar estos datos de los proveedores de datos para proporcionar servicios de valor añadido en función de los metadatos que ha recogido. Entre el tipo de ayuda que el proveedor de servicios puede proporcionar está la creación de una interfaz de búsqueda unificada en todos los repositorios. El siguiente esquema representa el modelo descrito.



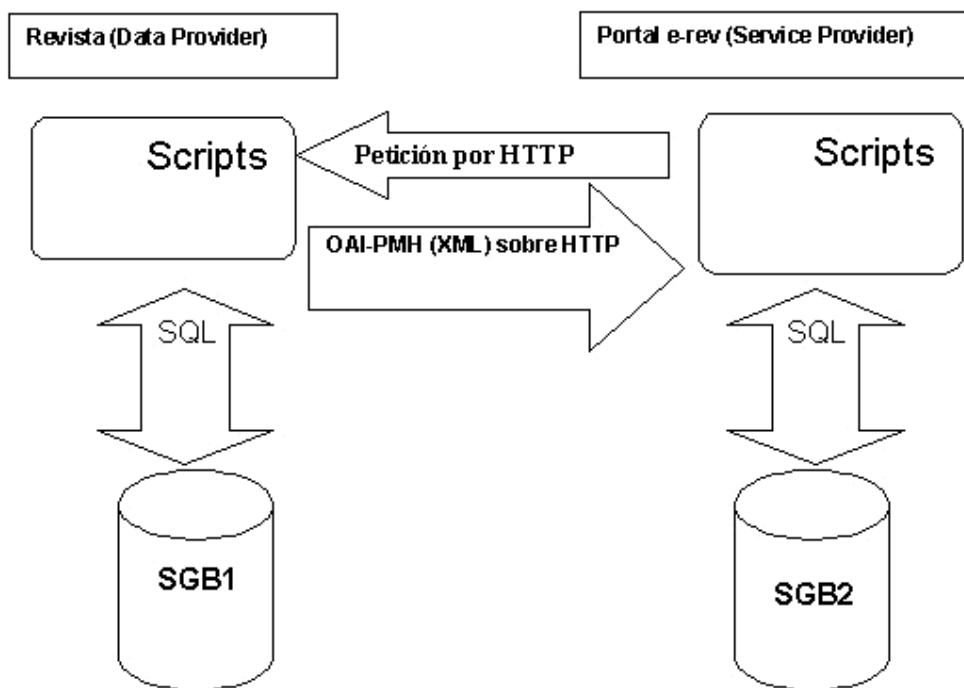
A continuación mostramos los distintos niveles del sistema:

- Nivel de presentación: La presentación y gestión de todo el modelo se hace a través del navegador Web, lo que nos permite desarrollar un entorno multiplataforma.
- Nivel de lógica de aplicación: Esta formado por implementaciones del protocolo OAI-PMH que se comunican mediante ficheros con formato XML que permite la interacción entre aplicaciones, independientemente de las peculiaridades concretas de cada una de ellas. Estas implementaciones pueden estar hechas en Java, C, o cualquiera de los lenguajes de

programación que se utilizan hoy en día en arquitecturas de este tipo

- Nivel de datos: Formado por los metadatos correspondientes a los recursos que mantiene cada repositorio o revista, estarán almacenados en bases de datos en un servidor web. Se recomienda el uso de bases de datos *open source* como MySQL, o Postgres, esto no será óbice para la utilización de software propietario como Access o SQL Server.

La solución propuesta se basa en la construcción de un conjunto de N capas de software, cada una de ellas encargada de ofrecer servicios a los otros elementos del sistema. Un ejemplo simplificado del flujo de trabajo de esta arquitectura podría ser el siguiente:

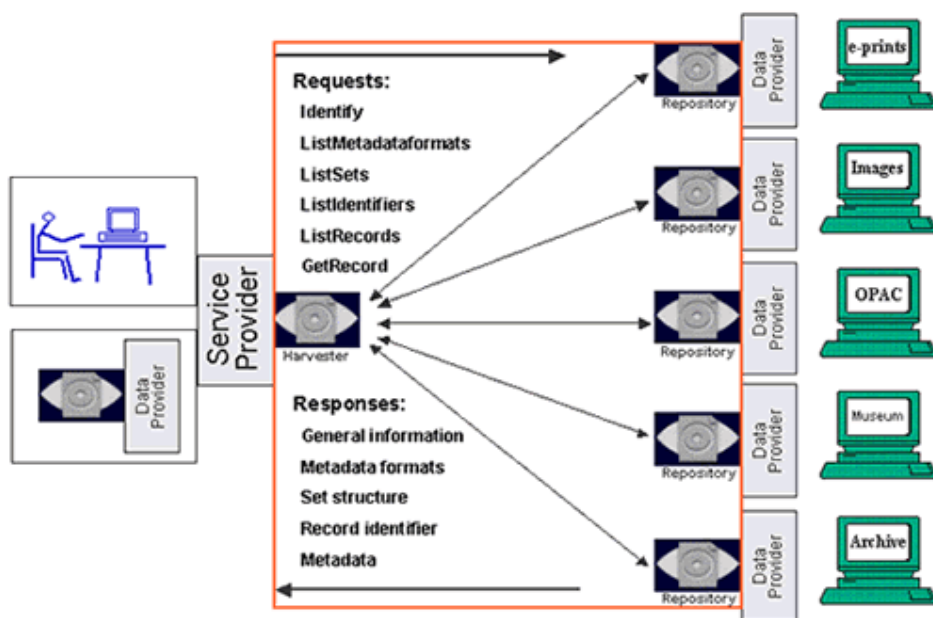


Si partimos del Portal de Revistas Electrónicas, que sería el proveedor de servicios, encontramos una primera capa de aplicación (JAVA, C, C++, PHP, Perl, ASP, JSP, C#) que contiene una implementación del protocolo OAI-PMH y que es la responsable de lanzar una petición de metadatos a través de HTTP a un repositorio o revista, reconocidos en el modelo como proveedores de datos o *data provider*.

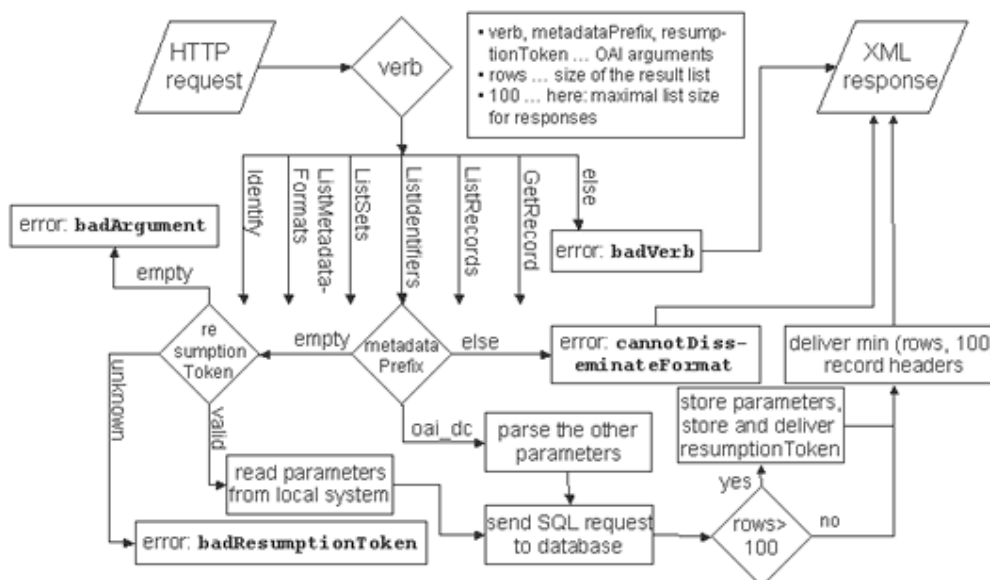
Cuando el servidor de datos recibe la petición de metadatos, la procesa a través de otra capa de aplicación programada y que contiene una implementación de OAI-PMH para realizar una consulta a su sistema de gestión de base de datos que contiene los metadatos referentes a los recursos que se encuentran en el repositorio o

revista. Mediante una consulta SQL se recopilan los metadatos. Una vez hecho esto la aplicación que ha recopilado los metadatos y que contiene una implementación de OAI-PMH se compone la respuesta a la petición realizada por el proveedor de datos en formato OAI-PMH con sintaxis XML y se envía nuevamente a través de HTTP. La capa de aplicación del *Service Provider* recoge la respuesta y se encarga de introducir en su propia base de datos los metadatos recibidos para componer posteriormente los servicios de valor añadido que presentará al usuario final.

El siguiente gráfico describe la estructura y el flujo de trabajo del sistema aquí descrito de forma detallada, especificando el tipo de peticiones HTTP:



La estructura mostrada por el gráfico anterior genera un flujo complejo de comunicación entre las distintas partes que forman parte del sistema. Este flujo de trabajo se muestra en total su amplitud y complejidad en el siguiente gráfico:



En esta figura se puede observar la forma en la que se realiza la comunicación entre los diferentes elementos del sistema. Esta comunicación se basa en una serie de verbos que forman parte del protocolo y que permiten la comunicación entre aplicaciones de forma transparente.

Si nos centramos en la arquitectura propia de cada uno de los elementos que componen esta arquitectura encontramos una subarquitectura propia para cada uno de ellos que será necesario respetar para la optimización del sistema.

3.3 DATA PROVIDER

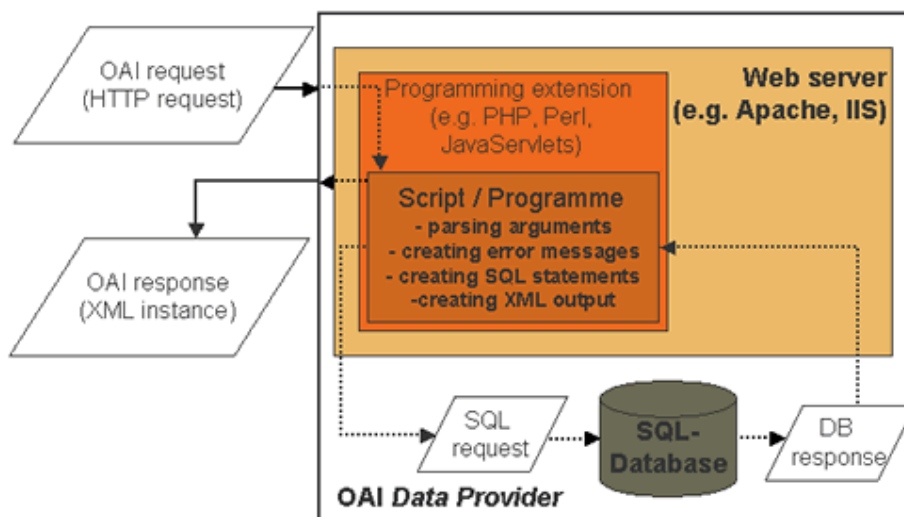
Estos son los requisitos que debe cumplir el Data Provider:

- Soporte de almacenamiento de metadatos: Es recomendable que los metadatos se almacenen en una base de datos relacional a la que se pueda acceder mediante consultas SQL, es necesario que exista un único identificador para cada recurso. También se puede utilizar un sistema de ficheros tipo LDAP
- Soporte Web: Un servidor web.
- Mantenimiento de un API para la posterior implementación del protocolo: Se puede utilizar cualquier lenguaje de programación que permita el desarrollo de aplicaciones web. Java, Perl, PHP son buenas elecciones a este respecto. El API debe contemplar el acceso a base de datos o al sistema de ficheros, si bien no es necesario implementar seguimiento de sesiones.
- URL: Mantenimiento de un identificador para el repositorio o revista único, basado en su URL.
- URI: Mantenimiento de un identificador único para cada recurso que forma parte del repositorio.
- Formato de los metadatos: Se pueden utilizar todos los formatos de metadatos que se desee. Como mínimo debe soportar Dublin Core sin cualificar.
- Datestamps: Se debe mantener un *datestamp* para la creación del recurso y otro para su modificación si es que es necesario.
- Soporte lógico para una jerarquía de conjuntos (*sets*): Permite el mantenimiento de clasificaciones de cara a una ordenación temática y/o conceptual de los recursos.
- Control de flujo: Se consigue a través de una implementación de la reanudación de la señal (*resumption token*) en la que se basa el diálogo entre el proveedor de metadatos y el servidor de datos. Este requisito no es obligatorio pero sí muy recomendable en cualquier proveedor de metadatos.

Respecto a sus componentes y arquitectura:

- Un analizador que valide los argumentos recibidos a partir de las peticiones OAI.
- Un generador de errores que cree respuestas XML para codificar los mensajes de error.
- Sistema de consulta a la base de datos o al sistema de ficheros para la recuperación de los metadatos.
- Generador de respuestas XML para la codificación de los metadatos que se enviarán como resultado de la petición de Service Provider o del Harvester.
- Control de flujo que realice una lista de secuencias incompleta cuando los repositorios sean muy grandes de cara a que el envío de la respuesta se haga poco a poco en función de la reanudación de la señal.

El siguiente diagrama muestra como interactúan estos componentes componiendo la arquitectura del Data Provider (en nuestro caso las revistas electrónicas).



3.4 SERVICE PROVIDER

Estos son los requisitos que debe cumplir Service Provider:

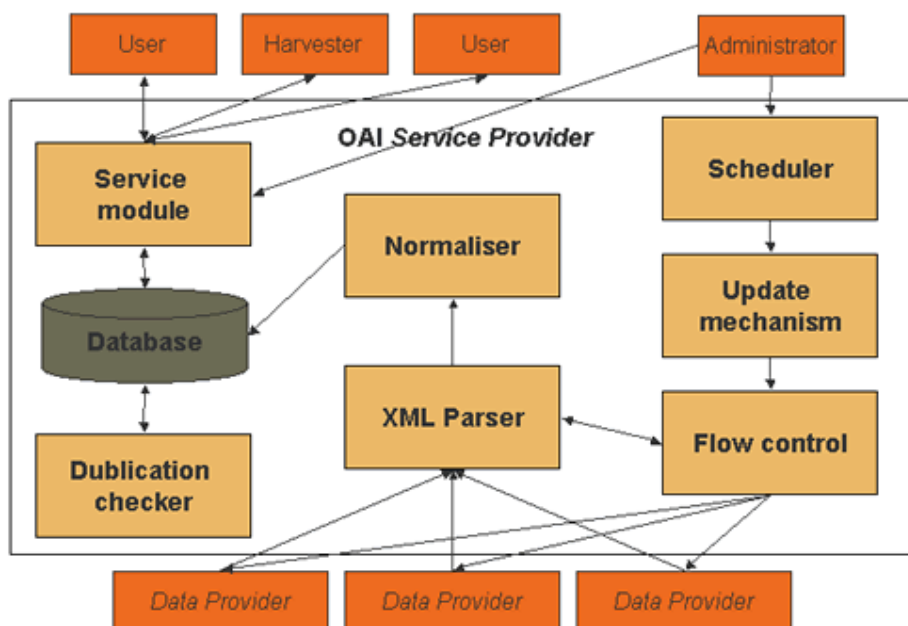
- Soporte de almacenamiento de metadatos: Es recomendable que los metadatos sean almacenados en una base de datos relacional a la que se pueda acceder mediante consultas SQL. También se puede utilizar un sistema de ficheros tipo LDAP o bases de datos XML.
- Soporte Web: Servidor web tipo Apache, IIS, etc.
- Mantenimiento de un API para la posterior implementación del protocolo: Se puede utilizar cualquier lenguaje de programación que permita el desarrollo de aplicaciones web. Java, Perl, PHP son buenas elecciones a este respecto. El API debe contemplar el acceso a base de datos o al sistema de ficheros, si bien no es necesario implementar seguimiento de sesiones.

Respecto a sus componentes y arquitectura:

- Gestión del archivo: Sistema de administración web que permita dar de alta a nuevas revista y repositorios (*Data Providers*) y gestionar la base de datos.
- Creador de peticiones: Se encarga de generar las peticiones para el proveedor de datos en función de los verbos OAI y de enviarlas a través de http.
- Temporizador (*Scheduler*): Encargado de comprobar si se han hecho modificaciones en los Data Provider que forman parte de la iniciativa. Este proceso se puede hacer a mano o automatizar a través de un cron, por ejemplo. Para el caso de las revistas que tienen periodicidad declarada la automatización de esta comprobación es absolutamente factible.
- Control de flujo (*Flow control*): Implementación para la reanudación de la señal en el caso de tener que recibir el listado de recursos de un proveedor de datos en varias veces debido a que este sea demasiado grande. El error de http 503 (que significa servicio no disponible) permite el análisis de la respuesta al mantener un periodo de reintento (*retry-after*).
- Mecanismo de actualización (*Update mechanism*): Mecanismo que se encargue de actualizar la información de una que se tiene en la base de datos sobre un proveedor de datos si es que el temporizador comprueba que se ha modificado.

- Analizador XML (Parser XML): Se encarga de analizar las respuestas extraídas de los proveedores de datos, con validación incluida.
- Herramienta de normalización (*Normaliser*): Se encarga de transformar los distintos formatos de metadatos que se reciben de los proveedores de servicios a una estructura homogénea, armonizando su posterior presentación.
- Base de datos: para almacenar la información sobre los proveedores de datos.
- Comprobador de redundancias (duplication checker): Se encarga de comprobar si dos proveedores de datos distintos repiten algún recurso y lo soluciona asignando un identificador distinto al recurso en función del proveedor de servicios al que pertenezcan.
- Modulo de servicios (Service module): Proporciona los servicios al usuario final, que únicamente interactúa con la base de datos que se ha creado después de recopilar los metadatos de los distintos Data Providers (revistas).

La interacción de estos componentes se materializa en la siguiente arquitectura:



4. PROTOCOLO OAI-PMH

OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting.

Básicamente OAI-PMH utiliza transacciones HTTP para emitir preguntas y obtener respuestas entre un servidor o archivo y un cliente o servicio recolector de metadatos. El segundo puede pedir al primero que le envíe metadatos según determinados criterios como por ejemplo la fecha de creación de los datos. En respuesta el primero devuelve un conjunto de registros en formato XML, incluyendo identificadores (URLs por ejemplo) de los objetos descritos en cada registro.

Las peticiones se emiten utilizando los métodos GET o POST del protocolo HTTP y constan de una lista de opciones con la forma de pares del tipo: clave=valor. Existen seis peticiones que un cliente puede realizar a un servidor:

- **Identify:** Utilizado para recuperar información sobre el servidor: nombre, versión del protocolo que utiliza, dirección del administrador, etc. Es el único verbo que no acepta parámetros.

<http://www.openarchives.org/OAI/openarchivesprotocol.html#Identify>

```
<?xml version="1.0" encoding="UTF-8"?>
<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>2002-02-08T12:00:01Z</responseDate>
  <request verb="Identify">http://memory.loc.gov/cgi-bin/oai</request>
  <Identify>
    <repositoryName>Library of Congress Open Archive Initiative
      Repository 1</repositoryName>
    <baseURL>http://memory.loc.gov/cgi-bin/oai</baseURL>
    <protocolVersion>2.0</protocolVersion>
    <adminEmail>somebody@loc.gov</adminEmail>
    <adminEmail>anybody@loc.gov</adminEmail>
    <earliestDatestamp>1990-02-01T12:00:00Z</earliestDatestamp>
    <deletedRecord>transient</deletedRecord>
    <granularity>YYYY-MM-DDThh:mm:ssZ</granularity>
    <compression>deflate</compression>
    <description>
      <oai-identifier
```

```

xmlns="http://www.openarchives.org/OAI/2.0/oai-identifier"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
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>lcoa1.loc.gov</repositoryIdentifier>
<delimiter>:</delimiter>

<sampleIdentifier>oai:lcoa1.loc.gov:loc.music/musdi.002</sampleIdentifier>
</oai-identifier>
</description>
<description>
  <eprints
    xmlns="http://www.openarchives.org/OAI/1.1/eprints"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/1.1/eprints
    http://www.openarchives.org/OAI/1.1/eprints.xsd">
    <content>

<URL>http://memory.loc.gov/ammem/oamh/lcoa1_content.html</URL>
  <text>Selected collections from American Memory at the Library
    of Congress</text>
  </content>
  <metadataPolicy/>
  <dataPolicy/>
</eprints>
</description>
<description>
  <friends
    xmlns="http://www.openarchives.org/OAI/2.0/friends/"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/friends/
    http://www.openarchives.org/OAI/2.0/friends.xsd">
    <baseURL>http://oai.east.org/foo/</baseURL>
    <baseURL>http://oai.hq.org/bar/</baseURL>
    <baseURL>http://oai.south.org/repo.cgi</baseURL>
  </friends>
</description>
</Identify>
</OAI-PMH>

```

- GetRecord. Utilizado para recuperar un registro concreto. Necesita dos argumentos: identificador del registro pedido y

especificación del formato bibliográfico en que se debe devolver.

<http://www.openarchives.org/OAI/openarchivesprotocol.html#GetRecord>

```
<?xml version="1.0" encoding="UTF-8"?>
<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>2002-02-08T08:55:46Z</responseDate>
  <request verb="GetRecord" identifier="oai:arXiv.org:cs/0112017"
    metadataPrefix="oai_dc">http://arXiv.org/oai2</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:arXiv.org:cs/0112017</identifier>
        <timestamp>2001-12-14</timestamp>
        <setSpec>cs</setSpec>
        <setSpec>math</setSpec>
      </header>
      <metadata>
        <oai_dc:dc
          xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
          xmlns:dc="http://purl.org/dc/elements/1.1/"
          xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
          http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
          <dc:title>Using Structural Metadata to Localize Experience of
            Digital Content</dc:title>
          <dc:creator>Dushay, Naomi</dc:creator>
          <dc:subject>Digital Libraries</dc:subject>
          <dc:description>With the increasing technical sophistication of
            both information consumers and providers, there is
            increasing demand for more meaningful experiences of
            digital
            information. We present a framework that separates digital
            object experience, or rendering, from digital object storage
            and manipulation, so the rendering can be tailored to
            particular communities of users.
          </dc:description>
          <dc:description>Comment: 23 pages including 2 appendices,
            8 figures</dc:description>
          <dc:date>2001-12-14</dc:date>
        </oai_dc:dc>
```

```

</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

- ListIdentifiers. Recupera los encabezamientos de los registros, en lugar de los registros completos. Permite argumentos como el rango de fechas entre los que queremos recuperar los datos. Agrupamientos mediante SETS, control de flujo (resumption Token) y especificación del formato bibliográfico en que se debe devolver.

<http://www.openarchives.org/OAI/openarchivesprotocol.html#ListIdentifiers>

```

<?xml version="1.0" encoding="UTF-8"?>
<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>2002-06-01T19:20:30Z</responseDate>
  <request verb="ListIdentifiers" from="1998-01-15"
    metadataPrefix="oldarXiv"
    set="physics:hep">http://an.oa.org/OAI-script</request>
  <ListIdentifiers>
    <header>
      <identifier>oai:arXiv.org:hep-th/9801001</identifier>
      <timestamp>1999-02-23</timestamp>
      <setSpec>physics:hep</setSpec>
    </header>
    <header>
      <identifier>oai:arXiv.org:hep-th/9801002</identifier>
      <timestamp>1999-03-20</timestamp>
      <setSpec>physics:hep</setSpec>
      <setSpec>physics:exp</setSpec>
    </header>
    <header>
      <identifier>oai:arXiv.org:hep-th/9801005</identifier>
      <timestamp>2000-01-18</timestamp>
      <setSpec>physics:hep</setSpec>
    </header>
    <header status="deleted">
      <identifier>oai:arXiv.org:hep-th/9801010</identifier>
      <timestamp>1999-02-23</timestamp>

```

```

<setSpec>physic:hep</setSpec>
<setSpec>math</setSpec>
</header>
<resumptionToken expirationDate="2002-06-01T23:20:00Z"
  completeListSize="6"
  cursor="0">xxx45abttzy</resumptionToken>
</ListIdentifiers>
</OAI-PMH>

```

- ListRecords. Igual que el anterior pero recupera los registros completos. Permite argumentos como el rango de fechas entre los que queremos recuperar los datos. Agrupamientos mediante SETS, control de flujo (resumption Token) y especificación del formato bibliográfico en que se debe devolver.

<http://www.openarchives.org/OAI/openarchivesprotocol.html#ListRecords>

```

<?xml version="1.0" encoding="UTF-8"?>
<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>2002-06-01T19:20:30Z</responseDate>
<request verb="ListRecords" from="1998-01-15"
  set="physics:hep"
  metadataPrefix="oai_rfc1807">
  http://an.oa.org/OAI-script</request>
<ListRecords>
<record>
  <header>
    <identifier>oai:arXiv.org:hep-th/9901001</identifier>
    <timestamp>1999-12-25</timestamp>
    <setSpec>physics:hep</setSpec>
    <setSpec>math</setSpec>
  </header>
  <metadata>
    <rfc1807 xmlns=
      "http://info.internet.isi.edu:80/in-notes/rfc/files/rfc1807.txt"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation=
        "http://info.internet.isi.edu:80/in-notes/rfc/files/rfc1807.txt
        http://www.openarchives.org/OAI/1.1/rfc1807.xsd">
      <bib-version>v2</bib-version>

```

```

<id>hep-th/9901001</id>
<entry>January 1, 1999</entry>
<title>Investigations of Radioactivity</title>
<author>Ernest Rutherford</author>
<date>March 30, 1999</date>
</rfc1807>
</metadata>
<about>
  <oai_dc:dc
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
  http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:publisher>Los Alamos arXiv</dc:publisher>
  <dc:rights>Metadata may be used without restrictions as long
as
  the oai identifier remains attached to it.</dc:rights>
  </oai_dc:dc>
</about>
</record>
<record>
  <header status="deleted">
    <identifier>oai:arXiv.org:hep-th/9901007</identifier>
    <datestamp>1999-12-21</datestamp>
  </header>
</record>
</ListRecords>
</OAI-PMH>

```

- ListSets. Recupera un conjunto de registros. Estos conjuntos son creados opcionalmente por el servidor para facilitar una recuperación selectiva de los registros. Sería una clasificación de los contenidos según diferentes entradas. Un cliente puede pedir que se recuperen solo los registros pertenecientes a una determinada clase. Los conjuntos pueden ser simples listas o estructuras jerárquicas. Acepta un parámetro de control de flujo (resumption Token).

<http://www.openarchives.org/OAI/openarchivesprotocol.html#ListSets>

```
<?xml version="1.0" encoding="UTF-8"?>
<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>2002-08-11T07:21:33Z</responseDate>
  <request verb="ListSets">http://an.oa.org/OAI-script</request>
  <ListSets>
    <set>
      <setSpec>music</setSpec>
      <setName>Music collection</setName>
    </set>
    <set>
      <setSpec>music:(muzak)</setSpec>
      <setName>Muzak collection</setName>
    </set>
    <set>
      <setSpec>music:(elec)</setSpec>
      <setName>Electronic Music Collection</setName>
      <setDescription>
        <oai_dc:dc
          xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
          xmlns:dc="http://purl.org/dc/elements/1.1/"
          xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
            http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
          <dc:description>This set contains metadata describing
            electronic music recordings made during the 1950ies
          </dc:description>
        </oai_dc:dc>
      </setDescription>
    </set>
    <set>
      <setSpec>video</setSpec>
      <setName>Video Collection</setName>
    </set>
  </ListSets>
</OAI-PMH>
```

- ListMetadataFormats. Devuelve la lista de formatos bibliográficos que utiliza el servidor. Acepta un parámetro para pasarle el identificador bibliográfico si no lo soporta devuelve un error: *idDoesNotExist*.

<http://www.openarchives.org/OAI/openarchivesprotocol.html#ListMetadataFormats>

```
<?xml version="1.0" encoding="UTF-8"?>
<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>2002-02-08T14:27:19Z</responseDate>
  <request verb="ListMetadataFormats"
    identifier="oai:perseus.tufts.edu:Perseus:text:1999.02.0119">
    http://www.perseus.tufts.edu/cgi-bin/pdataprov</request>
  <ListMetadataFormats>
    <metadataFormat>
      <metadataPrefix>oai_dc</metadataPrefix>
      <schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd
    </schema>

    <metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/
    </metadataNamespace>
    </metadataFormat>
    <metadataFormat>
      <metadataPrefix>olac</metadataPrefix>
      <schema>http://www.language-archives.org/OLAC/olac-
0.2.xsd</schema>
      <metadataNamespace>http://www.language-
archives.org/OLAC/0.2/
    </metadataNamespace>
    </metadataFormat>
    <metadataFormat>
      <metadataPrefix>perseus</metadataPrefix>
      <schema>http://www.perseus.tufts.edu/persmeta.xsd</schema>

    <metadataNamespace>http://www.perseus.tufts.edu/persmeta.dtd
    </metadataNamespace>
    </metadataFormat>
  </ListMetadataFormats>
</OAI-PMH>
```

El protocolo soporta múltiples formatos para expresar los metadatos, no obstante requiere que todos los servidores ofrezcan los registros utilizando Dublin Core simple (también conocido como no calificado), codificado en XML. Además de éste formato cada servidor es libre de ofrecer los registros en otro/s formatos adicionales (MARC por

ejemplo). Un cliente puede pedir que los registros se le sirvan en cualquiera de los formatos soportados por el servidor. La idea subyacente aquí es que en el futuro las diferentes comunidades que utilicen el protocolo definan sus propios formatos que sean más ricos y más precisos que el Dublin Core. Por ejemplo la comunidad de archivos de eprints está trabajando en un formato denominado AMF (Academic Metadata Format) <http://amf.openlib.org/doc/ebisu.html> que sea capaz de describir todos los elementos que intervienen en el proceso de comunicación científica: documentos, autores, instituciones y canales de distribución de documentos.

Las respuestas del servidor estarán formateadas según el protocolo HTTP con los adecuados encabezamientos. Serán documentos XML correctos que se podrán validar contra el esquema definido en el protocolo y disponible en la dirección:

<http://www.openarchives.org/OAI/openarchivesprotocol.html>

El protocolo OAI-PMH, aunque naciera centrado en publicaciones científicas, sirve perfectamente para la comunicación de metadatos sobre cualquier material almacenado en soporte electrónico. No hay nada en el protocolo que impida a los implementadores transmitir el contenido propiamente dicho de esos materiales.

El marco diseñado por OAI es intencionalmente simple con el propósito de proporcionar una mínima complicación para las instituciones que deseen implementarlo. Los protocolos como el Z39.50 tienen una funcionalidad más completa, por ejemplo, tratan cuestiones como el manejo de sesiones, gestión de conjuntos de resultados y permiten la especificación de predicados para filtrar los resultados obtenidos. Sin embargo, esta funcionalidad acarrea un incremento en la complejidad de la implementación y, en consecuencia, de los costes. Por lo tanto no se trata de reemplazar otras iniciativas, sino desarrollar una alternativa que sea fácil de implementar y de desarrollar para propósitos diferentes de los que ya tratan los sistemas de interoperabilidad existentes. El futuro juzgará si esta barrera mínima de interoperabilidad es realista y funcional.

La OAI no define o prescribe ningún esquema para la gestión de derechos. Los temas relacionados con restricciones en el acceso y gestión de la propiedad intelectual son la responsabilidad de los proveedores de datos.

Aspectos que no trata el protocolo son por ejemplo cuestiones de gestión o autorización para el acceso de los clientes. El servidor deberá recurrir a métodos externos si desea limitar los clientes a los que sirva información. En relación con este punto está la utilización que los clientes hagan de los datos. También queda fuera del

protocolo. Finalmente, tampoco trata el tema de cómo los clientes pueden localizar aquellos servidores que contengan los datos que necesitan.

Cambios de la versión 1.x a la 2.0:

Los orígenes del protocolo OAI-PMH fueron hacia el año 1999 en la reunión celebrada en Santa Fe (Nuevo México, USA), pero hasta el 2001 no vio la luz la versión 1.0, La intención era que este protocolo, con mínimas modificaciones, permaneciera estable al menos durante un año, mientras las distintas comunidades lo probaban y experimentaban con él. Después de un año y ya en el 2002 nace la versión 2.0 del protocolo OAI-PMH, una versión mucho más estable de las dos anteriores la 1.0 y la 1.1.

Cambios realizados en la versión 2.0

- Se ha clarificado el modelo de datos, eliminando las ambigüedades que existían en la versión 1.x.
- Se definió que el protocolo al menos tuviera un formato para los metadatos, Dublin Core sencillo o sin cualificar.
- Se definió un esquema XML único para validar todas las respuestas de todas las peticiones OAI-PMH.
- Se definieron más errores referidos a los problemas derivados de las conexiones, mejorando el control de flujo (resumption Tokens), mejorando considerablemente la transmisión de datos.
- Se añadieron nuevas funcionalidades, ahora OAI-PMH soporta harvesteos en función de los parámetros temporales que le pasemos.
- Manejo de errores mucho más limpio separando entre los errores surgidos del protocolo OAI-PMH o errores producidos por la comunicación o el propio servidor web.
- El verbo Identify manda una información mucho más clara que la que enviaba las versiones 1.x. permitiendo varias estrategias de harvesteo posibles.
- Se introdujeron nuevos verbos que permiten el harvesteo basados en grupos SET.
- Se generó nueva documentación, no sólo del protocolo OAI-PMH, también se hicieron guías para la implantación en las

distintas plataformas, sobre el desarrollo del software que implementara este protocolo, etc.

- Permite la duplicación de registros procedentes de distintos Data Provider
- Se hicieron esquemas MARCXML, que mejoraron los esquemas que traía la versión 1.0 (oai-marc)
- Se añadió la identificación basada oai-identifier, con la finalidad de poder saber que formatos bibliográficos soporta el Data Provider.

Entre los planes para el futuro están la creación de una versión SOAP (Simple Object Access Protocol) <http://www.w3.org/TR/SOAP> del protocolo.

5. DUBLIN CORE

En este punto explicaremos el formato principal de intercambio de registros mediante el protocolo OIA-PMH.

La Iniciativa de Metadatos Dublin Core DCMI es una organización dedicada a la promoción y difusión de normas interoperables sobre metadatos y el desarrollo de vocabularios especializados en metadatos para la descripción de recursos que permitan sistemas de recuperación más inteligentes.

Esta iniciativa se muestra mediante la definición de un esquema XML con los campos propuestos por la DCMI.

A continuación describiremos los elementos que forman dicho esquema:

Dublin Core	Etiqueta
Title	Título
Creator	Autor
Subject	Claves o Materia
Description	Descripción
Publisher	Editor
Contributor	Colaborador
Date	Fecha
Type	Tipo
Format	Formato
Identifier	Identificador
Source	Fuente
Language	Idioma
Relation	Relación
Coverage	Cobertura
Rights	Derechos

Los elementos poseen nombres descriptivos que pretenden transmitir un significado semántico a los mismos. Para promover una interoperabilidad global, una descripción del valor de algunos elementos podrá ser asociada a vocabularios controlados. Se asume que otros vocabularios controlados serán desarrollados para asegurar esta interoperabilidad en dominios específicos.

Cada elemento es opcional y puede repetirse. Además, los elementos pueden aparecer en cualquier orden.

Podemos clasificar estos elementos en tres grupos que indican la clase o el ámbito de la información que se guarda en ellos:

1. Elementos relacionados principalmente con el contenido del recurso: Title, Subject, Description, Source, Language, Relation, Coverage.
2. Elementos relacionados principalmente con el recurso cuando es visto como una propiedad intelectual: Creator, Publisher, Contributor, Rights
3. Elementos relacionados principalmente con la instanciación del recurso: Date, Type, Format, Identifier.

Descripción de los elementos:

- **Título** (Title): el nombre dado a un recurso, usualmente por el autor.
- **Autor o Creador** (Creator): la persona u organización responsable de la creación del contenido intelectual del recurso. Por ejemplo, los autores en el caso de documentos escritos, artistas, fotógrafos e ilustradores en el caso de recursos visuales.
- **Claves** (Subject): los tópicos del recurso. Típicamente, Subject expresará las claves o frases que describen el título o el contenido del recurso. Se fomentará el uso de vocabularios controlados y de sistemas de clasificación formales.
- **Descripción** (Description): una descripción textual del recurso, tal como un resumen en el caso de un documento o una descripción del contenido en el caso de un documento visual.
- **Editor** (Publisher): la entidad responsable de hacer que el recurso se encuentre disponible en la red en su formato actual, por ejemplo la empresa editora, un departamento universitario u otro tipo de organización.
- **Colaborador** (Contributor): una persona u organización que haya tenido una contribución intelectual significativa en la creación del recurso pero cuyas contribuciones son secundarias en comparación a las de las personas u organizaciones especificadas en el elemento Creator (por ejemplo, editor, ilustrador y traductor).
- **Fecha** (Date): una fecha en la que el recurso se puso a disposición del usuario en su forma actual. Esta fecha no ha de confundirse con la que pertenece al elemento Coverage, que

sería asociada con el recurso sólo en la medida en que el contenido intelectual está de algún modo relacionado con esa fecha.

- **Tipo del Recurso** (Type): la categoría del recurso, por ejemplo página personal, romance, poema, minuta, diccionario. Para asegurar la interoperabilidad, Type debería ser seleccionado de entre una lista de valores que actualmente se encuentra bajo desarrollo en un grupo de trabajo.
- **Formato** (Format): El formato de datos de un recurso, usado para identificar el software y posiblemente, el hardware que se necesitaría para mostrar el recurso. Para asegurar la interoperabilidad, los valores de Format deberían ser seleccionados de entre una lista de valores que actualmente se encuentra bajo desarrollo en un grupo de trabajo.
- **Identificador del Recurso** (Identifier): secuencia de caracteres usados para identificar unívocamente un recurso. Ejemplos para recursos en línea pueden ser URLs y URNs (cuando estén implementados). Para otros recursos pueden ser usados otros formatos de identificadores, como por ejemplo ISBN ("International Standard Book Number" - Número Internacional Normalizado para Libros)
- **Fuente** (Source): Secuencia de caracteres utilizado para identificar unívocamente un trabajo a partir del cual proviene el recurso actual. Por ejemplo, es posible usar Source con la fecha de 1603 como descripción de una película basada en una obra de Shakespeare, pero es preferible, en ese caso, usar Relation "IsBasedOn" con una referencia a un recurso distinto cuya descripción contenga el elemento Date con valor 1603.
- **Idioma** (Language): Lengua/s del contenido intelectual del recurso. Prácticamente el contenido de este campo debería coincidir con los de la RFC 1766
- **Relación** (Relation): un identificador de un segundo recurso y su relación con el recurso actual. Este elemento permite enlazar los recursos relacionados y las descripciones de los recursos.
- **Cobertura** (Coverage): la característica de cobertura espacial y/o temporal del contenido intelectual del recurso. La cobertura espacial se refiere a una región física (por ejemplo, sector celestial); uso de coordenadas (por ejemplo, longitud y latitud) o nombres de lugares extraídos de una lista controlada. La cobertura temporal se refiere al contenido del recurso en vez de a cuando fue creado o puesto accesible ya que este último

pertenece al elemento Date. Se usa el mismo formato basado en <http://www.w3.org/TR/NOTE-datetime>.

- **Derechos** (Rights): Una referencia (URL, por ejemplo) para una nota sobre derechos de autor, para un servicio de gestión de derechos o para un servicio que dará información sobre términos y condiciones de acceso a un recurso. Una especificación formal del elemento Rights se encuentra actualmente en discusión y por lo tanto su uso se considera experimental.

Siguiendo el esquema que podemos encontrar en:
http://www.openarchives.org/OAI/2.0/oai_dc.xsd

```
<schema targetNamespace="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns="http://www.w3.org/2001/XMLSchema"
  elementFormDefault="qualified" attributeFormDefault="unqualified">

<annotation>
  <documentation>
    XML Schema 2002-03-18 by Pete Johnston.
    Adjusted for usage in the OAI-PMH.
    Schema imports the Dublin Core elements from the DCMI schema for unqualified Dublin Core.
    2002-12-19 updated to use simpledc20021212.xsd (instead of simpledc20020312.xsd)
  </documentation>
</annotation>

<import namespace="http://purl.org/dc/elements/1.1/"
  schemaLocation="http://dublincore.org/schemas/xmls/simpledc20021212.xsd"/>

<element name="dc" type="oai_dc:oai_dcType"/>

<complexType name="oai_dcType">
  <choice minOccurs="0" maxOccurs="unbounded">
    <element ref="dc:title"/>
    <element ref="dc:creator"/>
    <element ref="dc:subject"/>
    <element ref="dc:description"/>
    <element ref="dc:publisher"/>
    <element ref="dc:contributor"/>
    <element ref="dc:date"/>
    <element ref="dc:type"/>
    <element ref="dc:format"/>
    <element ref="dc:identifier"/>
    <element ref="dc:source"/>
    <element ref="dc:language"/>
    <element ref="dc:relation"/>
    <element ref="dc:coverage"/>
    <element ref="dc:rights"/>
  </choice>
</complexType>

</schema>
```

Vemos que cada uno de los elementos se puede aparece cero o más veces así como el tipo de datos definido en: <http://dublincore.org/schemas/xmls/simpledc20021212.xsd>


```

<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns="http://purl.org/dc/elements/1.1/"
  targetNamespace="http://purl.org/dc/elements/1.1/"
  elementFormDefault="qualified"
  attributeFormDefault="unqualified">
  <xs:annotation>
    <xs:documentation xml:lang="en">
      Simple DC XML Schema, 2002-10-09 by Pete Johnston (p.johnston@ukoln.ac.uk),
      Carl Lagoze (lagoze@cs.cornell.edu), Andy Powell (a.powell@ukoln.ac.uk),
      Herbert Van de Sompel (hvdsomp@yahoo.com).
      This schema defines terms for Simple Dublin Core, i.e. the 15
      elements from the http://purl.org/dc/elements/1.1/ namespace, with
      no use of encoding schemes or element refinements.
      Default content type for all elements is xs:string with xml:lang
      attribute available.
      Supercedes version of 2002-03-12.
      Amended to remove namespace declaration for http://www.w3.org/XML/1998/namespace
      namespace, and to reference lang attribute via built-in xml: namespace prefix.xs:appinfo also
      moved.
    </xs:documentation>
  </xs:annotation>
  <xs:import namespace="http://www.w3.org/XML/1998/namespace"
    schemaLocation="http://www.w3.org/2001/03/xml.xsd">
  </xs:import>

  <xs:element name="title" type="elementType"/>
  <xs:element name="creator" type="elementType"/>
  <xs:element name="subject" type="elementType"/>
  <xs:element name="description" type="elementType"/>
  <xs:element name="publisher" type="elementType"/>
  <xs:element name="contributor" type="elementType"/>
  <xs:element name="date" type="elementType"/>
  <xs:element name="type" type="elementType"/>
  <xs:element name="format" type="elementType"/>
  <xs:element name="identifier" type="elementType"/>
  <xs:element name="source" type="elementType"/>
  <xs:element name="language" type="elementType"/>
  <xs:element name="relation" type="elementType"/>
  <xs:element name="coverage" type="elementType"/>
  <xs:element name="rights" type="elementType"/>

  <xs:group name="elementsGroup">
    <xs:sequence>
      <xs:choice minOccurs="0" maxOccurs="unbounded">
        <xs:element ref="title"/>
        <xs:element ref="creator"/>
        <xs:element ref="subject"/>
        <xs:element ref="description"/>
        <xs:element ref="publisher"/>
        <xs:element ref="contributor"/>
        <xs:element ref="date"/>
        <xs:element ref="type"/>
        <xs:element ref="format"/>
        <xs:element ref="identifier"/>
        <xs:element ref="source"/>
        <xs:element ref="language"/>
        <xs:element ref="relation"/>
        <xs:element ref="coverage"/>
        <xs:element ref="rights"/>
      </xs:choice>
    </xs:sequence>
  </xs:group>

  <xs:complexType name="elementType">
    <xs:simpleContent>
      <xs:extension base="xs:string">
        <xs:attribute ref="xml:lang" use="optional"/>
      </xs:extension>
    </xs:simpleContent>
  </xs:complexType>

</xs:schema>

```

Donde podemos observar que cada uno de los elementos es de tipo "String" con un atributo adicional de lenguaje.

Destacar que en nuestro caso toda la implantación que se ha llevado a cabo ha sido siguiendo el formato Dublin Core, pero podemos encontrar otro tipo de formatos a la hora de harvestear información del Data Provider.

6. SOFTWARE PARA LOS DATA PROVIDER

- [CDSware](#) Software desarrollado por el CERN (Apache, php/python, MySQL)
- [DLESE OAI](#) Software desarrollado por Digital Library for Earth System Education (JAVA empaquetado en fichero WAR, utiliza XML como almacén de datos)
- [Dspace](#) Software desarrollado por HP Labs y MIT Libraries (TomCat 4.0, JDK 1.3+, Ant 1.4 (si se realizan modificaciones en el código JAVA), PostgreSQL 7.2+)
- [OAICat](#) Software desarrollado por OCLS (JAVA Servlet, TomCat 4.0, JDK 1.3+, varias BD a través del driver JDBC o XML)
- <http://uilib-oai.sourceforge.net/> Enlace donde se describen varios Data Providers.
- [http://www.soros.org/openaccess/pdf/OSI Guide to Institutional Repository Software v2.pdf](http://www.soros.org/openaccess/pdf/OSI_Guide_to_Institutional_Repository_Software_v2.pdf) (pdf)
[http://www.soros.org/openaccess/software/OSI Guide to Institutional Repository Software v2.htm](http://www.soros.org/openaccess/software/OSI_Guide_to_Institutional_Repository_Software_v2.htm) (HTML)
Estudio comparativo de algunos DataProvider

7. SOFTWARE PARA LOS SERVICES PROVIDER

- [ARC Source](#): desarrollado por Old Dominion University. Requiere JDK 1.4, Tomcat 4.0x, y un servidor de base de datos relacionales (ha sido probado por Oracle y MySQL hasta ahora)
- [Celestial](#) desarrollado por la University of Southampton. Requiere oai-perl v2, MySQL, Perl 5.6x y un servidor web con posibilidad de ejecutar CGIs.
- [DP9](#) desarrollado por Old Dominion University. Requiere Tomcat y JDK 1.3

Para el portal e-revistas, se ha seleccionado el software *ARC harvester and search engine*, que tiene las siguientes características:

- **ARC** es un servicio experimental creado con el objetivo de investigar temas relacionados con la recolección de metadatos siguiendo el protocolo OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) y cómo hacerlos disponibles a los usuarios. Más que un servicio en sí mismo es un software que puede ser utilizado por instituciones que quieran crear sus propios servicios. El código fuente está disponible en la red de forma gratuita.
- URL: <http://arc.cs.odu.edu>
- Desarrollado por: Digital Library Research Group de la Old Dominion University
- Está desarrollado en Java y requiere JDK1.4
- Servidor de aplicaciones Tomcat 4.0x
- Sistema de gestión de base de datos con driver JDBC. Está probado con Oracle y MySQL
- Sistema Operativo: Windows, Linux, Solaris, etc

9. BIBLIOGRAFÍA Y ENLACES

- Tecnociencia: <http://www.tecnociencia.es/>
- E-Revistas: <http://www.tecnociencia.es/e-revistas/>
- Bien común y Open Access (Antonio Lafuente):
<http://www.madrimasd.org/informacionidi/debates/anteriores/debateActualidad-BienComun.asp?pagina=informacion&idForo=GlobalIDI-13>
- Iniciativa de Acceso abierto de Budapest:
http://www.geotropico.org/1_1_Documentos_BOAI.html
- Dublín Core: <http://www.dublincore.org/> y una descripción en español proporcionada por rediris:
<http://www.rediris.es/metadata/>
- OAI: <http://www.openarchives.org/>
- Protocolo OAI-PMH:
<http://www.openarchives.org/OAI/openarchivesprotocol.html>