

OAI-PMH: Protocolo para la transmisión de contenidos en Internet

José Manuel Barrueco
Biblioteca de Ciències Socials. Universitat de València.
46010 València. Spain
Jose.Barrueco@uv.es

Imma Subirats Coll
Biblioteca del Dept. de Política Territorial i Obres Públiques. Generalitat de Catalunya
08029 Barcelona. Spain
immasubirats@myrealbox.com

Resumen:

Se describe el protocolo OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting) utilizado para la transmisión de metadatos en Internet. Se analiza el contexto en el que nació, las comunidades de depósitos de documentos científicos y cómo se ha desarrollado y extendido su alcance a cualquier material en formato electrónico. Se describe brevemente su arquitectura basada en el modelo cliente – servidor donde los primeros, llamados archivos, ponen a disposición del público metadatos en formato Dublin Core para que puedan ser recuperados por los segundos. La comunicación se realiza mediante el protocolo http. Las respuestas están codificadas en XML. Finalmente se hace una revisión de las principales instituciones que lo han implementado, los servicios que se han basado en él y se dan una serie de herramientas que facilitan la creación de archivos abiertos.

Palabras Clave: Bibliotecas digitales ; Protocolos de Internet ; Metadatos ; XML ; Eprints

0.- Introducción

El trabajo que presentamos en este número monográfico tiene como objetivo divulgar entre los profesionales de nuestro país un nuevo protocolo para la transmisión de contenidos en Internet denominado OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting). <http://www.openarchives.org>

Si bien es un protocolo de aparición reciente, los primeros trabajos para su desarrollo se iniciaron en 1999, el interés que ha despertado entre la comunidad de bibliotecarios de todo el mundo ha sido muy grande. Este interés viene probado por la gran cantidad de jornadas celebradas en distintos países, por el número de artículos publicados en revistas especializadas, por la importancia de las instituciones que lo han apoyado desde el primer momento y por los numerosos proyectos de investigación financiados en el último año tanto por la Unión Europea como por la National Science Foundation de los USA.

Paradójicamente el interés entre los investigadores y profesionales de nuestro país parece escaso cuando no nulo. En las últimas jornadas técnicas sobre el

tema celebradas en Ginebra <http://library.cern.ch/Announcement.htm> y Lisboa http://www.oaforum.org/workshops/lisb_invitation.php el pasado año, la asistencia de participantes españoles fue simbólica. Las últimas jornadas sobre bibliotecas digitales JBIDI 2002 <http://mariachi.dsic.upv.es/jbidi/jbidi2002/> no incluyeron ninguna presentación sobre el mismo.

Con objeto de paliar esta deficiencia en un asunto que consideramos de vital interés para la investigación en bibliotecas digitales, presentamos este trabajo con un carácter eminentemente divulgativo. La importancia de OAI-PMH se puede resumir en una frase: OAI-PMH está llamado a ser a las bibliotecas digitales lo que HTTP es hoy al web.

El resto del trabajo se estructura de la siguiente forma: la sección 1 da una visión general del protocolo abordando cuestiones como cuáles son sus objetivos, cómo funciona o qué instituciones lo respaldan. La sección 2 hace una breve historia de su evolución desde su nacimiento en 1999 hasta la última versión publicada en Junio de 2002. La sección 3 entra, sin profundizar demasiado, en las cuestiones técnicas del protocolo. La 4, describe las principales instituciones que lo han adoptado y los servicios que se han creado utilizando las funcionalidades que aporta. Finalmente la sección 5 describe una serie de herramientas que nos pueden ayudar a la hora de implementarlo.

1.- ¿Qué es OAI-PMH?

En primer lugar hemos de matizar varios puntos respecto a su nombre. El término **archivo** refleja los orígenes de la iniciativa en el seno de las comunidades de eprints donde es sinónimo de depósito de documentos científicos a texto completo. No tiene nada que ver aquí con el concepto tradicional de archivo con connotaciones de preservación y conservación. Se utiliza por lo tanto con un sentido mucho más amplio, como un depósito para almacenar cualquier tipo de información. El término **abierto** se refiere al punto de vista de la arquitectura del sistema. Se trata de definir interfaces que faciliten la disponibilidad de contenidos procedentes de una variedad de proveedores. Apertura tampoco significa gratuidad o acceso ilimitado a dicha información

La Open Archives Initiative (OAi) se creó con la misión de desarrollar y promover estándares de interoperabilidad para facilitar la difusión eficiente de contenidos en Internet. Surgió como un esfuerzo para mejorar el acceso a archivos de publicaciones electrónicas (eprints), en definitiva, para incrementar la disponibilidad de las publicaciones científicas. Los trabajos iniciales se centraron en el desarrollo de marcos de interoperabilidad para la federación de archivos de eprints, pronto apareció evidente que dichos marcos (permitir el intercambio de múltiples formatos bibliográficos entre distintas máquinas utilizando un protocolo común) tenían aplicaciones más allá de esta comunidad. Por ello se adoptó un objetivo mucho más amplio: abrir el acceso a un rango de materiales digitales

Por lo tanto, la OAi no es solamente un proyecto centrado en publicaciones científicas, sino en 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. No obstante esto no es el objeto principal de OAI-PMH.

Los metadatos a transmitir vía OAI-PMH deberán codificarse en Dublin Core sin calificar con objeto de minimizar los problemas derivados de las conversiones entre múltiples formatos. Aunque se está investigando la creación de servicios tales como una interfaz de búsqueda a través de formatos heterogéneos de metadatos, una solución menos complicada y por lo tanto más fácil de implementar es requerir a los implementadores convertir sus datos a un formato común. Los quince elementos del Dublin Core han evolucionado a lo largo de los pasados años como el estándar de facto para los metadatos simples y multidisciplinares.

¿Qué relación existe con otros protocolos como el Z39.50? 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.

La OAi ha obtenido financiación en USA de la National Science Foundation. De la gestión administrativa y técnica se encargan dos comités que están coordinados por Herbert Van de Sompel y Carl Lagoze, ambos de la Universidad de Cornell.

2.- Un poco de historia, de la Convención de Santa Fe a la OAi

Los orígenes de OAi radican en un creciente interés en la búsqueda de alternativas a los modelos tradicionales de comunicación científica. En algunas disciplinas, principalmente en ciencias, comenzaron a surgir los llamados archivos o repositorios de documentos electrónicos como alternativa para la rápida comunicación de resultados de investigación. Esos documentos se han llamado eprints de forma genérica. Este nuevo concepto agrupa tanto aquellos documentos que no han pasado por un proceso de certificación o peer review (preprints) como aquellos que si han pasado por tal proceso o postprints (artículos,

libros, etc.). El más conocido de estos archivos es sin duda arXiv.org creado por Paul Ginsparg en Los Alamos (USA) para el área de Física.

En Octubre de 1999 se organizó una reunión en Santa Fe (Nuevo México, USA), con la idea de que la interoperabilidad de estos archivos de eprints era clave para aumentar su impacto entre la comunidad académica. Con ella se podrían federar varios archivos, intercambiar registros o realizar búsquedas en disciplinas relacionadas al mismo tiempo. Los participantes en esta reunión fueron especialistas en bibliotecas digitales, así como representantes de los principales archivos existentes:

- arXiv.org Considerado como el primer ejemplo de archivo de eprints. Fue fundado en 1991. Aunque comenzó como archivo de prepublicaciones ha evolucionado para incluir también artículos publicados en revistas tradicionales. Igualmente comenzó centrado en Física de Altas Energías pero ha incorporado otras disciplinas relacionadas como las Matemáticas, Informática, etc.
- CogPrints <http://cogprints.soton.ac.uk> Proyecto de la University of Southampton en el Reino Unido. Es una exportación del modelo de arXiv.org al campo de la Psicología y disciplinas relacionadas.
- NCSTRL <http://www.ncstrl.org> Es la Networked Computer Science Technical Reference Library, una colección de informes y documentos en Informática. Está basado en una arquitectura distribuida en la que los documentos son almacenados en archivos distribuidos y son hechos disponibles a través de servicios que se comunican utilizando el protocolo Dienst.
- ND LTD <http://www.ndtl.org> Es la Networked Digital Library of Theses and Dissertations. Su objetivo es construir una biblioteca digital de tesis en formato electrónico cuyos autores sean estudiantes de las instituciones miembros.
- RePEc <http://repec.org> Son las siglas de Research Papers in Economics. También se basa en un modelo distribuido. Proporciona a los autores la opción de remitir sus documentos de trabajo a un archivo local de su propia institución o, si no existe uno, al EconWPA <http://econwpa.wustl.edu>, un archivo mantenido por la Washington University at Saint Louis siguiendo el modelo de arXiv.org. Todos los archivos siguen el denominado Protocolo de Guildford que garantiza la interoperabilidad entre los archivos y los servicios a los usuarios finales.

La interoperabilidad de los archivos tiene varias facetas como son por ejemplo sistemas de identificación comunes, formatos de metadatos, modelos de

documentos o protocolos. Los participantes establecieron que una solución minimalista era imprescindible si se quería alcanzar una amplia adopción entre la comunidad de proveedores de eprints. La solución adoptada fue la recolección de metadatos (metadata harvesting). Esta solución permite a los proveedores de eprints exponer sus metadatos a través de una interfaz, con el objeto de que la misma pueda ser utilizada como la base para el desarrollo de servicios de valor añadido.

El resultado de la reunión fue un conjunto de acuerdos técnicos y organizativos conocidos como la Convención de Santa Fe. Los aspectos técnicos incluían tres puntos fundamentales: un formato para los metadatos, un protocolo basado en el antiguo Dients y un sistema de identificación.

Tras hacer públicos los resultados de la reunión, en Febrero de 2000, quedó claro que había un interés en esta iniciativa más allá de las comunidades de eprints. En principio bibliotecarios y museólogos se mostraron interesados en descubrir formas de hacer visibles a los motores de búsqueda en internet partes de las colecciones de bibliotecas y museos. Estas necesidades se expresaron en una serie de reuniones celebradas en el contexto de las principales jornadas sobre bibliotecas digitales celebradas tanto en USA como en Europa. Respondiendo a este amplio interés se procedió a la reconsideración de las decisiones tomadas en Santa Fe. Así se decidió ampliar el objeto de trabajo más allá de los eprints para incluir disciplinas que no tuvieran este tipo de documentación. Los aspectos técnicos aplicables exclusivamente a eprints fueron reconsiderados. Además la credibilidad del esfuerzo era incierta debido a la falta de una estructura organizacional. Los profesionales son lógicamente reacios a adoptar estándares cuando los responsables de promoción y mantenimiento de los mismos son cuestionables.

El último punto, credibilidad, fue el primero en tratarse y así en Agosto de 2000, la Digital Library Federation y la Coalition of Networked Information de los USA anunciaron que ofrecerían el soporte de su organización a la iniciativa. A partir de este momento comenzaron a funcionar dos comités, uno de gestión y otro técnico, que se encargarán de la coordinación de la iniciativa.

Las especificaciones revisadas fueron hechas públicas en Enero de 2001 con la publicación del Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH) 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.

Desde ese momento la implementación del protocolo comenzó, y aparecieron las primeras instituciones que lo utilizaron para poner en Internet sus metadatos. En su implementación, el OAI-PMH es una tecnología que sigue lo que Sapiro y Varian (Sapiro, 1999) denominan efectos de red, la adopción inicial es lenta y progresiva pero la respuesta positiva a la misma aumenta de forma dramática la tasa de adopción. Esto se ha cumplido en los dos años que lleva funcionando el

protocolo. Ya son más de 100 las instituciones que han creado archivos abiertos, el número de servicios basados en la utilización de la información almacenada en los anteriores no ha parado de crecer tanto en número como en calidad de los valores añadidos que ofrecen. En estos momentos hay registrados en el servidor de OAI unos 12 servicios. Igualmente han aparecido toda una serie de herramientas de software destinadas a facilitar la creación y mantenimiento de archivos. Han sido muchos los proyectos de investigación que se han concedido durante el pasado año para estudiar la aplicación del protocolo y temas relacionados. Así en USA está por ejemplo la Metadata Harvesting Initiative de la Fundación Mellon en el seno de la que se han financiado cuatro proyectos por valor de 1.5 millones de \$ con objeto de crear servicios basados en OAI-PMH. La National Science Digital Library, un proyecto de la National Science Foundation tiene como objeto la creación de lo que será la mayor biblioteca digital hasta el momento. Ha adoptado el protocolo como base para la comunicación de metadatos entre los participantes. En Europa se han financiado proyectos por parte de la UE como por ejemplo el Open Archives Forum cuyo objeto es la creación de una comunidad de interés en OAI en Europa por medio de la organización de jornadas y actividades de soporte a la implementación de archivos y servicios.

Inmediatamente después de la difusión de la versión 1 comenzó el trabajo del comité técnico para tratar los problemas de definición o funcionalidad que se fueran descubriendo. Ese trabajo desembocó en la elaboración de la versión 2 del protocolo anunciada en Junio de 2002. Los principales cambios que se introdujeron fueron relacionados con la clarificación de ambigüedades o mejores medios para expresar las funcionalidades existentes. Es decir, no se introdujeron cambios sustanciales.

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. Se espera que éste se convierta en una parte integral del trabajo en bibliotecas digitales. De hecho, en los pasados meses se ha pasado de hablar fundamentalmente del protocolo mismo, a hablar de proyectos en los que se usa el protocolo, y después a hablar de proyectos sin ni siquiera mencionar el protocolo. Otro área de interés son los formatos de metadatos, básicamente determinar si cumple su función el sistema utilizado actualmente (Dublin Core no calificado). También se estudiará la utilidad del protocolo más allá de la descripción de recursos, en cuestiones como certificación, estadísticas de uso, datos sobre citas, etc. Un aspecto que merece especial atención es una vuelta a la misión original de OAI, los eprints. Para ello se trabajará en un perfil de OAI-PMH para este tipo de documentos.

3.- El protocolo OAI-PMH

Los participantes en Santa Fe tomaron una decisión clave en cuanto a la arquitectura del protocolo. Adoptaron un modelo que rechazaba la búsqueda distribuida (como hace el Z39.50) a favor de simplemente tener servidores proporcionado metadatos, sujetos sólo a criterios de alcance bastante simples,

tales como por ejemplo proporcionar todos los registros añadidos o cambiados desde una fecha específica.

No vamos a entrar aquí en una descripción técnica del protocolo, pero 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:

- **GetRecord.** Utilizado para recuperar un registro concreto. Necesita dos argumentos: identificador del registro pedido y especificación del formato bibliográfico en que se debe devolver.
- **Identify.** Utilizado para recuperar información sobre el servidor: nombre, versión del protocolo que utiliza, dirección del administrador, etc.
- **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.
- **ListRecords.** Igual que el anterior pero recupera los registros completos.
- **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.
- **ListMetadataFormats.** Devuelve la lista de formatos bibliográficos que utiliza el servidor.

El protocolo soporta múltiples formatos para expresar los metadatos, no obstante requiere que todos los servidores ofrezcan los registros utilizando Dublin Core 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/2.0/>. Un ejemplo de petición y respuesta sería:

Petición:

```
http://an.oa.org/OAI-script?  
verb=GetRecord&identifier=oai:arXiv:hep-th/9901001&metadataPrefix=oai_dc
```

Respuesta:

```
<?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>2002-05-01T19:20:30Z</responseDate>  
<request verb="GetRecord" identifier="oai:arXiv:hep-th/9901001"  
  metadataPrefix="oai_dc">http://an.oa.org/OAI-script</request>  
<GetRecord>  
  <record>  
    <header>  
      <identifier>oai:arXiv:cs/0112017</identifier>  
      <datestamp>2001-12-14</datestamp>  
      <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>
```

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.

4.- Proveedores de datos y de servicios

De la sección anterior se desprende que la arquitectura de OAI-PMH se basa en clientes y servidores. Los primeros son los archivos que proporcionan la información y los segundos son los recolectores o servicios que toman los datos, con el objetivo de incorporarles algún valor añadido y presentarlos a los usuarios finales.

Desde Enero de 2001 se ha mantenido un registro de todos los archivos que han implementado el protocolo OAI-PMH. El registro no es obligatorio por lo que se supone que son muchos los archivos que existen y no se han registrado. Desde esa fecha el incremento del número de archivos ha sido constante llegando hasta 45 en la actualidad (sólo los que han adoptado la versión 2). Entre ellos tenemos como más destacables: arXiv.org junto con el resto de iniciativas que mencionamos en el punto 2, CERN que recoge informes y prepublicaciones en el área de Física o Citebase que proporciona datos sobre citas recibidas por los eprints almacenados en varios archivos.

Habría que destacar también los archivos que se están abriendo en el área de Biblioteconomía y Documentación. En estos momentos hay tres disponibles, aunque solamente uno de ellos está registrado:

- **@rchiveSIC** <<http://archivesic.ccsd.cnrs.fr/>> es un proyecto de colaboración entre varias instituciones francesas (universidades y centros de investigación como el CNRS). En estos momentos almacena unos 80 documentos, la mayor parte de ellos en francés. Incluye documentos de áreas relacionadas como Museología.

- **DLIST** (Digital Library of Information Science and Technology) <<http://dlist.sir.arizona.edu/>>. Es un archivo creado por la School of Information Resources and Library Science y Arizona Health Sciences Library (University of Arizona). Almacena más de 100 documentos. Su objetivo es recoger todo tipo de documentos científicos en Documentación pero con dos áreas temáticas de mayor énfasis: materiales educativos y bibliometría. Solamente aceptan documentos en Inglés.
- **E-LIS** (Eprints in Library and Information Science) <http://eprints.rclis.org>. Es el proyecto más reciente dado que aun no se ha hecho público. Es un esfuerzo internacional para crear un archivo multinacional y multilingüe de documentos científicos en las áreas de Biblioteconomía y Documentación. Ha sido financiado parcialmente por el Ministerio de Educación.

Igualmente se mantiene un registro de los servicios creados utilizando los datos proporcionados por los anteriores. Tampoco aquí es obligatorio el registro por lo que es imposible saber cuántos servicios existen realmente en la actualidad. Sin duda son muchos más que los 12 que aparecen en el registro oficial. Como más destacables:

- **ARC** <http://arc.cs.odu.edu/> Es un servicio experimental creado con el objetivo de investigar temas relacionados con la recolección de metadatos siguiendo el protocolo OAI-PMH y cómo hacerlos disponibles a los usuarios. Más que un servicio en sí mismo es un software que podría ser utilizado por instituciones que quieran crear sus propios servicios. El código fuente está disponible en la red de forma gratuita. Ha sido desarrollado por el Digital Library Research Group de la Old Dominion University.
- **OAister** <http://oaister.umdl.umich.edu/> Es un proyecto financiado por la fundación Mellon con el objetivo de crear una amplia colección de recursos digitales gratuitos, útiles y que previamente eran de muy difícil acceso y ponerla al alcance de cualquier usuario de la forma más sencilla posible. Es decir, trata de sacar a la luz colecciones que antes eran invisibles. Todos los recursos tienen el texto completo disponible en la red de forma que siempre se pueda llegar a los contenidos. Recoge datos de todos los archivos conocidos. En total 122 archivos con más de un millón de registros.
- **Perseus** <http://www.perseus.tufts.edu/> Es una biblioteca digital especializada en Humanidades. Su servicio OAI también recupera datos de todos los servicios conocidos. Está financiada por la National Science Foundation en los USA.

- **Cyclades** <http://www.ercim.org/cyclades> Es un proyecto financiado por la Unión Europea. Su objetivo no está directamente relacionado con OAI ya que es crear un marco de colaboración entre los investigadores de los centros que participan en el proyecto. Intenta fomentar la colaboración entre los mismos, emitir recomendaciones y crear servicios personalizados. Otros servicios también financiados por la UE son ICite <http://icite.sissa.it/> y TORII ambos dentro del proyecto: Tools for Innovative Publishing in Science <http://tips.sissa.it/>.

En resumen, los proveedores de servicios están proliferando y cada vez están proporcionando servicios más sofisticados. Se puede decir que existe un mercado donde los servicios pueden competir, por ejemplo existen hasta 10 interfaces diferentes a los datos proporcionados por arXiv.org, cada uno de ellos con unas características diferenciadoras.

5.- ¿Cómo crear un archivo abierto?

Como hemos visto en el punto tres OAI-PMH solamente es una interfaz sumamente sencilla para acceder a la información bibliográfica disponible en un archivo o repositorio. Por lo tanto cualquiera puede realizar una implementación del mismo para poner a disposición de la comunidad Internet los datos que hasta ahora estaban escondidos en bases de datos o catálogos. En este sentido bastaría con disponer de un servidor web y un programa CGI (en Perl o PHP) que recibiera las peticiones OAI-PMH, interrogara nuestra base de datos y devolviera la respuesta.

Por otro lado la iniciativa OAI nace del movimiento de eprints cuyo objetivo es el de poner a disposición del público documentos en formato electrónico vía repositorios de documentos. Para facilitar esta tarea han aparecido una serie de programas que permiten a cualquier institución (universidad o centro de investigación) crear su propio archivo al tiempo que hacerlo compatible con OAI-PMH. Ejemplos de algunos de estos programas son:

- **Eprints.** <http://www.eprints.org> El más popular de todos ya que está siendo utilizado en más de 30 instituciones. Es un software desarrollado en el seno del Open Citation Project dirigido por Stevan Harnad en la Universidad de Southampton (UK). Está diseñado con el objetivo de ser fácil, rápido de instalación y gratuito. Eprints se distribuye bajo la licencia GNU, lo cual significa que el código fuente es accesible y modificable por cualquier programador, con la condición que las modificaciones se hagan también accesibles públicamente. Eprints puede funcionar en cualquier ordenador con sistema operativo Linux.. Las principales características de Eprints son:
 - Facilidad de instalación y configuración. Este es un objetivo que no se ha alcanzado aún. Si bien el proceso está automatizado en gran parte se necesitan conocimientos técnicos para llevarlo a cabo. Es difícil que

en el estado actual pueda ser instalado por investigadores sin asistencia de administradores de sistemas.

- Permite almacenar documentos en cualquier formato, así como almacenar un mismo documento en varios formatos. La carga de ficheros se realiza mediante una interfaz web muy sencilla.
- Permite utilizar cualquier formato para almacenar la información bibliográfica sobre los documentos.
- Permite que los usuarios se registren como lectores o como autores para obtener un mayor aprovechamiento de sus funciones.
- **Dspace** <http://www.dspace.org> Es el más reciente de los programas ya que se anunció en Noviembre de 2002. Está desarrollado por la empresa HP y las bibliotecas del MIT. También es un software con las fuentes disponibles públicamente (open source) cuyo objetivo es permitir a una organización almacenar, describir y gestionar documentos electrónicos, distribuirlos a través del web a través de un sistema de búsqueda y recuperación de la información y finalmente proporcionar un sistema para el almacenamiento a largo tiempo de los documentos. Está pensado para funcionar en varias plataformas y soporta la versión 2 de OAI-PMH.
- **CDSware**. <http://cdsware.cern.ch> Su primera versión se hizo pública en Agosto de 2002. Está desarrollado, mantenido y utilizado por el CERN Document Server de Ginebra. Su objeto es permitir a una institución crear su propio servidor de eprints, catálogos de sus fondos o un sistema documental a través del web. Es compatible con OAI-PMH. Lo más destacado es que utiliza el formato MARC 21 para almacenar los registros bibliográficos. Igual que en los casos anteriores es un software gratuito distribuido bajo la licencia GPL (General Public License). En la biblioteca del CERN se utiliza para gestionar más de 350 colecciones formadas por más de 565.000 registros con unos 220.000 de ellos representando documentos a texto completo. El incremento se sitúa en torno a los 1000 registros semanales.
- **VT ETD-db**. <http://scholar.lib.vt.edu/ETD-db/> Creado en el Virginia Polytechnic Institute y la State University (USA). Como en los casos anteriores se trata de un software para crear depósitos de documentos. Está siendo usado por la Université Catholique de Louvain. Trata de proporcionar una interfaz para que los usuarios puedan introducir y gestionar información bibliográfica relativa a colecciones de tesis en formato electrónico.

6.- Para saber más

Shapiro, C; Varian, H. R. Information rules : a strategic guide to the network economy. Boston : Harvard Business School Press, 1999.

Van de Sompel, Herbert ; Lagoze, Carl. "The Santa Fe convention of the Open Archives Initiative". En: D-Lib, 2000, Febrero, v. 6, n. 2.

<http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>

Van de Sompel, Herbert ; Lagoze, Carl. "The Open Archives Initiative: Building a low-barrier interoperability framework". En: JCDL, 2001.

<http://www.openarchives.org/documents/oai.pdf>

Van de Sompel, Herbert ; Lagoze, Carl (ed.) "The Open Archives Initiative Protocol for Metadata Harvesting". Consultado el: 10-01-2003

<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>

Van de Sompel, Herbert ; Lagoze, Carl. "Notes from the Interoperability Front: A Progress Report on the Open Archives Initiative". En: ECDL, 2002.

<http://www.openarchives.org/documents/ecdl-oai.pdf>

Harnad, Stevan. "Free at Last: The Future of Peer-Reviewed Journals," En: D-Lib, 1999, v. 5, n. 12.

<http://www.dlib.org/dlib/december99/12harnad.html>