394

# RECOLECTOR OAI Y COLECCIONES DIGITALES

**Amparo Candelas Arnao**
Jefe de Servicio de Sistemas Informáticos
Subdirección General de Tecnologías y Sistemas de la Información
Ministerio de Cultura

**Mª Luisa Martínez-Conde**
Consejera
Subdirección General de Coordinación Bibliotecaria (DGLAB)
Ministerio de Cultura

**Amparo Candelas Arnao**
**Mª Luisa Martínez-Conde**

## Palabras clave

*Open Archive Iniciative, protocolos de intercambio de datos, recursos electrónicos, registro bibliográfico, digitalización*

## Resumen de su Comunicación

*El Ministerio de Cultura ha puesto en producción la aplicación Directorio y recolector de recursos digitales, que pone a disposición de los ciudadanos dos conjuntos de información: un directorio de proyectos de digitalización en España, incluyendo información académica editada en el formato e-prints, y un recolector OAI de recursos digitales que permite consultar de forma conjunta los registros incluidos en esos proyectos así como acceder al propio documento digitalizado.*

**Amparo Candelas Arnao**
**Mª Luisa Martínez-Conde**

# RECOLECTOR OAI Y COLECCIONES DIGITALES

El Ministerio de Cultura ha puesto en producción el servicio Directorio y recolector de recursos digitales, http://www.mcu.es/roai, que pone a disposición de los ciudadanos dos conjuntos de información: un directorio de proyectos de digitalización que recoge los proyectos e iniciativas de digitalización existentes en España, incluyendo información académica editada en el formato e-prints, y un recolector de recursos digitales que permite consultar de forma conjunta los registros incluidos en esos proyectos así como acceder al propio documento digitalizado. La aplicación utiliza técnicas basadas en los metadatos normalizados internacionalmente, así como la Iniciativa de Archivos Abiertos (Open Archive Iniciative).

Este servicio es resultado del trabajo del Grupo de Colecciones Digitales de las Comunidades Autónomas y el Ministerio de Cultura. Junto con la iniciativa de la Biblioteca Virtual de Prensa Histórica, http://www.mcu.es/prensahistorica, realizada en cooperación con las Comunidades Autónomas, el Ministerio de Cultura pretende contribuir a la Biblioteca Digital Europea y, además, crear los instrumentos necesarios para que otras colecciones digitales resultado de proyectos emprendidos en España puedan alcanzar la máxima visibilidad en la red y formar parte, asimismo de la futura Biblioteca Europea creando un instrumento de coordinación de los distintos proyectos que permitirá establecer planes de actuación cooperativos evitando solapamientos y optimizando así las inversiones que se lleven a cabo para aumentar el patrimonio digital español, preservándolo y difundiéndolo.

## 1. Los estándares del registro bibliográfico

Lo más probable al buscar información en INTERNET es experimentar la frustración al encontrar cientos o miles de ocurrencias con la imposibilidad de refinar o de hacer una búsqueda más precisa. El interés mundial por las normas de metadatos ha aumentado con el crecimiento de la publicación electrónica y las bibliotecas digitales, sobrecarga que resulta de las grandes cantidades de datos digitales disponibles en línea. El término “meta” viene de una palabra griega que significa “junto a, con, después, siguiente”. Los metadatos no son más que datos sobre los datos. No es algo reciente ni propio de la informática o INTERNET, los metadatos han estado con nosotros desde que el primer bibliotecario hizo una lista de los documentos que tenía en un estante. Actualmente es el término más usado para referirse a la información descriptiva sobre recursos de la Web.

Un conjunto de registros de metadatos que describen un libro u otra publicación contendrá elementos con la información del autor, el titulo, la fecha de publicación, y la signatura topográfica. ISO 2709, ISBD, ISSN, MAR21, MARC XML Schema, Z39.50, ... son parte de una extensa lista de las normas internacionales que rigen la catalogación e  intercambio de información bibliográfica.

En 1969, el Comité de Catalogación de la IFLA promovió la Reunión Internacional de Expertos en Catalogación (International Meeting of Cataloguing Experts), que resolvió el establecimiento de normas que regularan la forma y el contenido de las descripciones bibliográficas, elaborando los mecanismos que permitieran la cooperación y el intercambio de información bibliográfica. Publicada en 1971, la Descripción Bibliográfica Internacional Normalizada para Publicaciones Monográficas (ISBD(M)) fue la primera de las ISBD, que con una duración de casi treinta años, ha demostrado ser el esfuerzo más satisfactorio de la IFLA para promover la normalización en la catalogación.

ISO ha trabajado en la normalización de los formatos de la información automatizada. Aunque puedan establecerse reglas para conversión de uno a otro, el formato de almacenamiento de los datos por una aplicación no es un estándar, dando lugar a la incompatibilidad en el intercambio de los mismos. El formato ISO 2709, Information and Documentation - Format for Information Exchange, define la norma para transferencia de información en cintas, facilitando la comunicación entre sistemas.

Sin embargo, ISO 2709 no normaliza la compatibilidad entre las estructuras de datos, es decir, una aplicación no tiene que compartir la estructura de datos (campos) ni mucho menos los contenidos con otra. Es decir, ISO 2709 está pensado simplemente para intercambiar información.

Consecuentemente, el desarrollo de formatos ha exigido una normalización más amplia en cuanto a la aplicación de la norma ISO-2709 en relación con el diseño de la estructura. En la asignación de etiquetas, indicadores, limitadores, número y longitud de campos variables hay varios criterios como los empleados por la UNESCO con el Formato Común de Comunicación (FCC), la Biblioteca del Congreso de los Estados Unidos con el formato MARC y el formato de la CEPAL.

MARC es un formato creado a fines de los sesenta, pensado cuando todavía se hablaba de mecanización, no automatización, de bibliotecas. La estructura del registro MARC es una implementación de ISO 2709. MARC es un formato para cintas de datos, no para registro en línea, y por lo tanto podría haber sido abandonado hace ya mucho tiempo. La razón por la que sigue vigente es porque MARC no es sólo un formato de datos para registros informatizados sino un formato de información: no establece la manera en que deben ser extraídos los contenidos de cada campo, sino simplemente que contenidos son indispensables para una descripción bibliográfica.

El Formato MARC 21, contiene elementos de datos para los siguientes tipos de materiales: libros (ítems de texto de tipo monográfico, tales como libros encuadernados, electrónicos o microformas), seriadas (ítems de texto con un patrón recurrente de publicación, p.e. publicaciones periódicas, diarios, anuarios), archivos de computación (software, datos numéricos, multimedios en computación, sistemas o servicios en línea), mapas (todo tipo de materiales cartográficos, incluyendo mapas en hojas y globos en forma impresa, manuscritos, electrónicos, y microformas), música (partituras musicales impresas y manuscritas), registros sonoros (registros sonoros no musicales, y registros sonoros musicales.), material visual (imágenes y objetos, p.e., medios que se proyectan, películas, gráficos de dos dimensiones, objetos tridimensionales, objetos reales) y materiales mixtos (fundamentalmente colecciones de archivo y manuscritos con material de formas mixtas).

Con la necesidad del intercambio de información normalizada en formato XML, aparece MARC 21 XML Schema. El núcleo de MARC XML es simplemente un XML-Schema para contener datos MARC. Este formato puede ser usado como formato intermedio en transformaciones a Dublin Core o para validación de la información. Además, el MARC XML Schema no necesita ser modificado ante cambios en MARC21.
El modelo de metadatos Dublin Core (DC) o DCMI, es un esfuerzo internacional e interdisciplinar abocado a definir el conjunto de elementos básicos para describir los recursos electrónicos y facilitar su recuperación. El DC, surgido en 1995 en el seno del Online Computer Library Center, es hoy un esquema maduro de metainformación cuyo conjunto de elementos (DCMES) se ha formalizado, primero como norma ANSI/NISO Z39.85 en octubre de 2001, y recientemente, como estándar internacional ISO 15836-2003, desde el 8 de abril de este año.

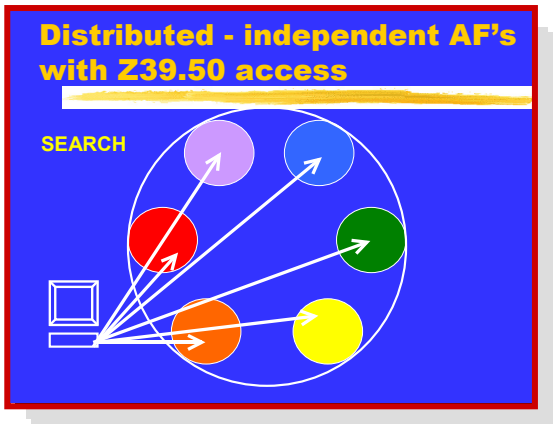
La norma del Dublin Core conlleva dos niveles: Simple y Cualificado. El Dublin Core Simple conlleva quince elementos; el Dublin Core Cualificado conlleva un elemento adicional, la audiencia [Audience], así como un grupo de elementos de matización, denominados por ello cualificadores, que refinan la semántica de los elementos de tal forma que pueden ser útiles para la recuperación/localización de recursos en Internet [resource discovery].

## 2. Los servidores de información bibliográfica

Además del típico formulario HTML de búsqueda que podemos encontrar en la página Internet de una base de datos bibliográfica, existen normativas especificas para que un servidor ofrezca consulta e intercambio de información on-line. La búsqueda a través de servicios HTML, está limitada por el propio buscador, la

dificultad de buscar hasta un determinado nivel jerárquico en páginas estáticas HTML y la aún más difícil búsqueda en las páginas dinámicas y bases de datos (deep web). Surge por tanto la necesidad de utilizar metadatos y protocolos que faciliten su recolección (metadata harvesting).

Desde el año 1995, uno de los protocolos ampliamente difundido en el sector bibliotecario es el Z39.50. El protocolo Z39.50 es un protocolo a nivel de aplicación para búsqueda y recuperación de información en base de datos bibliográficas. Este y otros servicios son definidos por la norma ANSI Z39.50 e ISO 23950.




Para el acceso a un servidor Z39.50 es necesario disponer de conexión a red TCP/IP y de un programa cliente Z39.50, que se suministra con los principales sistemas de gestión bibliotecaria o puede adquirirse independientemente. Para este tipo de conexión, se requerirá opcionalmente la identificación del usuario.
Uno de los inconvenientes del protocolo Z39.50 es que el usuario se ve en la necesidad de lanzar una consulta individualizada a cada uno de los servidores. Además, el usuario debe conocer la existencia del servidor y como conectar a él. Por ello, han surgido iniciativas de información que a través de una página web recogen listas con la los datos de dichos servidores.

Otro problema es el pago de las licencias dependiendo del número de conexiones que el servidor pretenda atender, es decir, el número de clientes Z39.50 que puedan consultar simultáneamente. Cuando todas las conexiones disponibles están en uso, el próximo cliente Z39.50 que quiera acceder obtendrá una negación de servicio como respuesta.

En Octubre de 1999, tuvo lugar un congreso en Santa Fe para decidir mecanismos que alentaran el desarrollo de soluciones E-Print. El grupo de participantes en este congreso estaba de acuerdo en que la interoperatividad de INTERNET proporcionaba nuevas oportunidades para la diseminación de información. Como resultado de este congreso se creó la Open Archives Initiative (OAI) y fué el comienzo del establecimiento de un marco para facilitar la federación de servidores de contenidos en la Web.

El nombre Open Archives Initiative refleja el origen E-Prints donde el término archivo es generalmente aceptado como un sinónimo para un repositorio documentos educativos, pero OAI usa el termino “archivo” en un sentido más amplio: como un repositorio para información almacenada. El original enfoque en E-Prints fué ampliado para cubrir provedores de contenido de muchos dominios. De esta forma, la Open Archives Initiative (OAI) desarrolla y promueve soluciones de interoperatividad que ayudan a facilitar la diseminación eficiente de contenidos.

La implantación de este nuevo estándar está ya tan difundida que actualmete Google está usando OAI-PMH para recolectar información del National Library of Australia (NLA) Digital Object Repository.

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

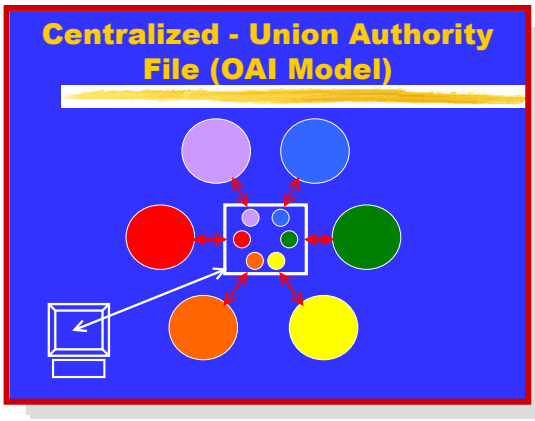
El protocolo para recolección de metadatos de la Iniciativa de Archivos Abiertos (Open Archives Initiative Protocol for Metadata Harvesting), referido como OAI-PMH en el resto de este artículo, proporciona un marco de interoperatividad independiente de la aplicación basado en recolección de metadatos (metadata harvesting). Existen dos clases de participantes en el marco OAI-PMH:

• Proveedores de datos, sistemas que soportan el OAI-PMH para exponer los metadatos de la información de los recursos que administran; un proveedor de datos puede escoger registrarse y así publicitar el hecho de haber adoptado el protocolo OAI-PMH.

• Proveedores de servicio, usan recoleccion de metadatos via el OAI-PMH como base para construir servicios de valor añadido; un proveedor de servicio puede escoger registrarse y así publicitar su existencia.

Un recolector es una aplicación cliente que resuelve peticiones OAI-PMH. A recolector está soportado por un proveedor de servicio con el objetivo de obtener metadatos de los repositorios. Un repositorio es un servidor accesible por la red que puede procesar las 6 peticiones OAI-PMH descritas por la norma. Un repositorio es gestionado por un proveedor de datos para exponer los metadatos para su recolección. Para permitir varias configuraciones del repositorio, el OAI-PMH distingue entre distintas entidades relacionadas con el metadato al que se accede por OAI-PMH.

Profundizar en el protocolo OAI-PMH escapa por su amplitud al objetivo de este árticulo. Resumiendo mucho, entre las carácterísticas que proporciona un proveedor de datos OAI debemos destacar:




• Permite la transmisión de metadatos en Internet
• Los metadatos están en formato Dublin Core / ISO 15836
• El usuario sólo necesita un navegador web para realizar una consulta
• La comunicación se realiza sobre el protocolo HTTP
• Las respuestas están codificadas en XML

Y en cuanto a los recolectores OAI, es decir, los metabuscadores de proveedores de datos OAI, podemos destacar:

• El usuario consulta un único recolector obteniendo la información de múltiples repositorios

• El acceso a los repositorios y los recursos es inmediata mediante hipervínculos incluidos en la respuesta

Como ejemplo de una consulta OAI, veamos la respuesta a una pregunta de identificación del servicio (http://www.mcu.es/prensahistorica/OAIHandler?verb=Identify) al acceder a la Biblioteca Virtual de Prensa Histórica, que se ofrece a través del Ministerio de Cultura:

<?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>2006-03-13T16:14:53Z</responseDate>
<request verb="Identify">http://www.mcu.es/prensahistorica/OAIHandler</request>
- <Identify>
<repositoryName>Biblioteca Virtual de Prensa Histórica</repositoryName>
<baseURL>http://www.mcu.es/prensahistorica/OAIHandler</baseURL>

<protocolVersion>2.0</protocolVersion>
<adminEmail>mailto:prensahistorica@dglab.mcu.es</adminEmail>
<earliestDatestamp>2000-01-01</earliestDatestamp>
<deletedRecord>no</deletedRecord>
<granularity>YYYY-MM-DD</granularity>
<compression>gzip</compression>
<compression>deflate</compression>
+ <description>
+ <description>
- <description>
- <eprints xmlns="http://www.openarchives.org/OAI/1.1/eprints" xsi:schemaLocation="http://www.openarchives.org/OAI/1.1/eprints http://www.openarchives.org/OAI/1.1/eprints.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
- <content>
<text>La Biblioteca Virtual de Prensa Histórica está formada por la versión digital de publicaciones periódicas españolas, fundamentalmente prensa, editadas en los siglos XVIII, XIX y XX y conservadas en las Bibliotecas Públicas del Estado y otras instituciones de memoria.</text>
<URL>http://www.mcu.es/prensahistorica</URL>
</content>
- <metadataPolicy>
<text>Los metadatos de la Biblioteca Virtual de Prensa Histórica son de libre acceso y utilización, debiéndose citar la procedencia de los mismos para cualquier tipo de intercambio. No se realizará ninguna venta o intercambio comercial de estos datos.Las copias digitales de la Biblioteca Virtual de Prensa Histórica son de libre acceso y con fines de consulta e investigación. Cualquier reutilización de las mismas con estos fines debe incluir la marca de agua con el copyright del Ministerio de Cultura. No se realizará ninguna venta o intercambio comercial con estos datos.</text>
</metadataPolicy>
- <submissionPolicy>
<text>Las instituciones de memoria participantes en el programa envían anualmente los títulos seleccionados para su digitalización y son incorporadas tanto en la base de datos como en el repositorio elaborado según las directrices OAI.</text>
</submissionPolicy>
<comment>El servicio es ofrecido por la Subdirección General de Coordinación Bibliotecaria, de la Dirección General del Libro, Archivos y Bibliotecas, y la Subdirección General de Tratamiento de la Información del Ministerio de Cultura, en la dirección http://www.mcu.es/prensahistorica</comment>
</eprints>
</description>
+ <description>
</Identify>
</OAI-PMH>

Y la respuesta a una consulta de un recurso concreto (http://www.mcu.es/prensahistorica/ OAIHandler ?verb=GetRecord&identifier=oai:prensahistorica.mcu.es:1226&metadataPrefix= oai_dc):

<?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>2006-03-13T16:27:07Z</responseDate>
<request identifier="oai:prensahistorica.mcu.es:1226" metadataPrefix="oai_dc" verb="GetRecord">http://www.mcu.es/prensahistorica/OAIHandler</request>
- <GetRecord>

- <record>
- <header>
<identifier>oai:prensahistorica.mcu.es:1226</identifier>
<datestamp>2005-02-07</datestamp>
</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>La lealtad [Recurso electrónico]</dc:title>
<dc:creator>Tipografía de Mengual y Muntaner (Palma de Mallorca) , imp.</dc:creator>
<dc:type>text</dc:type>
<dc:publisher>[S.l. : s.n.]</dc:publisher>
<dc:date>, 1913-</dc:date>
<dc:language>spa</dc:language>
<dc:description>Copia digital</dc:description>
<dc:description>Descripción basada en: Año I, n. 6 (jun. 1913)</dc:description>
<dc:coverage>España -Illes Balears -Illes Balears -Palma de Mallorca</dc:coverage> <dc:identifier>http://www.mcu.es/prensahistorica/es/consulta/registro.cmd?id=1226</dc:identifier>
</oai_dc:dc>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

## 4. El servicio Directorio y recolector de recursos digitales

El grupo de trabajo creado en el marco de las Jornadas de Cooperación Bibliotecaria del Ministerio de Cultura y las Comunidades Autónomas estableció unos objetivos sobre el nuevo servicio que se debía proporcionar al ciudadano en relación a los proyectos de digitalización en España:

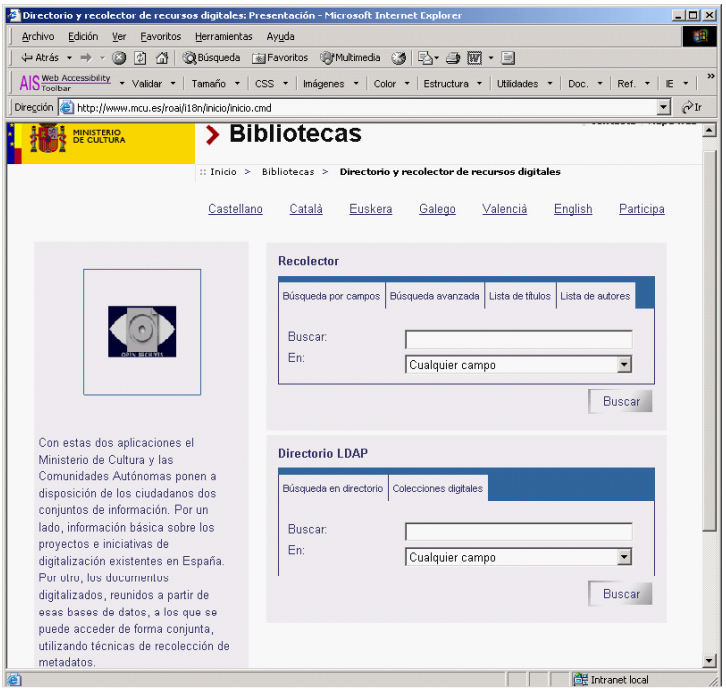
- Crear un recolector OAI de recursos digitales en España
- Crear un directorio LDAP de colecciones digitales en España
- Ofrecer una interfaz web de búsqueda en ambos servicios

Fruto de estos objetivos es el desarrollo de la aplicación Directorio y recolector de recursos digitales que el Ministerio de Cultura acaba de poner en INTERNET. Esta aplicación consta de recursos propios para los administradores y gestores del servicio, es decir, restringido al grupo de trabajo y a la Subdirección General de Coordinación Bibliotecaria (SGCB).

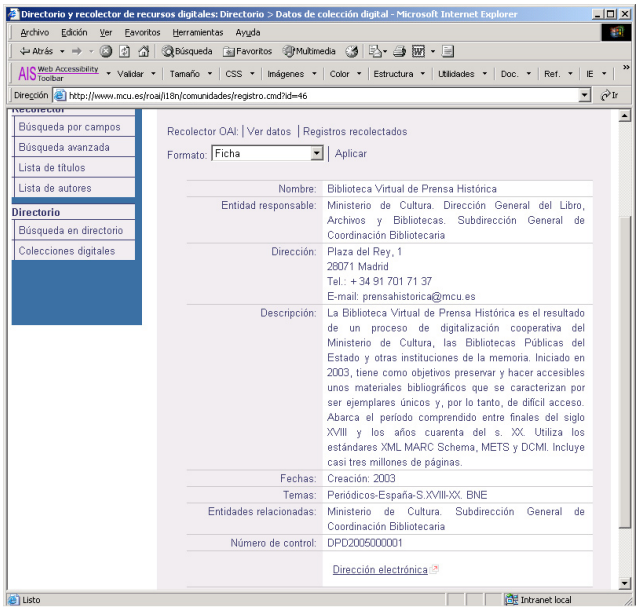
Este grupo es el que crea las colecciones digitales que serán publicadas a través de un directorio LDAP. Los campos y datos que se pueden definir siguen la estructura recogidos en el estándar MARC para Community Information.

Por otra parte, la SGCB es quien define los servidores proveedores de datos OAI para dichas colecciones y establece los parámetros para la recolección.
Además, la aplicación ofrece un servicio abierto al público para la consulta del directorio LDAP de colecciones digitales y para consulta del repositorio OAI.

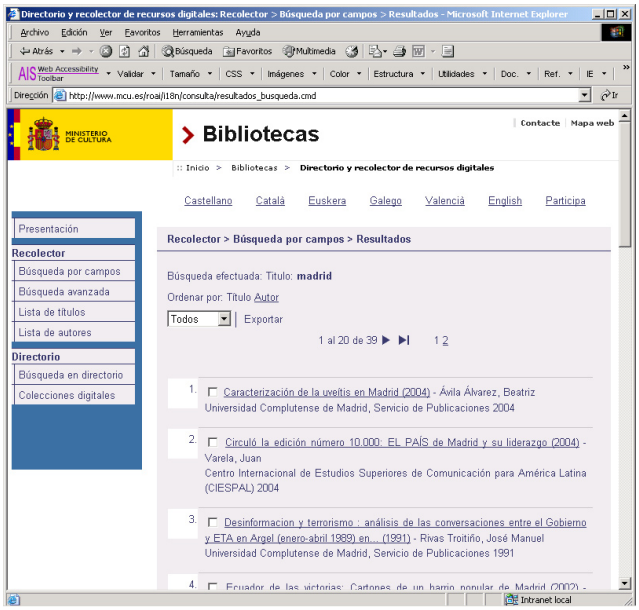
Para filtrar los datos de todas las colecciones existentes en el sistema se permite hacer búsqueda sobre ellas. Esto se hace mediante la introducción de una palabra o palabras clave en los distintos campos.
Los resultados obtenidos de la búsqueda se mostrarán ordenados por el nombre del proyecto. Al ir al enlace correspondiente a un proyecto se accede a ver todos sus campos.




Se podrá cambiar el formato de presentación de estos datos entre tres posibles: Ficha, MARC Etiquetado, XML-MARC Schema. En todas las colecciones digitales al ver sus detalles se incluye un enlace hacia la página externa al sistema correspondiente a dicha colección. Dependiendo del perfil asignado que posea el usuario autenticado tendrá mas o menos opciones de control sobre la colección.

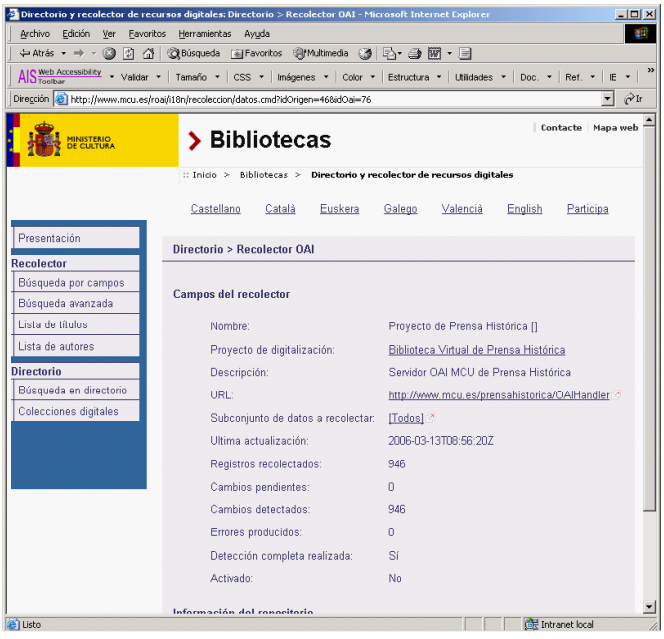

Las consultas relacionadas con todos los registros de metadatos recolectados de las diferentes colecciones digitales programadas, es decir, de las obras en el repositorio, puede hacerse limitando ciertos campos que coincidan con una cadena de caracteres introducida. Tras realizar la búsqueda el sistema se nos mostrarán los resultados en caso de haber alguna entrada coincidente. Estos resultados consisten en una lista con un resumen de la obra (título, autor, año).




Estos resultados obtenidos pueden ser ordenados por los campos Título o Autor y pueden ser exportados todos o solo algunos registros del resultado a un fichero en diferentes formatos: ISBD, Ibermarc, Onix, etc. y el juego de caracteres en el que estará escrito.
Si accedemos a los datos propios del registro descargado del repositorio además de ofrecernos un enlace a la copia digital del mismo. Estos datos se pueden ver en diferentes formatos de presentación (ISBD, Onix, Ibermarc Etiquetado, etc.). También se puede exportar la ficha usando el enlace Exportar.
Por último, incidir en la necesidad de que el servicio se ha implantado cumpliendo las normas de accesibilidad nivel AA y ofreciendo sistema multilingüe en 6 lenguas: castellano, gallego, catalán, valenciano, euskera e ingles.

## Proveedores de datos OAI

Para poder crear un proveedor de datos OAI, se deberá conocer una URL válida al almacén de metadatos correspondiente al proyecto de digitalización. El resto de los datos no son necesarios inicialmente. El campo Subconjunto de datos a recolectar permite discriminar la recolección por uno de los subconjuntos de datos que posea el almacén de metadatos. Si no se indica nada considerará que debe coger todos. Si se desea que la actualización sea gradual para no sobrecargar el recolector se puede indicar un Límite de registros por sesión que serán actualizados. Por último, se puede suspender la actividad de recolección si la casilla Activado se encuentra sin marcar.




Cuando se ha llevado a cabo una recolección aparecerá información adicional sobre el resultado del proceso:



Volver

• Ultima Actualización: Indica la fecha en la que se produjo la última actualización de los datos.

• Registros Recolectados: Son los registros que se han recolectado de otros proyectos mediante este recolector.

• Cambios pendientes: Son el número de cambios que aún no se han descargado y aplicado del repositorio de metadatos.

• Cambios detectados: El número de cambios que se han detectado hasta ahora en el almacén de datos del proyecto al que corresponde el recolector, esto incluye tanto nuevos registros como cambios en registros existentes.

• Errores producidos: Número de errores producidos mientras se intentaban obtener los datos del almacén.

• Detección completa realizada: Indica si se han detectado al menos una vez todos los cambios en el repositorio, aunque estos no hayan sido descargados.

## Arquitectura de la aplicación

El sistema de software propuesto se compone de diferentes repositorios de datos y procesos, como se puede observar en el diagrama de la siguiente figura.

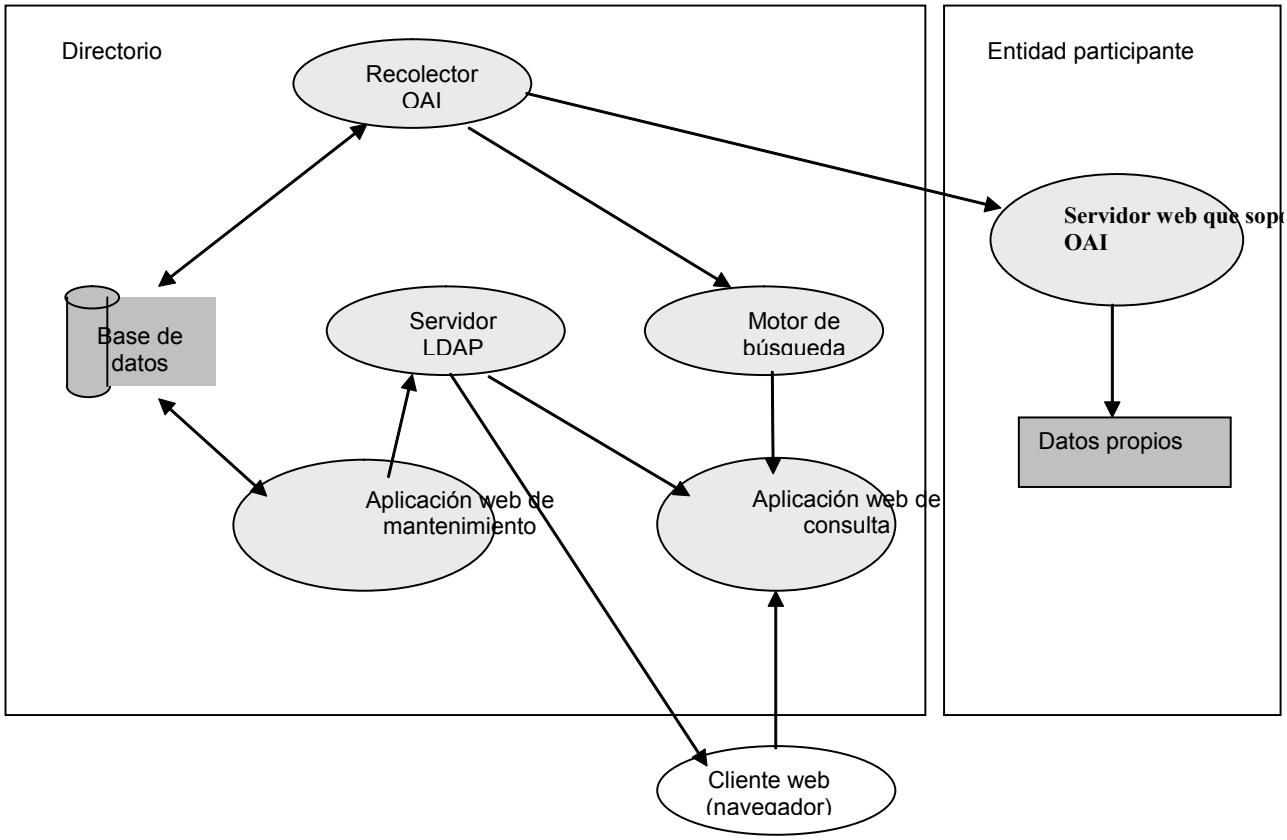



**Fig. 1:** Visión general de los componentes del sistema

La descomposición del sistema en múltiples componentes permite que se distribuya en diversos ordena-

dores o que se centralice en un único servidor, aumentando la escalabilidad total del sistema. El hecho de que sea una aplicación web pura permite acceder tanto a la parte de consulta como a la de mantenimiento desde cualquier ordenador a través de un navegador sencillo.

El entorno tecnológico es el usual en el Ministerio de Cultura:

• Base de datos: Oracle Release 9

• Aplicaciones web de consulta y de mantenimiento: desarrollado en Java para su ejecución en un servidor J2EE, Tomcat.

• Proceso recolector OAI: una aplicación Java estándar, programada mediante cron, que accederá a la base de datos Oracle y a los repositorios OAI externos a través del protocolo http.

• Motor de búsqueda a texto libre: desarrollado en C++, disponible actualmente para sistemas Windows, Linux y Solaris, se accede a él como un servicio web SOAP.

• Servidor LDAP: se utiliza el servidor OpenLDAP, que sirve como implementación de referencia del estándar, desarrollado bajo código libre. Existe un proceso batch que actualiza periódicamente el directorio con los cambios almacenados en la base de datos.

El proceso Recolector OAI se dispara diariamente de forma programada, y recorre la tabla de proyectos de digitalización. Para todos aquellos que puedan actuar como fuente de datos OAI, el recolector establecerá una conexión con los mismos, y pedirá que se le envíen las altas y modificaciones que se hayan realizado desde la última actualización. Podrán configurarse diferentes frecuencias de actualización para distintas fuentes de dato, de forma que aquellos proyectos que se actualicen con mayor frecuencia se exploren más a menudo.

La comunicación con los servidores de los proyectos de digitalización se realizará conforme al proyecto OAI (ver [OAI2]), intercambiándose la información en formato XML conforme al esquema Dublin Core.
Las altas y modificaciones se grabarán en la base de datos y se indexarán mediante el motor de búsqueda a texto libre, y serán accesibles desde ese mismo instante en la página web de consulta de la aplicación.

## 5. Perspectiva de futuro

El servicio Directorio y recolector de recursos digitales está desarrollado para permitir un crecimiento auditado, es decir, las colecciones digitales que se recogen y los proveedores de datos OAI dados de alta son establecidos por el grupo de trabajo creado entre las Comunidades Autónomas y el Ministerio de Cultura. De esta forma, el crecimiento del servicio está garantizado según vayan surgiendo nuevas colecciones digitales y/o proveedores OAI de interés.