

Implantación de la *Open Archives Initiative* en España

[Fernanda Peset](#) and [Antonia Ferrer](#)

Instituto de Diseño y Fabricación, Universidad Politécnica de Valencia, Spain

Resumen

Introducción. Se aborda la situación en España de las acciones relacionadas con el protocolo *Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH)*, usado para la recolección y difusión de metadatos de documentos. El tipo de bibliotecas digitales determinado por dicho protocolo está contando con un auge inusitado a escala mundial, con repositorios institucionales llevados a cabo principalmente por las organizaciones académicas.

Método. Se han consultado registros internacionales y se han contrastado los datos mediante dos tipos de encuestas.

Análisis. Estudiar el estado de implantación de *OAI-PMH* en España.

Conclusiones. La situación en el año 2006 resulta alentadora en cuanto a número de proyectos, pero es bastante deficitaria en cuanto a la cantidad de datos almacenados; es necesario subsanar las omisiones y los errores existentes en los registros oficiales internacionales; en España se necesitan acciones de coordinación que permitan rentabilizar las inversiones; y finalmente, la falta de estudios como el que aquí se presenta dificulta el conocimiento de este área y, por tanto, la aplicación de indicadores para evaluar la difusión de la ciencia mediante *OA*.

[Abstract in English](#)

Introducción

Las definiciones de biblioteca digital resultan algo confusas debido a la evolución de este concepto a lo largo de la última década. Se ha llamado *biblioteca digital* a una variada cantidad de desarrollos, desde los catálogos automatizados hasta la pura digitalización de una colección. Según Peset:

...es difícil acotar qué se entiende por biblioteca digital teniendo en cuenta las diferentes técnicas, materiales que se difunden, agentes implicados y servicios que ofrecen. Lo que está claro es que implica unos documentos o información en formato electrónico, la utilización de una tecnología y una forma de difusión'. ([Peset 2003](#): 241)

En este artículo nos ocupamos de un determinado tipo de biblioteca digital vinculado con la filosofía de acceso abierto a la publicación científica -Open Access-, en tanto que ha supuesto una revolución en el mundo de la publicación científica. Al utilizar el protocolo Open Archive Initiative-Protocol for Metadata Harvesting ([OAI-PMH](#)) cuentan con la ventaja de permitir no sólo el depósito de los trabajos en los archivos (para luego suministrarlos como 'proveedores de datos'), sino también el establecimiento de servicios de valor añadido reutilizando la información almacenada en esos archivos. Esto, que sin duda favorece la visibilidad de la investigación, contribuye además, tal y como detectó la OCDE ([Quint 2004](#)), a que se recupere parte de la inversión pública y se maximice su retorno social al no pagar dos veces por la misma información: la primera abonando el sueldo de los investigadores cuando escriben, y la segunda cuando las bibliotecas suscriben los títulos de revista donde se publica la investigación.

[ArXiv.org](#), creado en 1991, es un ejemplo pionero que se tomó como referencia en una famosa reunión que tuvo lugar en Santa Fe, New Mexico, en 1999, donde se acordaron los aspectos técnicos y organizativos que actualmente son directrices obligadas para los repositorios OA.

Posteriormente el movimiento Open Access se ha ido consolidando con diversas declaraciones hechas por grupos de expertos reunidos en Budapest (2002), Berlín (2003), Bethesda (2003) y Valparaíso (2004), que tuvieron una resonancia muy amplia, así como con la creación de repositorios tanto temáticos como institucionales, y editoriales que publican en OA.

En todas las ramas de las ciencias se ha extendido esta forma de comunicación que, gracias a la separación entre proveedores de datos y proveedores de servicios, se obtiene versatilidad y flexibilidad para explotar la información, creando nuevas funciones sin tener que migrar datos, introducirlos de nuevo, etc. Es decir, se rentabiliza el uso de la información a largo plazo. Se entiende pues el interés que ha suscitado entre la comunidad bibliotecaria, cuyos sistemas todavía se nutren de información recurrente que se introduce una y otra vez en cada uno de los catálogos individuales de cada organización.

En España está teniendo cierta repercusión en los últimos dos años en las Bibliotecas Universitarias. En cambio el impacto en las bibliotecas públicas podemos considerarlo aún marginal. En general, el escaso desarrollo de estas iniciativas en nuestro país, trae como consecuencia que parte de la bibliografía profesional quede oculta a la comunidad internacional que difunde sus resultados mediante repositorios OAI-PMH.

Es difícil determinar qué tipo de objetos contienen las colecciones de las bibliotecas digitales existentes, es decir los formatos que soportan la información, o cuál es el origen de los metadatos o quién debe conservar el objeto digital.

La distribución de contenidos pivota sobre tres factores críticos:

1. Las colecciones y los usuarios

Actualmente son accesibles grandes cantidades de información y los formatos de los activos digitales, así como su naturaleza, son muy variados: textos, imágenes, sonidos... Pero además no es igual disponer de una imagen fija publicitaria que de una fotografía artística; o no es lo mismo tratar textos de divulgación que científicos a texto completo. Por ello cada uno de estos objetos digitales necesita estar descrito para poder ser útil para el público al que va dirigido. Como se verá en el siguiente capítulo, existen diferentes lenguajes de descripción según la finalidad que se pretenda. Así, por ejemplo, el Learning Object Metadata (1) (LOM) está dirigido al tratamiento de datos para el aprendizaje, mientras que el Machine Readable Cataloging (MARC)(2) se usa para el intercambio de registros bibliográficos entre bibliotecas, así como para describir documentos de cara a su conservación.

Pero no sólo han aumentado las colecciones, sino también los usuarios, tal y como describe Jones ([2007](#)), pues cualquier ciudadano se puede convertir en usuario intensivo y habitual de información. En el mundo académico hay dos factores que favorecen la mayor disponibilidad de información: la edición electrónica y la aparición de depósitos de objetos digitales. El impacto de este nuevo escenario en la investigación y la enseñanza universitaria es indudable. Especialmente ahora que nuestro país ha iniciado el camino de

convergencia con Europa, la introducción de métodos docentes más prácticos obliga al estudiante a utilizar los recursos de información de manera intensiva. Las clases de tipo magistral son paulatinamente sustituidas por nuevos métodos docentes más interactivos, vivos y de estudio en grupos, cosa que amplía el número potencial de usuarios de las bibliotecas digitales.

El tipo de información que consumen la investigación y la enseñanza, la publicación científica, cuenta con el valor añadido de su calidad, que ha sido probada por los sistemas tradicionales de evaluación o certificación (publicación en revistas *peer reviewed*, básicamente). Pero también cuenta con la desventaja de su relativamente escasa distribución, pues reside en bases de datos estructuradas, que forman la parte mayoritaria de lo que se llama Internet Invisible (*invisible* porque los buscadores no pueden rastrear la información almacenada en ellas debido a la necesidad de usar códigos de acceso y lenguajes de búsqueda especiales). Esta invisibilidad constituye una pérdida todavía mayor porque la calidad de dicha información científica es superior en dos aspectos: por un lado, - como se ha dicho - existe un proceso de evaluación en su producción -en revistas y congresos-, y unos criterios para seleccionarla e introducirla en las bases de datos. Por otro, se somete a un proceso de normalización formal que mejora la recuperación, y se le incluye metainformación sobre el contenido de los documentos mediante la indización y la clasificación, algo que ahora se promociona en la web semántica.

2.- Las tecnologías y los protocolos

No sólo se han multiplicado las colecciones de piezas digitales y los usuarios potenciales. También se han creado múltiples protocolos y técnicas para su distribución. Por ejemplo, en el ámbito de las bibliotecas, el citado formato de descripción MARC aplica las clásicas normas ISBD (*International Standard Bibliographic Description*). Las tecnologías que soportan MARC son diversas y se encuentran ya muy implantadas, entre ellas la mayoría de softwares de gestión de bibliotecas.

Cuando pasamos del campo propiamente bibliotecario de gestión de colecciones al mundo de la gestión de contenidos, el número de formatos se amplía. Debemos diferenciar entre los que describen el objeto (metadatos) y los que permiten intercambiar dichos metadatos.

Existen diversos *formatos de descripción* o de asignación de metadatos a los objetos digitales. Uno de los más implantados en el sector documental es [Dublin Core](#), creado en 1995 con el propósito de estandarizar la descripción (y la recuperación) de documentos en la Web. Proporciona la información descriptiva básica de un recurso independientemente del formato en que se encuentre.

Entre los *sistemas de intercambio* de información entre distintas plataformas se encuentra XML (Extended Markup Language), en cuya filosofía se integran los formatos MPEG 21 (3) DIDL (Packaging digital content) (MPEG-21 Digital Item Declaration Language), METS (Metadata Encoding and Transmission Standard) (4), Scorm (Shareable Content Object Reference Model) (5) y el propio protocolo OAI-PMH. La *tecnología* para gestionar contenidos con estos formatos es también de lo más variado según estemos en un escenario u otro, y se tienen versiones para buscadores, para integradores de recursos de una institución -SFX (6)-, etc.

En cualquier caso, la clave para compartir datos reside en dos factores: la *interoperabilidad* entre los sistemas, y los protocolos que se usan para describir los datos. En el ámbito de la Open Archives Initiative (OAI) el software libre open source, es decir, no propietario y de código abierto, ha demostrado ser muy atractivo para los promotores de nuevas instalaciones, por lo que lo hallamos casi omnipresente. Algunos programas son: Fedora, DSpace, E-Prints, para operar archivos, y PKP (Public Knowledge Project) para servicios.

La interoperabilidad entre sistemas permite el fácil intercambio de datos, lo que propicia que los propios autores autoarchiven sus publicaciones. Ello obliga a que los estamentos políticos aporten directrices y velen por los derechos de difusión y reproducción ([Vives 2005](#)). Y a los superados problemas técnicos clásicos, como el diseño básico de sistemas de información, se suman nuevas necesidades derivadas de la sociedad-red. Por ejemplo el proyecto PURL (Persistent Uniform Resource Locator) acomete el problema de la estabilidad de la dirección URL de los recursos en internet; y la iniciativa DOI (Digital object

identifier) soluciona la cuestión de su identificación única.

Las bibliotecas digitales tienen dos desafíos:

1. *Añadir valor a los servicios de información.* Tener en funcionamiento sistemas interoperables permite la reutilización de contenidos para otros servicios, evitando trabajos rutinarios como la asignación de metadatos. Esta reutilización es una de las bases de la Open Archives Initiative, puesto que los archivos o proveedores de datos almacenan y preservan los objetos digitales, mientras que los proveedores de servicios recolectan esos datos para crear servicios de valor añadido. Por ejemplo, la Royal Netherlands Academy of Arts and Sciences (KNAW) gestiona el programa Digital Academic Repositories (DARE), que cuenta con [The Cream of Science](#), un servicio que ofrece libre acceso al 60% de los trabajos en texto completo de los científicos holandeses más relevantes.

2. *Preservar sus fondos.* La necesidad de establecer políticas que aseguren la conservación de los objetos digitales se manifiesta con fuerza día a día. Así, por ejemplo, en el entorno OAI tenemos el proyecto [Sherpa DP](#) (Creating A Persistent Preservation Environment For Institutional Repositories), que estudia la aplicación del Reference Model for an Open Archival Information System (OAIS) ISO 14721:2003, a la preservación de los fondos de los repositorios institucionales. La responsabilidad administrativa de la conservación es uno de los temas básicos a plantear cuando se comienza un archivo abierto ([Keefer 2006](#)).

3. *Crecimiento del acceso abierto.* El acceso abierto a la literatura científica tiene una especial relevancia en el ámbito de las universidades y centros de investigación ([Barrueco y Subirats 2003](#); [Ferrer et al. 2005](#)). Con la utilización del protocolo OAI-PMH estas organizaciones y su personal se benefician de la libertad de elección de software, del libre acceso a las publicaciones científicas a través del desarrollo de nuevos modelos de comunicación que clarifican los derechos de los autores sobre sus obras ([Villamon et al. 2006](#)) y la reutilización de los datos almacenados en diferentes repositorios con el objetivo de ofrecer servicios de información de valor añadido.

Así pues, el movimiento Open Access ha experimentado un aumento exponencial tanto en el número de los repositorios que se inscriben en los registros a escala mundial como en el número de documentos depositados en ellos.

El presente trabajo describe y evalúa las iniciativas OAI en España aportando distintos resultados estadísticos. Para conseguirlo se ha realizado un estudio sistemático y exhaustivo de las acciones que se están llevando a cabo en nuestro país ([Serrano y Prats 2005](#); [Frías et al. 2007](#); [Rodríguez y Alvíte 2007](#)). Las carencias que presentan los registros internacionales, como se expondrá a lo largo del texto, ha determinado la creación de un registro exhaustivo -refOA- que permite contar con datos fiables, fomentar la visibilidad de las iniciativas españolas, explotar estadísticamente los datos en aras de perfeccionar las políticas de promoción y obtener los indicadores científicos que se relacionan con una correcta difusión de la información. En definitiva, las conclusiones del trabajo han permitido diseñar los contenidos de OA-LATINO, web de referencia para OAI de hablas española y portuguesa que aún no ha visto la luz. Partiendo de la certeza de que la OAI se encuentra eclosionando en España, consideramos que es el momento de moldear estrategias que rentabilicen los esfuerzos que se inviertan en el sector de la e-ciencia.

Material y método

Desde mayo de 2006 hasta enero de 2007 se efectuó la recopilación de los repositorios y proveedores de servicios creados y mantenidos en España, Portugal y países latino americanos. Se identificaron las principales fuentes de información existentes en la Red: los registros o directorios de repositorios OAI más conocidos internacionalmente y con más número de registros como [Registry of Open Access Repositories-ROAR](#), [The Directory of Open Access Repositories-OpenDOAR](#), [Open Archives.org](#) y [Experimental OAI Registry at UIUC](#). Fueron completados con los datos de Oliver y Swain ([2006](#)).

En segundo lugar se usaron las listas de los proveedores de datos que son recolectados por parte de los proveedores de servicios o agregadores. Para descubrir nuevos proveedores de servicios se usaron los registros de [Open Archives.org](http://OpenArchives.org) y de la Universidad de Illinois (UIUC).

Los resultados del análisis se almacenaron en una base de datos llamada *refOA* (véase Anexo). Para cada uno de los elementos localizados se describió una serie de campos que permitieran el análisis estadístico. Tras recoger la información aparecida en cada uno de los recursos identificados se contactó a los responsables de los proyectos españoles para corroborar los datos recogidos. Además se envió una breve encuesta a las dos listas de correo españolas con mayor influencia en el ámbito de Biblioteconomía y Documentación y Open Access respectivamente: IweTel y OS-Repositorios. La encuesta ayudó a localizar los proyectos que no eran hasta el momento suficientemente visibles en la Red y en el ámbito OA, así como corregir los errores fruto de la recolección de datos previa. También se conocieron otras iniciativas a través de bibliografía, informes como el de la Red Española de Bibliotecas Universitarias ([Rebiun 2006](#)), contactos personales, etc. De esta base de datos se extrajo la información relativa a los proyectos nacionales y los internacionales con participación española para tabularlos estadísticamente.

Resultados

Las acciones identificadas suman un total de 67, incluyendo tres proyectos internacionales con alta participación española (E-LIS, HAL/REDIAL y SCIELO Spain). Entre ellas se cuentan no sólo repositorios y servicios/agregadores, sino también todo tipo de acciones relacionadas con la implantación del protocolo OAI y Open Access o al menos que estén inscritas en los registros internacionales mencionados en el apartado anterior. Por ejemplo, se recogen acciones de integración que tienen la finalidad de compartir conocimiento entre los administradores y profesionales vinculados a los proyectos, como sería el caso de portales o wikies -Grupo de Usuarios de Dspace en España-, y los congresos y reuniones de expertos. Como ya se ha comentado, este nivel de exhaustividad ha sido posible gracias a la ayuda de numerosos profesionales que han visto el proyecto que estamos describiendo como una oportunidad de dar mayor visibilidad al trabajo realizado en España en OAI. Especialmente los trabajos de [Rebiun \(2006\)](#), [Frías \(2006\)](#) o [Prats \(2006\)](#) ha permitido detectar la existencia de 20 iniciativas en estado embrionario, de manera que algunos de los datos que se presentan, en general, harán referencia a las 45 acciones más consolidadas. Esta es la distribución según la naturaleza de la acción.

Tipo	Número de acciones
Listas de Correo	2
Blog/Portal	2
Congresos	2
Proveedores de datos	50*
Proveedores de Servicios	8
Revista	6
Total	67
* Dos de ellos son simultáneamente Proveedores de datos y Proveedores de Servicios	

Tabla 1. Distribución del número de acciones por tipo

Como hemos comentado, la naturaleza de cada una de las acciones es diferente, aunque todas están relacionadas con la OAI. En esta tabla se ha incluido cualquier acción de la que se ha tenido

conocimiento. Los que se encuentran en desarrollo son Proveedores de datos, la mayoría -exactamente 18- de ámbito universitario. Cabe señalar que la mayor parte de acciones son repositorios universitarios -treinta y tres incluidos los que están en desarrollo-. Su procedencia académica convierte la normalidad de nuestro país un repositorio institucional que recoge la producción propia de la organización. Cabe destacar el caso de Dialnet que tiene un carácter globalizador, ya que es un proveedor de datos y de servicios que recoge la producción científica hispana. Contamos también con dos listas de correo -OS-Repositorios y GUDE-Grupo de Usuarios de Dspace en España-, dos reuniones de expertos, un blog, y un portal tipo wiki para GUDE.

Por el contrario, nuestro país no está representado en el registro de mandatos institucionales que obligan a archivar en repositorios abiertos los trabajos financiados con dinero público: ROARMAP ([Registry of Open Access Repository Material Archiving Policies](#)).

En cuanto a la distribución geográfica, los resultados muestran una gran preponderancia de la actividad en Madrid y Cataluña. Esta última reúne casi el 50% de las acciones. En el otro extremo, Baleares, Castilla-La Mancha, Asturias, y las ciudades autónomas de Ceuta y Melilla no cuentan con ninguna acción. Obviamente, Asturias y estas últimas no sorprenden pues no disponen de universidad, pero en los otros casos se echa en falta.

Nombre	Nº real	Nº total (incluyendo los que están en proyecto)
Desconocido	0	1
Andalucía	1	5
Aragón	1	2
Islas Canarias	1	1
Castilla y León	1	2
Cataluña	20	22
Galicia	2	4
Comunidad de Madrid	12*	16
Región de Murcia	1	1
Navarra	2	4
País Vasco	0	1
La Rioja	2	2
Extremadura	1	2
Comunidad Valenciana	3*	4
Total	47	67
*Se han asignado E-LIS a Valencia, HAL/REDIAL y Scielo Spain a Madrid		

Tabla 2. Distribución del número de acciones por Comunidad Autónoma



Figura 1. Distribución del número de acciones por Comunidad Autónoma

El primer proyecto datado es TDX/TDR, una biblioteca digital de tesis de todas las universidades españolas iniciada por el Consorcio de Bibliotecas Universitarias de Catalunya en 2001. En general no es fácil conocer la fecha de inicio de prestación de los servicios, pues en los registros a menudo consta la fecha de inscripción. Conocemos las fechas de inicio de 40 de las 45 acciones, lo que supone un 88% del total. La distribución permite observar un despegue que comienza en 2004 y que esperamos que se consolide en los próximos años si realmente se materializan todos los proyectos que dicen estar en desarrollo.

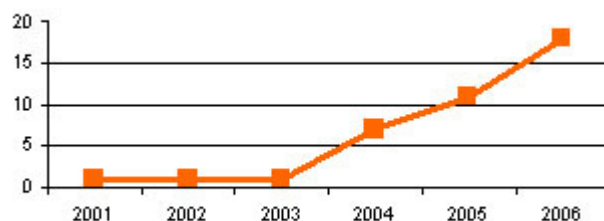


Figura 2. Distribución cronológica del número de acciones

Tras esta breve presentación de resultados hagamos un análisis detallado de los proveedores de datos y de revistas (considerando éstas como proveedores), que suman 36 en total. La naturaleza de cada uno de los proyectos difiere en muchos aspectos: unos distribuyen referencias y textos completos, otros dependen de una sola institución, otros son una pasarela temática, o un simple título de revista. Como ya se ha comentado, la mayoría son repositorios institucionales que recogen la producción científica de una organización, por lo que distribuyen contenidos de materias variadas. También Dialnet, que recoge toda la producción española o TDR -Tesis doctorales en red- cuentan con esta naturaleza interdisciplinar. A continuación, los proveedores de datos suman quince y los títulos de revista cinco. El análisis de si se trata de un proveedor temático o interdisciplinar queda como sigue:

Cobertura	Nº de proveedores de datos
Interdisciplinar	23
Específica	13
Total	36

Tabla 3: Distribución por ámbito temático

Como cabe suponer es casi imposible determinar el tipo de público final al que van dirigidos estos repositorios. En el caso de proveedores de datos vinculados a la universidad debemos considerar que se utilizan para fines corporativos además de científicos, como puede ser la recopilación de datos para las memorias anuales o la evaluación de la producción científica del personal. En el caso de las revistas, además de los suscriptores habituales quedan abiertas a la consulta de cualquier persona interesada en sus materias.

También puede resultar de interés analizar el tipo documental que distribuyen. Casi el 50% son artículos de revista/comunicaciones a congresos, entre las que se cuentan las revistas que vuelcan su contenido en la red utilizando el protocolo OAI. De nuevo constatamos que en España la implantación de este protocolo tiene su presencia mayoritaria en la universidad, donde el mayor número de documentos pertenecen al tipo documental mencionado. Tímidamente se aplica a otros soportes documentales, tal y como está ocurriendo en otros lugares debido a la potencia de distribución que el protocolo posee.

Tipo documental	Nº de proveedores de datos
Artículos/comunicaciones	15
Multisoporte	13
Títulos de revista	2
Tesis	2
Gráfico/vídeo	2
Proyectos fin de carrera	1
Manuscritos/incunables	1
Total	36

Tabla 4: Distribución por tipo documental

Los contenidos distribuidos por proveedores de datos y revistas suman 160.441 registros en diciembre de 2006. Hay que señalar que se han obtenido los datos de 28 de los 36 proveedores en activo recogidos, lo que significa un 81%. Por otra parte, dos proveedores de datos son a su vez servicios, lo que está deformando la estadística, pues suman ellos solos un 17% del total distribuido, además de que resulta imposible discernir los solapamientos. Así también, algún proveedor de datos como Dialnet sirve indistintamente registros con sus textos completos y registros compuestos únicamente por metadatos. El mayor es Dialnet artículos, con 80.000 registros. Le sigue [RACO](#), el portal de las revistas catalanas, con 25.000, y el depósito de la [Universidad Autónoma de Barcelona](#) con 15.000, la revista [ConoZe](#), y [E-LIS](#) (archivo abierto dedicado a las ciencias de la documentación e información) que contiene actualmente 7.200 documentos y [Scielo-Spain](#) con 5.000 aproximadamente. Como se ve, ninguno es representativo de la normalidad española por las razones siguientes. El primer caso, ya hemos comentado, distribuye algunos registros sin el texto completo, por lo que sería una biblioteca híbrida. Almacenados en su archivo a texto completo poseen 12000 registros. El segundo es un proveedor de servicios a la vez que de datos. El depósito de la UAB sería por tanto el primero, aunque con materiales de muy diversa índole -desde exámenes a Memorias, de enlaces externos a actas de congresos completas-. Por su parte, ConoZe almacena pequeños artículos de opinión, mientras que E-LIS y Scielo Spain son proyectos internacionales. En todo caso, sin soslayar estos sesgos en la estadística, el tamaño medio de un repositorio en España sería de 3.000 registros a texto completo.

La distribución de datos por comunidad autónoma arroja el resultado de las dos primeras columnas de datos:

Comunidad autónoma	Nº de documentos	Nº de repositorios con contenidos conocidos	Tamaño medio de los repositorios (tm)	Índice de eficiencia (tm/3.009)	PIB 2006 en millones de euros	Índice de desarrollo en OA (r/PIB)
Andalucía	2.408	1	2.408	0,800	136.472	18
Islas Canarias	1.663	1	1.663	0,553	39.547	42
Castilla y León	0	1	0	0,000	52.438	0
Cataluña	55.964	12	4.663	1,550	183.821	304
Galicia	342	2	171	0,057	50.410	7
Comunidad de Madrid	9.182	6	1.530	0,508	171.988	53
Región de Murcia	24	1	24	0,008	25.499	1
Navarra	671	1	671	0,223	16.475	41
La Rioja	12.234	2	6.117	2,033	7.143	1.713
Extremadura	0	1	0	0,000	16.109	0
Comunidad Valenciana	4.785	1	4.785	1,590	95.589	50
Total	87.273	29	3.009	1,000	*795.491	110
*Avance PIB España 2006: 980.954 millones de euros						

Tabla 5: Distribución del número de registros, del número de archivos (conocidos sus contenidos), tamaño medio, índice de eficiencia, PIB por Comunidad Autónoma e índice de desarrollo en OA.

Para analizar la implantación de la OAI utilizaremos un Índice de eficiencia en la distribución de contenidos. Se trata de un nuevo indicador que señala la eficacia de una zona en la distribución de contenidos a texto completo, en relación con el área general o el país al que pertenece. En nuestro caso lo hemos obtenido dividiendo el tamaño medio de los repositorios de cada comunidad autónoma por el tamaño medio de todos los repositorios de España.

Empezaremos explicando los resultados más notorios con valor 0. En Castilla León el proyecto de Incunables y Manuscritos de la Universidad de Valladolid no ha funcionado durante el tiempo que se investigó. Mientras que el repositorio del Instituto Universitario de Ciencias de la Educación de la Universidad de Salamanca no contenía ningún registro, -aunque sí se había inscrito en el registro internacional WIKI Dspace, distorsionando la fiabilidad de los registros internacionales-.

En todo caso la mayoría se encuentra significativamente por debajo del índice 0,5, excepto Cataluña, Rioja y Valencia -ya se ha explicado que E-LIS es un proyecto internacional, en absoluto vinculado únicamente a Valencia-. Teniendo en cuenta otros parámetros como el PIB de cada comunidad, algunas de ellas (en especial Galicia y Murcia) deberían revisar su política de inversión en estos proyectos -por supuesto, además de Castilla-La Mancha y Extremadura, con 0 documentos, y del resto de CA que no aparecen en la tabla por no poseer ni repositorios-.

En cuanto a los proveedores de servicios hemos identificado ocho iniciativas cuyo alcance y naturaleza difieren mucho entre sí. Generalmente incluyen limitaciones geográficas (Ministerio de Cultura; Recercat, depósito de la investigación catalana), o institucionales (UPCommons, portal de acceso abierto de la [Univesidad Politécnica de Cataluñ](#)). Estos servicios están constituidos por recolectores (harvesters) que permiten las búsquedas federadas (simultáneas en varios repositorios a la vez). Los servicios que recogen más archivos son [ROAI](#) y Rebiun. [E-Ciencia](#) (buscador de archivos abiertos de la Comunidad de Madrid) se alimenta básicamente de los vídeos de la UNED, que no son servidos directamente por ésta-.

En un análisis más detallado de los servicios se ha intentado comprobar la duplicación de esfuerzos y el solapamiento en la provisión de contenidos. Entre todos los proveedores sirven alrededor de 400.000 registros, sin que haya constancia de la redundancia. Se ha medido el solapamiento de los repositorios recolectados por los servicios ROAI y Rebiun: un 70%.

Quizá el futuro de esta investigación y por ende de la iniciativa OAI será atajar el problema del solapamiento en la provisión y difusión de contenidos, pues los de los repositorios no-institucionales como DIALNET o E-LIS pueden fácilmente solaparse con los de los institucionales.

Conclusiones

En primer lugar se han detectado algunos errores en los registros considerados "oficiales" del ámbito OA. El hecho de que la inscripción en dichos registros sea libre y voluntaria da lugar a que a veces se consignen datos poco fiables. También existe el problema de tener que asignar cada proyecto a un único país, cosa que no es adecuada para los proveedores de datos multinacionales, como E-LIS ([Peset et al. 2003](#)) o HAL/Redial. A modo de ejemplo señalamos el número de iniciativas que ROAR identificaba como españolas: un único proyecto en abril 2005 ([Peset y Ortín 2007](#)), 13 identificados por Muñoz Tinoco ([2006](#)), 18 en octubre de 2006, y 19 un mes después. En este estudio se han detectado 14 más (casi un 50% del total). Otro aspecto necesario es el establecimiento de filtros en la inscripción de proyectos (7), de manera que sólo se puedan registrar los que estén en marcha y tengan un número mínimo de datos.

La normalidad de una acción OA en España está constituida por un repositorio universitario, creado en 2006, con un tamaño medio de 3.000 registros aproximadamente.

Existe ya un cierto número de proveedores de datos y servicios en España, pero sus contenidos se difunden desde plataformas de muy diferente naturaleza y entidad, de manera que es difícil cuantificar en qué medida OA está aumentando la visibilidad de la ciencia.

Cabría apuntar algunas recomendaciones en consonancia con la evolución internacional del movimiento:

1. Incrementar la distribución de nuevos contenidos a texto completo y la provisión de servicios de valor añadido.
2. Promocionar el self-archiving ayudando a las instituciones a establecer políticas de mandato que obliguen a los autores a archivar su producción científica financiada.
3. Apoyar los proyectos de calidad con el fin de garantizar su mantenimiento, la preservación de los datos, la calidad de los contenidos, evitar solapamientos y, en definitiva, racionalizar el gasto.
4. Promover acciones coordinadas, compartiendo esfuerzos y conocimientos, en estos momentos sería un factor de rentabilización y aumentaría la visibilidad de la e-ciencia.
5. Crear un Web de Referencia para las iniciativas OAI en español y portugués, OA-LATINO, el cual podría contener el:
 - Registro Oficial descriptivo (refOA)
 - Barómetro anual OA (bOA)
 - Recomendaciones (recOA): Guías de buenas prácticas; Tutoriales; Checklist para poner en marcha y evaluar un archivo o un servicio; Uso de vocabularios, clasificaciones y metadatos comunes; Recomendaciones de firma a los autores hispanos para que no malogren su posible impacto internacional (IraLIS)...

Sólo contando con análisis estadísticos diacrónicos es posible detectar las tendencias y líneas de futuro. Y son las únicas herramientas que permiten aplicar indicadores científicos que contemplen aspectos de difusión mediante OA, tal y como apunta la Comisión Europea ([EURAB 2006](#)).

Notas

(1) LOM se centra en el conjunto mínimo de propiedades que permiten gestionar, ubicar y evaluar objetos educacionales. Este esquema base se compone de 9 categorías y 47 elementos. Sus categorías son: General, Ciclo de vida, Meta-información, Técnica, Uso educativo, Derechos, Relación [con otros recursos], Observaciones, y Clasificación. http://www.iaa.upf.es/~jblat/material/doctorat/students/jccbis/Estandares_LOM.htm.

(2) El formato MARC permite el intercambio de información, el acceso a catálogos colectivos y la catalogación compartida. Comenzó su desarrollo en los 60 y ha tenido distintas variaciones y adaptaciones hasta nuestros días. Actualmente está regulado por la norma ISO 2709/73.

(3) Estándar para el intercambio de contenidos multimedia que respeta los derechos de autor y de distribución, diferenciando además los de propiedad intelectual y de utilización que tiene el usuario.

(4) Estándar para la codificación descriptiva, administrativa y estructural de metadatos de bibliotecas digitales, expresado en XML. Lo mantiene la Oficina de las normas MARC de la Biblioteca del Congreso, siendo desarrollado por la Federación de Bibliotecas Digitales. (<http://www.loc.gov/standards/mets/>).

(5) Especificaciones de objetos pedagógicos estructurados. Permite la importación en sistemas de aprendizaje distintos.

(6) Integrador de recursos electrónicos que permite la búsqueda simultánea en distintas revistas electrónicas a través de sencillos formularios de búsqueda, permitiendo el acceso al documento completo si la biblioteca está suscrita.

(7) No obstante debe observarse que en el caso de este estudio se han tenido en cuenta todas las iniciativas existentes.

Referencias

- Barrueco, J. M. and Subirats, I. (2003). [Open Archives Initiative protocol for metadata harvesting \(OAI-PMH\): descripción, funciones y aplicación de un protocolo](#). *El profesional de la información*, **12**(2), 99-106 Retrieved 8 December, 2008 from <http://eprints.rclis.org/177/2/cardedeu.pdf> (Archived by WebCite® at <http://www.webcitation.org/5cwCcWV1t>)
- European Research Advisory Board. (2006). [Scientific publication: policy on open access](#). Brussels: European Commission. (EURAB 06.049) Retrieved 8 December, 2008 from <http://bit.ly/FdBS> (Archived by WebCite® at <http://www.webcitation.org/5cwKG5FGT>)
- Ferrer, A., Peset, F., Moreno, M. and Lloret, N. (2005). *Guía metodológica para la implantación de una biblioteca digital universitaria*, Gijón, Spain: Trea
- Frías, J. A., Santiago, T. and Travieso, Crispulo (2007). *La difusión de la investigación académica: panorama general de repositorios institucionales en las universidades españolas*. Presentada en el Seminario Internacional de Bibliotecas Digitais de Brasil, 18 a 20 de setembro de 2007.
- Jones, C. (2007). *Institutional repositories: contents and culture in an open access environment*, Oxford: Chandos Publishing
- Keefer, A. (2006). [Preservación digital y repositorios institucionales](#). *Anuario ThinkEPI*, 2007, 196-199. Retrieved 8 December, 2008 from <http://www.thinkepi.net/15/> (Archived by WebCite® at <http://www.webcitation.org/5cwLFJyBD>)
- Muñoz Tinoco, C. (2006). [Iniciativas y tendencias de la edición electrónica de la literatura científica: acceso abierto \(open access\)](#). *BIBLIODOC: Anuari de biblioteconomia, documentació i informació*, 2005, 65-90. Retrieved 2 February, 2007 from <http://eprints.rclis.org/archive/00005483/01/OA-4.pdf> (Archived by WebCite® at <http://www.webcitation.org/5cwNuGPz8>)
- Oliver, K.B. and Swain, R. (2006). [Directories of institutional repositories: research results and recommendations](#). Paper presented at the 72nd IFLA General Conference and Council, 20-24 August 2006, Seoul. Retrieved 8 December, 2008 from http://www.ifla.org/IV/ifla72/papers/151-Oliver_Swain-en.pdf (Archived by WebCite® at <http://www.webcitation.org/5cwLd6v4w>)
- Peset, F. (2003). [Bibliotecas digitales en Internet de libro raro, antiguo e incunables](#). *Anales de Documentación*, **6**, 241-260. Retrieved 8 December, 2008 from <http://www.um.es/ojs/analesdoc/article/viewFile/1931/1921> (Archived by WebCite® at <http://www.webcitation.org/5cwNLGpad>)
- Peset, F. and Ortín, A. (2007). [Tendencias internacionales en el acceso libre a la documentación científica digital: OAI-Open Archives Initiative](#). In *Derecho, historia y universidades: estudios dedicados a Mariano Peset*. Vol. 2. (pp. 441-452). Valencia: Univ. de Valencia. Retrieved 4 Juny, 2007 from http://eprints.rclis.org/4131/1/ortin_y_peset.pdf (Archived by WebCite® at <http://www.webcitation.org/5cwLwFH4>)
- Peset, F., Subirats, I., Barrueco, J. M. & Noverges, N. (2003). [Políticas editoriales en la documentación española e implantación de E-LIS \(E-Prints in Library and Information Science\)](#). In *Proceedings Contenidos y Aspectos Legales en la Sociedad de la Información (CALSI), Valencia (Spain)*. Retrieved 23 May, 2004 from <http://eprints.rclis.org/416/1/CALSI.pdf> (Archived by WebCite® at <http://www.webcitation.org/5cwNccjVo>)
- Quint, B. (2004, 9 February). [OECD Ministers support open access for publicly funded research data](#). *News Breaks & The Weekly News Digest*. Retrieved from <http://newsbreaks.infotoday.com/nbreader.asp?ArticleID=16519> (Archived by WebCite® at <http://www.webcitation.org/5cwOFevAK>)
- REBIUN (2006). [Resumen resultados encuesta línea 3 REBIUN: Recursos electrónicos editados por las Universidades españolas](#). Retrieved 7 October, 2006 from <http://bibliotecna.upc.es/Rebiun/nova/archivosNoticias/112.pdf> (Archived by WebCite® at <http://www.webcitation.org/5cwQ2qQc2>)
- Rodríguez Bravo, B. and Alvite Díaz, M^a L. (2007). E-science and open access repositories in Spain. *OCLC Systems and Services*, **23** (4), 363-371
- Serrano I & Prats I (2005) [Repositorios abiertos: el libre acceso a contenidos](#). *Revista de*

Universidad y Sociedad del Conocimiento (RUSC), 2 (2). Retrieved 8 December, 2008 from <http://www.uoc.edu/rusc/2/2/dt/esp/serrano.pdf> (Archived by WebCite® at <http://www.webcitation.org/5cwWJ936o>)

- Villamón, M., Devís, J. & Valenciano, J. (2006). Análisis de las 'Instrucciones para autores' de las revistas españolas de Ciencias de la Actividad Física y el Deporte. *Motricidad, European Journal of Human Movement*, 16, 133-150
- Vives Gràcia, J. (2005). Aspectos de propiedad intelectual en la creación y gestión de repositorios institucionales. *El profesional de la información*, 15 (4), 267-278

Abstract in English

Introduction. The article approaches the situation in Spain of actions related to the *Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)*, used for the harvesting and diffusion of metadata for documents. The type of digital libraries, under this protocol, is experiencing worldwide development with institutional repositories developed mainly by academic organizations.

Method. International inventories have been consulted and have contrasted the data through two types of surveys.

Analysis. Studying the state of implantation of *OAI-PMH* in Spain.

Conclusions. The situation in 2006 was encouraging in terms of number of projects, but less so in terms of the amount of stored data. It is necessary to remedy the omissions and errors in the official international registers and measures are required in Spain for coordination to enable profitable investments. Finally, the lack of studies such as presented here hinders the understanding of this area and therefore the application of indicators to assess the dissemination of science through open archives.

How to cite this paper

Peset, F. & Ferrer, A. (2008). "Implantación de la Open Archives Initiative en España". *Information Research*, 13(4) paper 385. [Available at <http://InformationR.net/ir/13-4/paper385.html>]

Find other papers on this subject

Scholar Search

Google Search

Windows Live

■ [Bookmark This Page](#)

Anexo: Lista de iniciativas analizadas, URL e institución

Nombre	URL	Institución promotora
ConoZe.com	http://www.conoze.com/interfaz/oai/	ConoZe.com
Biblioteca Virtual de Andalucía	http://www.juntadeandalucia.es/cultura/bibliotecavirtualandalucia	Junta de Andalucía. Consejería de Cultura
Repositorio CBUA Tesis según REBIUN	En desarrollo	CBUA Consorcio Bibliotecas Universitarias Andalucía
Futuro repositorio Universidad de Cádiz	En desarrollo	Universidad de Cádiz
Futuro repositorio Universidad de Granada	En desarrollo	Universidad de Granada
Futuro repositorio Universidad de Sevilla	En desarrollo	Universidad de Sevilla
Universidad de Las Palmas de Gran Canaria	http://bdigital.ulpgc.es/mdc	Universidad de Las Palmas de Gran Canaria
Universidad de Valladolid Manuscritos e Incunables	http://gutenberg.dcs.fi.uva.es:8080/OAI_UVA/Oai_incunables	Universidad de Valladolid
Futuro repositorio Universidad de Burgos	En desarrollo	En desarrollo
GUDE	gude@listserv.rediris.es , http://www.rediris.es/list/info/gude.es.html	Universitat Politècnica de Catalunya
TDR / TDX	http://www.tesisenred.net	CBUC/CESCA (Consorti de Biblioteques Universitaries de Catalunya)
Temaria	http://temaria.net	Universitat de Barcelona
ARCA (Arxiu de Revistes Catalanes Antiques)	http://www.bnc.cat/digital/arca/index.html	Biblioteca de Catalunya-CBUC-Universitat Rovira i Virgili
RECERCAT	http://www.recercat.net	CBUC/CESCA
RACO	http://www.raco.cat	CBUC/CESCA
GUDE Grupo de Usuarios de Dspace de España	http://sod.upc.es/gude/index.php/Instalaciones_de_Dspace	Universitat Politècnica de Catalunya

Statistics and Operations Research Transactions - SORT	http://www.idescat.net/sort/	IDESCAT Instituto de Estadística de Cataluña
Intangible Capital	http://www.intangiblecapital.org/	Pep Simó Guzmán
Tiempos Modernos: Revista Electrónica de Historia Moderna	http://www.tiemposmodernos.org	Asociación Mundos Modernos y REDIRIS
UPCommons: portal d'accés obert al coneixement de la UPC, antes DSpace.UPC	http://upcommons.upc.edu	Universitat Politècnica de Catalunya
Recolector de Recursos digitales	http://Eprints.upc.es/rebiun	REBIUN. Red de Bibliotecas Universitarias
UPC-PFC, TFC i Tesines	http://biblioteca.upc.es/pfc/	Universitat Politècnica de Catalunya
UPC-Revistas / DSpace.Revistas UPC	UPC-Revistas / DSpace.Revistas UPC	Universitat Politècnica de Catalunya
UPC-Eprints / DSpace.E-prints UPC	http://e-prints.upc.edu	Universitat Politècnica de Catalunya
UABPubD Dipòsit Digital de Documents	http://ddd.uab.es/	Universitat Autònoma de Barcelona
DUG: Dipòsit Universitat de Girona	http://diobma.udg.es:8080/dspace/index.jsp	Universitat de Girona
Videoteca digital de la UPC	http://biblioteca.upc.edu/video	Universitat Politècnica de Catalunya
Arxiu Gràfic de l'ETSAB	http://eprints.upc.es/adminbustia/bib210/	Universitat Politècnica de Catalunya
Depósito de archivos visuales	En desarrollo	Universitat Politècnica de Catalunya
Depósito materiales docentes / La Factoría	http://biblioteca.upc.edu/factoria	Universitat Politècnica de Catalunya
Futuro repositorio Universidad Rovira i Virgili	En desarrollo	Universitat Rovira i Virgili
avalia-t Axencia de Avaliación de Tecnoloxías Sanitarias de Galicia ·	http://avalia-t.sergas.es	Axencia de Avaliación de Tecnoloxías Sanitarias de Galicia · Subdirección Xeral de Aseguramento e

Subdirección Xeral de Aseguramento e Planificación Sanitaria		Planificación Sanitaria - Consellería de Sanidade · Xunta de Galicia
Futuro repositorio de tesis BUGALICIA	En desarrollo	BUGALICIA Consorcio de Bibliotecas Universitarias de Galicia
DSpace da UDC	http://dspace.udc.es/	http://dspace.udc.es/
Futuro repositorio Universidad de Vigo	En desarrollo	Universidad de Vigo
Open Access. Acceso abierto a la información científica	http://weblogs.madrimasd.org/openaccess	Madrid ImasD
Invenia Repository for Technological Innovation	http://www.invenia.es/	Invenia (empresa), con colaboración de AEDHE Asociación de Empresarios del Henares y Agencia Navarra de Innovación y Tecnología S. A., navactiva.com
Hal / REDIAL (Red Europea de Información y Documentación sobre América Latina)	http://halshs.archives-ouvertes.fr/REDIAL/es/	CNRS Francia internacional
Biblioteca digital de prensa histórica	http://www.mcu.es/prensaHistorica/es/consulta/busqueda.cmd	http://www.mcu.es/prensaHistorica/es/consulta/busqueda.cmd
Scielo Spain	http://wwwscielo.isciii.es/scielo.php	Instituto de Salud Carlos III
e-revistas Tecnociencia	http://www.tecnociencia.es/e-revistas	FECYT - Fundación Española para la Ciencia y la Tecnología
Recolector del Ministerio de Cultura	http://roai.mcu.es/es/inicio/inicio.cmd	Ministerio de Cultura
E-ciencia Comunidad de Madrid y Consorcio Madroño	http://www.madrimasd.org/informacionidi/e-ciencia/default.asp	Consorcio Madroño
E-Prints Complutense	http://www.ucm.es/eprints	Universidad Complutense de Madrid
e-Spacio UNED	http://e-spacio.uned.es/fez/index.php	Universidad nacional de Educación a Distancia

Universidad Carlos III de Madrid - Departamento de Biblioteconomía y Documentación	http://hipatia.uc3m.es:8080/dspace/	Universidad Carlos III de Madrid
Portal de revistas científicas UCM	En desarrollo	Universidad Complutense de Madrid
Futuro repositorio Universidad Alcalá de Henares	http://dspace.uah.es/dspace/	Universidad de Alcalá de Henares
Futuro repositorio tesis Universidad Cardenal Herrera	En desarrollo	Universidad Cardenal Herrera
Futuro repositorio Universidad Rey Juan Carlos	En desarrollo	Universidad Rey Juan Carlos
Universidad Carlos III de Madrid - E-archivo Universidad Carlos III	http://e-archivo.uc3m.es:8080/dspace	Universidad Carlos III de Madrid
TONOSDIGITAL- Universidad de Murcia	http://www.tonosdigital.es/ojs/index.php/journal/oai	Universidad de Murcia. Facultad de Letras
Proveedor de Servicios grupo DELI	http://www.deli.deusto.es/oai/provServ/buscar	Grupo DELI
Revistas electrónicas de la Universidad de Navarra	http://dspace.unav.es/index.jsp	Universidad de Navarra
Universidad de Navarra	En desarrollo	Universidad de Navarra
Futuro repositorio tesis Universidad de Navarra	En desarrollo	Universidad de Navarra
Futuro repositorio Universidad de Mondragón	En desarrollo	Universidad de Mondragón
Dialnet Artículos	http://dialnet.unirioja.es	Universidad de la Rioja
Dialnet Tesis	http://dialnet.unirioja.es	Universidad de la Rioja

Instituto Universitario de Ciencias de la Educación DSpace en la Universidad de Salamanca	http://noesis.usal.es:443/dspace/?	Universidad de Salamanca
Futuro repositorio departamental, Universidad de Salamanca	En desarrollo. http://bomarzo.rec.usal.es:8080/dspace	Universidad de Salamanca
OS-Repositorios	os-repositorios@listserv.rediris.es, http://www.rediris.es/list/info/os-repositorios.es.html	IATA-Rediris
E-LIS	http://eprints.rclis.org	Internacional
Universidad de Alicante-RUA	En desarrollo	Universidad de Alicante
Futuro repositorio tesis Universidad de Zaragoza	En desarrollo	Universidad de Zaragoza
Repositorios institucionales: una vía hacia el acceso, la visibilidad y la preservación de la producción científica	http://redsicura.iata.csic.es/xarxa/ocs/index.php?cf=4	Universidad de Zaragoza-IATA-UNED
Servicios de información basados en la iniciativa OAI Workshop CALSI	http://www.calsi.org/workshop/index.htm	Universidad Politécnica de Valencia