

Fundamentos tecnológicos del acceso abierto: Open Archives Initiative y Open Archival Information System

Por Teresa Silió

Resumen: El propósito de este artículo es estudiar las tecnologías que hacen posible el acceso abierto y la preservación de los documentos digitales. Por un lado, se analiza la Open Archives Initiative (OAI) — que se ha desarrollado para posibilitar el acceso al contenido de archivos distribuidos—, y sus componentes: los Proveedores de Datos, los Proveedores de Servicios y el protocolo OAI-PMH. Éste se creó para la extracción normalizada de metadatos de los archivos electrónicos, asegurando con ello la infraestructura para la preservación y la interoperabilidad entre archivos, independientemente de las organizaciones a las que pertenezcan o de donde se encuentren. También se analiza el papel de los enlaces y los metadatos en el acceso abierto, en concreto CrossRef, DOI y OpenURL. Por otro lado, se analiza el estándar del modelo de referencia Open Archival Information System (OAIS) para un sistema abierto de información en archivos, cuya doble finalidad es la preservación y el acceso a largo plazo de la información en formato digital, obviando los previsibles cambios tecnológicos. Estas dos iniciativas, OAI y OAIS, tienen un interés generalizado para las distintas instituciones académicas y para la información científica de las diferentes áreas del conocimiento.



Teresa Silió es licenciada y doctora por la UAM. Actualmente está en el Laboratorio de Internet Cindoc-CSIC y es profesora de cursos de la CAM de biblioteconomía y documentación. Desde 2000 ha dado clases de documentación en ciencias de la salud en la UAX. Ha trabajado como Jefe del Servicio de Biblioteca y Documentación del CNIC (ISCIII). Durante cinco años ha trabajado en el CIB (CSIC). Además, ha coordinado el diseño y los contenidos de la Biblioteca Virtual de la Red Temática Cardiovascular (Recava) del ISCIII.

Palabras clave: Acceso abierto, OAI, OAIS, OAI-PMH, E-prints, Archivos, Repositorios, Metadatos, Interoperabilidad, Preservación, Difusión.

Title: Technological foundations of open access: Open Archives Initiative and Open Archival Information System

Abstract: The aim of this paper is to analyse the technological foundations of both open access and digital archives preservation. First, a discussion is offered of the Open Archives Initiative, which was developed to facilitate the discovery of content stored in distributed archives. The main components of this initiative are also described: Data Providers, Service Providers, and the OAI-PMH protocol, which was created for the standardised metadata harvesting of the archives. By this means the infrastructure for preservation and interoperability of archives is assured, regardless of their physical location or their parent organisations. Links and metadata related to open access are also analysed, specifically CrossRef, DOI, and OpenURL. Subsequently the article explores the Reference Model for an Open Archival Information System (OAIS), which was created to assure long term access and preservation of information, regardless of future technological changes. These two initiatives, OAI and OAIS are extremely important for academic institutions and for scientific information.

Keywords: Open access, OAI, OAIS, OAI-PMH, E-prints, Archives, Repositories, Metadata, Interoperability, Preservation, Dissemination, Information.

Silió, Teresa. “Los fundamentos tecnológicos del acceso abierto: Open Archives Initiative y Open Archival Information System”. En: *El profesional de la información*, 2005, septiembre-octubre, v. 14, n. 5, pp. 365-380.

EL ACCESO ABIERTO (OA) y la Open Archives Initiative (OAI) tienen como finalidad la interoperabilidad (o compatibilidad) entre archivos, la preservación a largo plazo y el acceso universal a la información. Open access se refiere a la política de acceso abierto de las publicaciones científicas, mientras que OAI está en relación con la tecnología necesaria para llevarla a cabo gracias a la creación de repositorios compatibles de publicaciones electrónicas.

Los problemas de interconexión de archivos, al igual que el de las múltiples interfaces de búsqueda, las búsquedas distribuidas en varios archivos o las complicaciones con la sintaxis, han sido solucionados gracias a los metadatos, que son recopilados mediante el protocolo OAI-PMH. Esto ha supuesto la demostración de que es posible crear una biblioteca digital basada en la interoperabilidad y que es posible poner en común distintos dominios del conocimiento.

Open Archives Initiative (OAI)

Se lanzó con el objetivo de diseñar un sistema que posibilite la interoperabilidad entre archivos abiertos de *e-prints*, que son documentos de investigación científicos o técnicos publicados electrónicamente que sirven para facilitar el intercambio entre pares y el avance científico. <http://www.eprints.org>

Abarcan los siguientes tipos:

—*Pre-prints*: trabajos pendientes de la revisión por pares (*peer review*).

—*Post-prints*: ya han sido evaluados.

—*Reprints*: separatas o copias de los documentos ya evaluados y publicados, que la editorial manda a los autores para su difusión.

Los archivos abiertos de *e-prints* son repositorios electrónicos donde los autores depositan (autoarchivan) las versiones completas de sus artículos ya revisados o las versiones preliminares que están pendientes de la revisión por pares (Bailey, 2005). Se trata de que los materiales almacenados estén disponibles para su búsqueda y su descarga, tanto en el caso de los artículos no publicados como en el de los textos de investigación ya dados a conocer en revistas científicas. La idea de que un *preprint* sea revisado antes de su publicación por algunas personas no es una novedad, pero sí lo es el nuevo alcance actual, que sean enviados a un servidor desde donde pueden ser descargados y revisados por otros científicos antes de pasar el filtro definitivo de la revisión por pares por parte de la revista (Franklin, 2002-2003).

Existen dos grandes métodos de OA a las publicaciones científicas: por un lado los repositorios de documentos electrónicos y por otro las revistas que proporcionan acceso abierto a los lectores gracias a un sistema de cobro a los autores por la revisión y la publicación de sus artículos. Según un informe del *Inist/CNRS*, el triunfo del acceso abierto puede depender solamente de que se asegure la realización de *peer review* o evaluación por pares (Franklin, 2002-2003), por lo que cabe esperar que en el futuro se establezcan servicios de revisión en los repositorios de *e-prints* y en los archivos institucionales.

Se han creado archivos abiertos de *e-prints* que permiten la rápida difusión de los resultados de la investigación (McKiernan, 2004). El sistema de repositorios, aceptado por gran parte de la comunidad académica, fue utilizado por primera vez en *Los Alamos National Laboratory* en 1991, dando origen al archivo de física y matemáticas *ArXiv* (Ginsparg et al., 1999). Posterior-

mente se han desarrollado otros casos: tanto los ordenados por disciplinas como los archivos académicos que pueden contener tanto *pre-prints* como *e-prints* (Hitchcock, 2003).

Los archivos de *e-prints* no significan simplemente un crecimiento en la escala y en el número de los archivos, sino la introducción del concepto de “archivos distribuidos”, que permite considerar varios archivos como si fuera uno solo de manera que se puedan ofrecer servicios como las búsquedas precisas o la medida del impacto de las publicaciones (Hitchcock, 2003). La infraestructura para la creación de archivos distribuidos se estableció en 1999 con la *Open Access Initiative* y al protocolo *OAI-PMH* para la recuperación de metadatos. *OAI* evolucionó a partir del prototipo *UPS (Universal Preprint Service)* (Ginsparg et al., 1999), cuyo origen está en la reunión de Santa Fe (México), donde se trató el tema de la creación de un servicio universal para el autoarchivo de *e-prints* por parte de los autores (Fox, 2000) para comunicar los resultados de una investigación antes de que fuese sometida a la revisión por pares y antes de su publicación en una revista científica (*OAForum.org*). *UPS* está pues en la base del protocolo *OAI-PMH* y de hecho, *OAI-PMH* ha heredado la arquitectura básica de *UPS* con sus dos roles lógicos: el suministrador de los datos y quien ofrece los servicios.

Bastantes editores permiten el autoarchivo en depósitos (institucionales, temáticos, etc.) de copias de los artículos ya publicados, como por ejemplo la editorial *Cambridge University Press*, que ha lanzado el repositorio *DSpace@Cambridge*. Es conveniente que los autores sepan qué editoriales permiten el archivo en repositorios, y para ello es necesario acudir a páginas como *Romeo (Rights Metadata for Open archiving)* donde aparece in-



baratz

En primera línea en
Sistemas de Información y Gestión del Conocimiento



Raimundo Fernández Villaverde, 28 28003 Madrid (España) Teléfono +34 91 456 03 60 · Fax +34 91 533 09 58 www.baratz.es · E-mail: informa@baratz.es

- Soluciones para bibliotecas:
Absys, absysNET, Absys *express*
- Gestión documental y del conocimiento:
BKM, Baratz Windows, BRSCGI
- Soluciones para archivos: **Albalá**
- Servicios de Catalogación Retrospectiva
- Edición en CD-Rom



UNE-EN ISO 9001:2000

Direcciones web y blogs de interés

- ARC (A Cross Archive Search Service)
<http://arc.cs.ou.edu/>
- Archon
<http://archon.cs.ou.edu/>
- ARNO Academic Publications Online
<http://arno.uvt.nl/~arnosite/index.html>
- ArXiv
<http://www.arxiv.org>
- BioMed Central
<http://www.biomedcentral.com/>
- CERN Document Server Software (CDSware)
<http://cdsware.cern.ch/>
- ChemWeb.com
<http://www.chemweb.com/preprint>
- Citebase
<http://citebase.eprints.org/cgi-bin/search>
- Citeseer Scientific Literature Digital Library
<http://citeseer.ist.psu.edu/>
- CogPrints
<http://cogprints.soton.ac.uk/>
- CompSciPrePrints
<http://www.sciencedirect.com/preprintarchive>
- Computer Research Repository (CoRR)
<http://uk.arxiv.org/archive/cs/intro.html>
- Creative Commons
<http://creativecommons.org/license/?lang=es>
- CrossRef Multiple Resolution
<http://www.crossref.org/mr/index.html>
- DigiLib: University of Groningen: a weblog on scholarly online publishing, open access and library related technology
<http://db.weblog.ub.rug.nl/archives/research/>
- Digital Commons Proquest
http://www.proquest.com.br/espanhol/produtos/digital_commons.htm
- Digital Library of MIT Theses
<http://theses.mit.edu/>
- Digital Library of the Commons (DLC)
<http://dlc.dlib.indiana.edu/view/subjects/>
- DList (Digital Library Information Science & Technology)
<http://dlist.sir.arizona.edu>
- Dspace
<http://www.dspace.org/>
- Dublin Core metadata element set
<http://es.dublincore.org/documents/dces>
- Dublin Core qualifiers
<http://dublincore.org/documents/2000/07/11/dces-qualifiers/>
- E-LIS (Library and Information Sciences)
<http://eprints.rclis.org/>
- E-prints
<http://www.eprints.org>
- E-PrintsUCM, Universidad Complutense de Madrid
<http://www.ucm.es/eprints>
- European Patent Office
<http://www.european-patent-office.org/index.en.php>
- Fedora: an open-source digital repository management System
<http://www.fedora.info>
- Greenstone
<http://www.greenstone.org>
- Google Print Google Project
<http://print.google.com/googleprint/library.html>
- Highwire
<http://highwire.stanford.edu/>
- Ibiblio
<http://www.ibiblio.org/collection>
- Infomine Scholarly Internet Resource Collection
<http://infomine.ucr.edu/>
- Institute of Physics
<http://www.iop.org/EJ/>
- Institution Archives Registry
<http://archives.eprints.org/eprints.php>
- ISO Archiving Standards-Overview
<http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html>
- i-Tor (NIWI)
<http://www.i-tor.org/en/>
- JISC Information Environment Service Registry (IESR)
<http://iesr.ac.uk/>
- JISC Open Access Briefing Paper
<http://www.jisc.ac.uk>
- Kepler
<http://kepler.cs.ou.edu/>
- LinkFinderPlus
<http://www.endinfosys.com/prods/linkfinderplus.htm>
- Lockss
<http://lockss.stanford.edu/>
- Mathematics Preprint Server
<http://www.ams.org/global-preprints/special-server.html>
- Marxcml
<http://www.loc.gov/standards/marxcml/>
- MIT Repository
http://www.rlg.org/en/images/projects/MIT_Infrastructure+OAIS.jpg
- MyCoRe
<http://www.mycore.de/eng/index.html>
- MyOAI
<http://myoai.com>
- NASA Technical Report Server
<http://ntrs.nasa.gov/?method=advanced>
- Ncstrl (Networked Computer Science Technical Reference Library)
<http://www.ncstrl.org/>
- NetPrints: Clinical Medicine & Health Research
<http://clinmed.netprints.org/home.dtl>
- NSDL National Science Digital Library
<http://nsdl.org>
- OAI Open Archives Initiative
<http://www.openarchives.org>
- OAI-PMH Open Archives Initiative Protocol for Metadata Harvesting
<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>
- OAI Registered Data Providers
<http://www.openarchives.org/Register/BrowseSites>
- OAI Repository Explorer
<http://re.cs.ucl.ac.za/>
- OAI Service Providers
<http://www.openarchives.org/service/listproviders.html>
- OAI Tutorial
<http://www.oaforum.org/tutorial/>
- OAIcat OCLC
<http://www.oclc.org/research/software/oai/cat.htm>
- OAIS Modeled Repositories
http://www.rlg.org/en/page.php?Page_ID=377
- OAIster
<http://oaister.umdl.umich.edu/oaister/>
- Oasis (Organization for the Advancement of Structured Information Standards)
<http://www.oasis-open.org>
- OCLC Digital Archive
http://www.rlg.org/en/images/projects/OCLC_Digital_Archive.gif
- OpCit
<http://arabica.ecs.soton.ac.uk/code/doc/ReadMe.html>
- OpCit Metalist of Open Access Eprint Archives
<http://opcit.eprints.org/explorearchives.shtml>
- Open Archives Forum
<http://www.oaforum.org>
- Open Language Archives Community
<http://www.language-archives.org>
- OpenSource
<http://www.opensource.org>
- OpenURL
<http://library.caltech.edu/openurl>
- OpenURL, a tutorial
http://www.endinfosys.com/pdf/openurl4_02.pdf
- Open Video Project
<http://www.open-video.org/>
- OSTI Office of Scientific and Technical Information
<http://www.osti.gov/eprints/>
- PubMed Central
<http://www.pubmedcentral.gov>
- Rapid Visual OAI Tool
<http://rvot.sourceforge.net/>
- Registering as a Data Provider OAI-PMH version 2.0
<http://www.openarchives.org/data/registerasprovider.html>
- RePEc
<http://www.repec.org/>
- Repository Explorer
<http://oai.dlib.vt.edu/cgi-bin/Explorer/oi2.0/testoai>
- Romeo
<http://romeo.eprints.org>
- SciELO
http://scielo.isciii.es/scielo.php?lng_es
- Science Commons
<http://science.creativecommons.org/>
- ScienceDirect
<http://www.sciencedirect.com>
- Scirus
<http://www.scirus.com/srsapp/aboutus/#sources>
- Simple Dublin Core metadata
<http://dublincore.org/usage/terms/dc/current-elements/>
- Sparc Open Access News Letter
<http://www.arl.org/sparc/index.html>
- Sparc Publishing Resources
<http://www.arl.org/sparc/resources/pubres.html>
- Scopus Open Access Journals List
http://www.info.scopus.com/aboutscopus/documents/oa_list.xls
- Scopus Content Coverage
<http://www.info.scopus.com/aboutscopus/content-coverage/>
- Sherpa
<http://www.sherpa.ac.uk/romeo.php?all=yes>
- Sourceforge
<http://sourceforge.net>
- Tesis Doctorales en Red de las Universidades de Cataluña
http://www.tdr.cesca.es/index_tdx_cs.html
- Xml
<http://www.xml.org>
- Zmarco
<http://zmarco.sourceforge.net/>

formación acerca de si la editorial lo permite, si es posible hacerlo antes o después de la publicación, si admite los dos casos o no permite en absoluto el archivo. En webs como *Sherpa* se recoge información sobre los repositorios, las editoriales y las políticas detalladas que éstas siguen. Así, por ej., la editorial *Kluwer*, una división de *Springer*, permite que el autor archive el *pre-print* (*pre-refereeing*, antes de la evaluación) con la posibilidad, además, de que el autor archive el *post-print* (post-evaluación).

<http://www.lboro.ac.uk/department/s/dis/disresearch/romeo/index.html>
<http://www.sherpa.ac.uk/>

Para ilustrar la diferencia vamos a comparar un artículo ya publicado en la revista *Serials review*, de *Elsevier*, y su correspondiente *e-print*, que ha sido autoarchivado por el autor en *E-LIS*, un repositorio abierto de documentos sobre biblioteconomía y documentación.

Para ello accedemos a la publicación a través *ScienceDirect*, la plataforma electrónica de pago de *Elsevier* que contiene centenares de revistas. El artículo aparece en la citada publicación periódica, correlativamente con los otros artículos de ese número (figura 1):

The screenshot shows the Elsevier Serials Review website interface. At the top, there are navigation tabs: Home, Journals, Books, Abstract Databases, My Profile, Alerts, and Help. Below this is a search bar and a 'WELCOME GUEST USER' message. The main content area displays the journal title 'Serials Review' with the copyright notice 'Copyright © 2005 Elsevier Inc. All rights reserved.' and the issue information 'Volume 30, Issue 4, Pages 257-381 (2004)'. A 'display checked items' button and an 'e-mail articles' button are visible. The article list shows four items, with the third item being the article of interest: 'The Criteria for Open Access' by David Goodman, pages 258-270. The article is listed as 'ARTICLE' and includes links for 'SummaryPlus', 'Full Text + Links', and 'PDF (184 K)'. The page number 'articles 1 - 24' is shown at the bottom right of the article list.

Figura 1. Artículo revisado y publicado por una editorial

The Criteria for Open Access

Goodman, David (2004) The Criteria for Open Access. *Serials Review* 30(4): pp. 258-270.

Full text available as:
 Other (Microsoft Word)
 PDF - Requires Adobe Acrobat Reader or other PDF viewer.

Abstract

Each proposal for Open Access (OA) has its unique combination of features; each argument for or against OA focuses on particular features or criteria. This article is intended to discuss these criteria, both individually and also as each of them contributes to the different proposals for OA. Evaluation of the proposals themselves is not attempted. This discussion is intended to be of value to the supporters of OA, in choosing which plan to adopt, and to those opposed to OA, in showing where the weaknesses do and do not lie. In other words, this article intends to improve the level of factual understanding in the ongoing discussions.

Keywords: Open Access, Open Archives, Scholarly Publishing, Scholarly Communication

Subjects: H. Information sources, supports, channels. > HN. e-journals. E. Publishing and legal issues. > ED. Intellectual property: author's rights, ownership, copyright, copyleft. B. Information use and sociology of information. > BF. Information policy

ID Code: 3542

Deposited By: Goodman, David

Deposited On: 02 April 2005

Figura 2. E-print correspondiente al artículo

Goodman, D., *The criteria for open access*, *Serials Review*, Volume 30, Issue 4, 2004, Pages 258-270.

disponible en:

<http://sciencedirect.com/science/journal/00987913>

Después, utilizando un buscador como *Google*, que rastrea toda clase de contenidos, incluidos los *open access* (Suber, 2005a), se llega a *E-LIS*, donde el autor ha autoarchivado o depositado el *e-print*

correspondiente al artículo anterior, como se ve en la figura 2:

Deposited By: Goodman, David

Deposited On: 02 April 2005

Al efectuar su descarga aparece más información sobre él:

This is the author's version of this paper, as published in Serials review, volume 30, no. 4: Special Focus on Open Access: Issues, Ideas, and Impact.

doi:10.1016/j.serrev.2004.09.009, copyright: Elsevier, Inc 2004.

Existen redes electrónicas de *e-prints* (*e-print networks*) creadas por científicos para ser usadas por otros profesionales y estudiantes avanzados. En realidad se trata de portales que integran sitios web y bases de datos, y están provistos de herramientas especializadas que permiten:

—Hacer búsquedas *deep web*, combinado consultas a texto completo en documentos pdf en sitios web con la búsqueda distribuida en bases de datos.

—Servicio de alertas automáticas sobre nuevos *e-prints* de interés.

—Información sobre sociedades y autores.

La interfaz de búsqueda de esta red permite seleccionar dónde y cómo hacer la búsqueda: por autor, institución, disciplina, fecha, archivo, etc.

Componentes de la OAI

La OAI está encaminada a ofrecer el acceso libre a la web científica en forma de macro-archivos. An-

te la multiplicidad de archivos y de interfaces de búsqueda se ha recurrido a los metadatos y al lenguaje de marcado xml de cara a conseguir la interoperabilidad entre archivos (así como, a largo plazo, la preservación de la información).

«OpenURL es una dirección web que contiene los metadatos del artículo»

El carácter “abierto” de esta iniciativa se refiere al hecho de permitir el *metadata harvesting*, es decir, la localización (eventualmente automática) y la recopilación de los metadatos que describen los registros contenidos en los repositorios mediante el protocolo OAI-PMH. Además OAI promueve los estándares necesarios para crear los registros de información en sistemas electrónicos distribuidos y descentralizados para ofrecer servicios, gracias a la interoperabilidad entre sistemas de metadatos. Debido a la extracción de estos metadatos, los proveedores de servicios OAI ofrecen servicios complejos de búsqueda.

<http://www.openarchives.org>

OAI promueve el acceso abierto a publicaciones digitales académicas, mediante el desarrollo de unos estándares de interoperatividad. Funciona según el esquema siguiente: los proveedores de datos (OAI data providers) ponen su información a disposición de los proveedores de servicios (OAI service providers), que localizan los metadatos por medio de OAI-PMH y construyen servicios de búsqueda. Según esto, tres son los elementos principales de OAI:

1. Proveedores de datos

Les corresponde crear, conservar y depositar los recursos en un repositorio y exponer sus metadatos. De cara a la interoperabilidad, los repositorios deben tener la in-

Biblioteconomía y documentación	E-LIS. http://eprints.rclis.org DList (Digital Library Information Science & Technology). http://dlist.sir.arizona.edu
Biomedicina	BioMed Central. http://www.biomedcentral.com PubMed Central. http://www.pubmedcentral.gov Highwire. http://highwire.stanford.edu/ Scielo. http://scielo.isciii.es/scielo.php?lng_es NetPrints: Clinical Medicine & Health Research. http://clinmed.netprints.org/home.dtl
Ciencia y tecnología	National Science Digital Library (NSDL). http://nsdl.org/ Citebase. http://citebase.eprints.org/cgi-bin/search
Economía	REPEC. http://www.repec.org/
Física	ArXiv. http://es.arxiv.org/ Archon. http://archon.cs.odu.edu/ Institute of Physics. http://www.iop.org/EJ/ NASA Technical Report Server. http://ntrs.nasa.gov/?method=advanced CERN Document Server (CDS). http://cdsweb.cern.ch/?ln=en
Informática	Ncstrl. http://www.ncstrl.org/ Citeseer. http://citeseer.ist.psu.edu/ Computer Research Repository (CoRR). http://uk.arxiv.org/archive/cs/intro.html
Lingüística y psicología	CogPrints. http://cogprints.soton.ac.uk/ Open Language Archives Community. http://www.language-archives.org/
Matemáticas	Mathematics Preprint Server. http://www.ams.org/global-preprints/special-server.html
Química	ChemWeb.com. http://chemweb.com/
Tesis doctorales	Digital Library of MIT Theses. http://theses.mit.edu/ Tesis Doctorales en Red de las Universidades de Cataluña. http://www.tdr.cesca.es/index_tdx_cs.html
Varios	D-Space, MIT. https://dspace.mit.edu/index.jsp E-PrintsUCM, Universidad Complutense de Madrid. http://www.ucm.es/eprints European Patent Office. http://www.european-patent-office.org/index.en.php Digital Library of the Commons (DLC). http://dlc.dlib.indiana.edu/view/subjects/ Ibiblio Archivo Digital (por CDU). http://www.ibiblio.org/collection The Infomine Scholarly Internet Resource Collection. http://infomine.ucr.edu/ JISC Information Environment Service Registry (IESR). http://iesr.ac.uk/ Oaister. http://oaister.umdl.umich.edu/o/oaister/
Vídeo	Open Video. http://www.open-video.org/

Tabla 1. Algunos ejemplos de OAI data providers





content + tools + services

OVID

Ovid LinkSolver™

La próxima generación de Enlaces Universal

Acceso al texto completo, recursos web y bases de datos con un clic del ratón: LinkSolver de Ovid organiza el mundo de información a su alcance

Lanzamiento
inminente:
Funcionalidades
de Metabusador

www.ovid.com

spain@ovid.com +34 91 418 62 75

<i>ARC (A cross archive search service)</i>	Permite buscar información en varios repositorios <i>OAI</i> desde un único interfaz de búsqueda.
<i>ARNO</i>	Herramienta flexible para la implementación y gestión de repositorios institucionales. Permite unir estos repositorios con otros localizados en otras partes del mundo. Es capaz de funcionar como un gestor de contenidos. Necesita software adicional de terceros.
<i>CERN Document Server Software (CDSware)</i>	Permite construir un servidor de <i>pre-prints</i> propio, un catálogo en línea o un sistema de gestión documental en la web. Este sistema cumple con el protocolo <i>OAI-PMH</i> y usa <i>Marc21</i> como estándar bibliográfico subyacente. <i>CDSware</i> está diseñado para acomodarse a manejar grandes cantidades de información con varios formatos distintos.
<i>CiteBase</i>	Este recurso supone una fuente de información y de navegación basada en el impacto de artículos y de autores. Aparecen las referencias, las citaciones y las co-citaciones de un artículo. Además tiene la funcionalidad "correlation generator", que mide la correlación entre el impacto de sus citas y el impacto de sus descargas.
<i>Digital Commons (ProQuest)</i>	Herramienta compatible con <i>OAI</i> que permite construir repositorios institucionales personalizables y publicar una extensa variedad de tipos de publicaciones, incluidos documentos de conferencias, impresiones previas y posteriores, capítulos de libros, presentaciones, informes, monografías, archivos de vídeo, conjuntos de datos y programas ejecutables, así como otros recursos de "literatura gris" de forma rápida y sencilla.
<i>DSpace (MIT)</i>	<i>Managed services.</i> Captura, almacena, indexa, recupera y preserva en formato digital todo el material de investigación de grandes organizaciones como super-archivos. Es capaz de integrar contenidos distribuidos y de preservarlos. <i>End-user services.</i> Está diseñado para que las comunidades de usuarios envíen información y para que también accedan a ella. Fue creado para capturar la producción intelectual de organismos de investigación multidisciplinares.
<i>Eprints</i>	Software que ha sido implementado muy frecuentemente. Es relativamente sencillo y además cuenta con una red de apoyos. Se ha ido mejorando según las demandas locales hasta incorporar búsquedas avanzadas, metadatos extendidos y otros.
<i>Fedora</i>	Sistema flexible de gestión de repositorios y de bibliotecas digitales, con capacidad de gestionar de forma eficiente un millón de objetos. Tiene tres interfaces, una de administración del repositorio, otra para detectar y distribuir objetos en el repositorio y otra de acceso web.
<i>Greenstone</i>	Paquete de software para construir bibliotecas digitales. Permite organizar y publicar la información en internet independientemente de cuál sea el formato de los archivos o de dónde se encuentren ubicados. Permite integrar tanto archivos web, como cd-rom o intranets, que pueden ser consultados a través de un navegador
<i>i-Tor (NIWI)</i>	Puede presentar varios tipos de información a través de un solo interfaz web, independientemente de dónde se encuentren los archivos almacenados o de cuál sea su formato. En este repositorio independiente de los datos está separado el contenido de la interfaz de usuario. Funciona al tiempo como un Proveedor de servicios <i>OAI</i> y como un Proveedores de datos <i>OAI</i> . Es capaz de publicar datos de bases de datos relacionales, sistemas de archivos y sitios web. Permite a los usuarios añadir contenidos directamente a través de un navegador.
<i>Kepler</i>	Es un software que sirve para convertirse en un proveedor de datos <i>OAI</i> . Contiene un "archivelet" autoinstalable para autoarchivar publicaciones. Pretende integrar las bibliotecas digitales y los grupos de investigación, que así pueden tener un control constante sobre su biblioteca digital. Contiene dos posibilidades, una enfocada a grandes comunidades y otra para edición individual. En el caso de las comunidades se distribuye un software a cada miembro, un archivelet; después el servidor reúne toda la información distribuida en archivelets y realiza las búsquedas. En el caso de los editores individuales que quieran instalar un repositorio que cumpla con <i>OAI</i> , existen unos archivelets genéricos.
<i>Lockss</i>	Permite a las bibliotecas elaborar colecciones digitales con acceso permanente a su contenido. Es un software <i>peer-to-peer</i> que funciona como un sistema de preservación. La información se distribuye vía web y se almacena con un sistema fácil de usar.
<i>MarcXml</i>	Conjunto de programas <i>Java</i> que permiten la conversión de otros formatos a formato <i>Marc</i> y la conversión desde éste a otros formatos. Ha sido desarrollado por la <i>Library of Congress</i> y por <i>Marc Standards Office</i> ; constituye un marco para trabajar de forma flexible y extensible con <i>Marc</i> en un entorno xml. Contiene varios componentes, como <i>schemas</i> , hojas de estilo, y herramientas de software. Requiere <i>Java</i>
<i>MyCoRe</i>	Conjunto de herramientas para construir bibliotecas y archivos digitales. Consta de una sola aplicación basada en un núcleo funcional para implementar el repositorio; incluye búsquedas distribuidas en repositorios dispersos, integra audio y vídeo, editores de metadatos y demás. Se puede adaptar localmente a las necesidades de los usuarios y contiene instrucciones sobre cómo construir aplicaciones.
<i>My.OAI</i>	Interfaz muy completa de búsqueda de metadatos.
<i>NCSTRL</i>	Versión modificada de <i>Dienst</i> .
<i>OAIcat OCLC</i>	Aplicación en <i>Java</i> para dotar al repositorio de un marco modificable que se ajuste al protocolo <i>OAI-PMH</i> v2.0.
<i>OAIster</i>	Motor de búsqueda. Permite crear colecciones de recursos electrónicos académicos que pertenezcan a instituciones diferentes y que tengan formatos diferentes.
<i>OpCit</i>	<i>Open Citation Project</i> es una herramienta de enlace y análisis de citaciones para archivos abiertos.
<i>Rapid Visual OAI Tool</i>	Sirve para construir gráficamente un repositorio <i>OAI-PMH</i> a partir de una serie de archivos que contengan registros que estén en cualquier formato aceptable, como <i>Marc</i> . Permite pasar del formato nativo al formato <i>oai_dc</i> y una vez que esto se ha llevado a cabo, puede responder a las solicitudes <i>OAI-PMH</i> . Viene con un servidor http y un sistema de gestión de <i>OAI-PMH</i> . Está escrito en <i>Java</i> .
<i>Repository Explorer</i>	Contiene un interfaz que permite comprobar si un archivo cumple con el <i>OAI-PMH</i> , el protocolo para la extracción normalizada de metadatos.
<i>Zmarco</i>	Proveedor de datos que cumple <i>OAI-PMH</i> . La 'Z' significa Z39.50; 'Marc' significa <i>MACHine-Readable Cataloging</i> ; y la 'O' significa <i>OAI, Open Archives Initiative</i> , ya que <i>Zmarco</i> permite hacer que los registros <i>Marc</i> que están disponibles para un servidor Z39.50 lo estén también para <i>OAI-PMH</i> .

Tabla 2. Principales sistemas de software relacionados con repositorios

formación en formato *Unqualified Dublin Core*. De esta manera, los proveedores de servicios pueden recopilar los metadatos y ofrecer servicios complementarios a partir de esa información (tabla 1).

<http://www.dublincore.org>

Las instituciones del entorno académico pueden convertirse en proveedores de datos *OAI* de una forma relativamente sencilla, como la que sugiere *OAIster*, que anima a las instituciones potenciales de serlo.

<http://oaister.umdl.umich.edu>

2. Proveedores de servicios

Su función es usar los metadatos que han extraído por medio de *OAI-PMH* de los proveedores de datos y construir servicios de búsqueda complejos con un interfaz único:

—*ARC* (*a cross archive search service*): permite extraer metadatos y hacer búsquedas cruzadas entre archivos diferentes.

—*MyOAI*: proporciona un interfaz de búsqueda entre las distintas bases de metadatos *OAI* y funciones como *SDIs* o el historial de búsqueda.

—*OAIster*: posibilita las búsquedas de metadatos entre repositorios utilizando operadores de proximidad y otros mecanismos (McKiernan, 2004).

En cuanto a los proveedores de servicios *OAI* cabe destacar por la potencia de sus resultados *Scopus*, el buscador de la editorial *Elsevier*, que ha ampliado la cobertura de *Scirus*, el cual por su parte ya era el más potente buscador científico en internet. *Scirus* indizaba la mayor parte de las principales revistas y editores, bases de datos, más de 160 millones de sitios *.edu*, *.org*, *.ac.uk*, *.com*, *.gov*, de medicina, ciencia, tecnología, universidades, etc. Con ello, *Scopus*, se ha convertido en la mayor base de datos de sumarios

del mundo abarcando tanto recursos que requieren suscripción como gratuitos, repositorios, bases de datos y revistas con acceso abierto: *CogPrints*, *ArXiv.org*, *US Patent and Trademark Office*, *E-Print ArXiv*, *Chemistry Preprint Server*, *BioMed Central*, *PubMed Central*, *Scitation* y otros desde el año 1966, haciendo búsquedas simultáneas en todas las áreas de conocimiento, con herramientas de consulta bastante avanzadas.

<http://info.scopus.com>

<http://scirus.com>

«Es conveniente que los autores sepan qué editoriales permiten el archivo en repositorios, y para ello se puede acudir a webs como Romeo»

En cuanto al software abierto, las tecnologías más habituales son *Java*, *PERL*, *APS*, *PHP*, *GNU*, etc. Con relación a este tema es muy útil el informe detallado *Digicult report: core technologies for the cultural and scientific heritage sector* (2005), donde se explica la diferencia entre el *free software*, que se refiere al que es accesible públicamente, y el *open software* (*OSI*) de código abierto, donde el código fuente es mejorado de forma continua de cara a una mayor calidad de las aplicaciones (Zurita, 2003). Éste sería el modelo “bazar”, frente al sistema “catedral” del software propietario (Raymond, 1998). Las principales fuentes de información están en *Sourceforge.net* y en *Opensource.org* (tabla 2).

3. *OAI-PMH* para la extracción normalizada de metadatos

Un protocolo es un lenguaje utilizado para comunicar equipos entre sí. *OAI-PMH* por su parte está basado en *http* y en *xml*, y fue creado para la extracción de normalizada de los metadatos que están accesibles en los repositorios que

cumplen el estándar *Dublin Core* (Weibel, 2000). Por un lado, *xml* proporciona una sintaxis para la estructuración de documentos; por otro, los metadatos proporcionan la infraestructura para el proceso de preservación digital.

El proveedor de servicios lleva a cabo la extracción de los metadatos en los registros del proveedor de datos (figura 3) para construir servicios de búsqueda complejos.

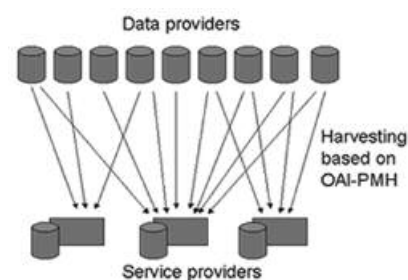


Figura 3. Varios proveedores de servicios pueden extraer información de varios proveedores de datos

OAI-PMH soluciona el problema inicial de los múltiples interfaces de búsqueda y el hecho de que no hubiera una manera automática de poner en común los metadatos correspondientes. Por ello se puso el énfasis en la compatibilidad. De esta forma se ideó la recopilación por parte de una máquina de los metadatos de los archivos, para proporcionar servicios centralizados de búsqueda.

<http://oforum.org>

El *Universal Preprint Service* (*UPS*), predecesor de *OAI*, como ya se ha dicho, supuso la demostración de que era posible crear una biblioteca digital que diera acceso a servicios basados en colecciones de metadatos recopilados en múltiples archivos, es decir, se había avanzado en la interoperabilidad y en la preservación. Desde la reunión de Santa Fe se demostró que mediante el protocolo *UPS* era posible hacer búsquedas cruzadas a partir de la colección de metadatos recopilados en múltiples archivos, lo que permitía determinar qué servidor utilizar en cada consulta y obviar proble-

mas en los lenguajes de interrogación, en la sintaxis y los atributos. Se se había dado un paso más hacia el desarrollo de marcos que permitiesen poner en común distintos dominios del conocimiento.

La arquitectura básica de la OAI es: el *Service provider* hace una petición al *Data provider* mediante http, el cual le responde enviándole los metadatos en xml gracias a *OAI-PMH* (figura 4).

Los archivos se pueden consultar directamente, pero podrán ser indizados por los motores de búsqueda como *OAIster* en la medida en que los datos cumplan el formato *OAI (OAI-compliant)*, que es la especificación del *Dublin Core Metadata Initiative (DCMI)* de quince elementos (*Unqualified Dublin Core*), adoptado como norma *ISO 15836*. Los metadatos proporcionados por los proveedores de datos no son obligatorios y todos son repetibles. Algunos de los metadatos que forman el *Dublin Core Metadata Element Set* son: *Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage* y *Rights*. Se suelen presentar divididos en tres grupos que indican la clase o alcance de la información que incluyen, con la

capacidad de describir cualquier objeto de información (**Méndez, 2002**). Con el avance en materia de homogeneidad por parte de *OAI* se cumple la razón de ser del *Dublin Core*: “definir marcos para la interoperabilidad de los conjuntos de metadatos” (**Weibel, 2000**).

Dublin Core era ya en 2000 el estándar *de facto* de metadatos en la web (**Weibel, 2000**). Sin embargo, desde *Los Alamos National Laboratory* sugirieron el uso de otros formatos como *MPEG 21 DIDL, METS* o *Scorm*, que tienen la capacidad de describir relaciones múltiples entre recursos.

La razón por la que es preferible el uso de *OAI-PMH* es que se lleva un registro de cada entrada de metadatos, pero no del recurso, de manera que al cambiar los metadatos cambian los registros. Si se modifican los recursos, también lo hacen los metadatos y los registros de los metadatos, de manera que hay una sincronización completa entre todos. Pero el protocolo no sólo controla los metadatos, sino que también los propios recursos. Por esta razón introducen formatos complejos de objetos para describir recursos digitales y expanden el ámbito de los metadatos más allá de *Dublin Core* o Marc a los llamados

complex object formats (21 DIDL, METS y Scorm), que en combinación con *OAI-PMH* proporcionan un marco para la recopilación de recursos digitales (**Van de Sompel et al., 2004**).

Enlaces y metadatos: OpenURL y CrossRef

Las urls no son enlaces persistentes. Por ello, *OpenURL* y *CrossRef* se complementan.

Los *DOIs (digital object identifiers)* constituyen la url permanente de un objeto digital, por ejemplo una imagen o un artículo. Sobre esta base se llevó a cabo el desarrollo de *CrossRef*, una iniciativa para implementar identificadores únicos para los recursos digitales que permitiesen su localización, independientemente de dónde se encontrase el recurso en internet (**Samuels, 2001**). La iniciativa *CrossRef* fue lanzada conjuntamente por los principales editores asociados expresamente en *PILA (Publishers International Linking Association, Inc.)* en el año 2000 para enlazar directamente las citas con los respectivos artículos mediante *DOI*.

Los editores usan un proceso automático para depositar los registros de los metadatos en la base de datos de *CrossRef*, donde cada registro debe incluir un *DOI* y una url, así como los metadatos siguientes: *Journal title, Issn, First author, Year, Volume, Issue, y Page number*. Después se deposita un registro con los metadatos y se registra el par *[doi-url]* en un directorio central (**Brand, 2004**). *CrossRef* incorporó una mejora en sus hiperenlaces llamada *multiple resolution* de los *DOIs*, por la que se pueden asociar elementos múltiples a cada *DOI*, de manera que al pinchar en un determinado enlace aparece un menú de posibilidades: todos los webs donde se encuentra el objeto o artículo buscado, otros formatos del mismo artículo, recursos relacionados... El directorio *DOI* debe conte-

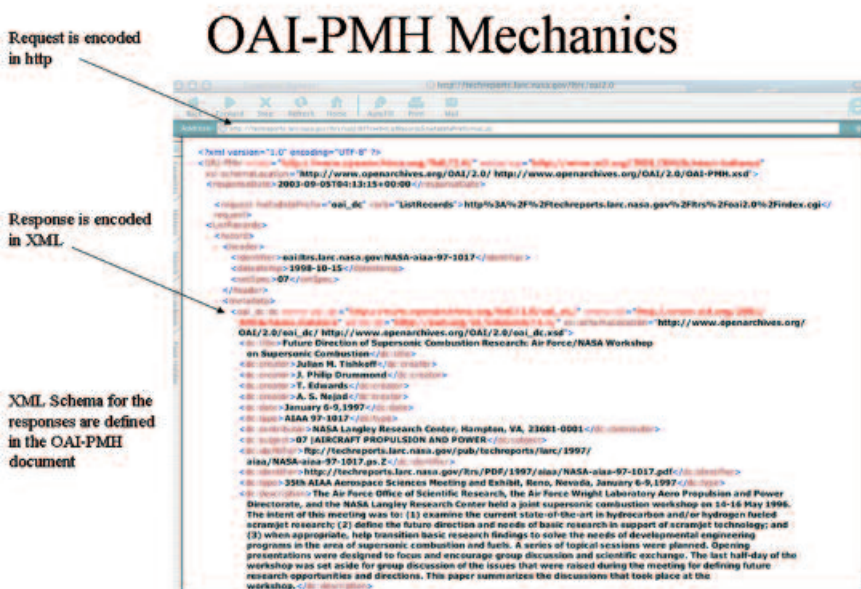


Figura 4. Funcionamiento del protocolo OAI-PMH (Nelson)



50



CUENTA AÑOS
GESTIONANDO
DOCUMENTACIÓN
CIENTÍFICA

www.cindoc.csic.es

Bases de datos del CSIC

Portales científicos

Biblioteca con más de 7.000 títulos de revistas científicas y 23.000 monografías

Suministro de documentos primarios

Revista Española de Documentación Científica

Búsquedas bibliográficas

Publicaciones especializadas

novedad
editorial

TESAURO ISOC DE
TOPÓNIMOS 2004

EDICIÓN
ABREVIADA



URBADOC. Base de datos vía web con más de 700.000 referencias bibliográficas sobre, arquitectura, urbanismo y ordenación del territorio, producida por la asociación Urbadata

www.urbadoc.com

La OpenURL siguiente: <i>http://LinkFinderPlus.library.edu?genre=article&issn=12345678&volume=99&issue=1&date=20020101&spage=27&atitle=What Is An OpenURL&title=Harry's White Papers</i>	
Se descompone en los siguientes metadatos:	
"genre=article"	Género del objeto descrito, un artículo de revista.
"&"	Sirve para separar los elementos de los metadatos entre sí. Aparece tantas veces como sea necesario.
"issn=12345678"	Issn de la revista que contiene el artículo, el 12345678.
"volume=99"	Volumen de la revista, el 99.
"issue=1"	Número de la revista, el 1.
"date=20020101"	Fecha de publicación del artículo, 20020101 (1 de enero de 2002).
"spage=27"	Página en la que comienza el artículo, la 27.
"atitle=What is an openURL"	Título del artículo, "What is an openURL".
"title=Harry's white papers"	Título de la revista, "Harry's white papers".

Tabla 3. OpenURL formada por metadatos

ner identificadores para todos los tipos de material (artículos, patentes, informes, imágenes) gracias al registro de múltiples agencias de registro de *DOIs* para que se produzca una red de enlaces del tipo *comprehensive linking network* (Brand, 2004).

«CrossRef fue lanzado por los principales editores agrupados en PILA en el año 2000 para enlazar directamente las citas con los respectivos artículos usando DOI»

Por su parte, *OpenURL* se desarrolló a raíz del enorme crecimiento de las colecciones de publicaciones electrónicas. También influyó el hecho de que los sistemas de información fuesen distribuidos, así como que las bases de datos se encontrasen tanto en un entorno local como remoto. Por todo ello se vio la necesidad de interrelacionar la información disponible para solventar los problemas derivados de que los links proporcionados por los proveedores no cubrían la totalidad de las colecciones y no eran sensibles al contexto del usuario. *OpenURL* constituye la solución a estos inconvenientes por ser un link que transporta los metadatos del registro a un servidor de enlaces, que proporciona un link que es sensible

al contexto, de manera que *OpenURL* transforma los metadatos y los identificadores en servicios (Nelson, 2003). Eso es precisamente *OpenURL*, una dirección web que contiene los metadatos del artículo con un formato determinado.

En los sistemas de enlaces basados en *OpenURL* la información sobre la citación de un recurso aparece codificada como una dirección web. Un servidor de enlaces puede construir direcciones web a partir de los metadatos; pero debe saber cómo obtenerlos, en qué formato aparecen en la url, y ése precisamente es el papel de *OpenURL*: ser un estándar para transmitir metadatos en una url, de manera que dos sistemas se los puedan mandar entre sí con una sintaxis específica, con un formato determinado (Samuels, 2001). En la tabla 3 se puede ver un ejemplo de cómo se lleva a cabo esta distribución; el enlace podría contener además el *DOI* del artículo. *OpenURL* puede funcionar de forma persistente mediante *DOI* sin tener que actualizar los metadatos. Se convierte de esta forma en un mecanismo de transporte de metadatos y de identificadores que describen una publicación para hacer un *context-sensitive linking*, es decir, enlaces sensibles al contexto del usuario.

OpenURL contiene, pues, tanto una url básica como metadatos; manda los metadatos a gestores de

enlaces (*link resolvers*) como *SFX* o *LinkFinder Plus*, que enlazan con el recurso o con el texto completo de los artículos (Samuels, 2001), dependiendo del contexto de los usuarios. Gracias a los gestores de enlaces, se pueden relacionar las *OpenURLs* entrantes y los privilegios locales de acceso de cada institución y desplegar así los enlaces de los recursos apropiados como catálogos, textos completos, motores de búsqueda, bases de datos, etc. (Brand, 2004), de tal manera que filtran el acceso que cada institución tiene a cada recurso.

Open Archival Information System (OAIS)¹

Ha sido publicado como una recomendación del *Ccsds (Consultative Committee for Space Data Systems)* y como norma *ISO 14721:2003*. *OAIS* focaliza su actividad en la preservación a largo plazo de la información científica en formato digital, como garantía de que será accesible en el futuro (Hirtle, 2001). Este doble propósito, conservación y acceso, parte del hecho de que la información sobrevive a su soporte y por ello es preciso adelantarse a la obsolescencia de la tecnología (Holdsworth, 1999).

Los rápidos cambios que se producen en la industria informática y la naturaleza efímera del almacenamiento electrónico de datos ha-

cen necesario planificar su preservación. Los registros pueden ser migrados (mediante refresco, replicación, re-empaquetamiento o transformación) hacia otros soportes o hacia diferentes entornos de software para que continúen siendo accesibles (*Ccsds 650.0-B-1*). Un sistema de información debe considerar el hardware, el software y los recursos humanos necesarios para la adquisición, preservación y difusión de la información. *OAIS* consiste en un modelo lógico que abarca todas las funciones de un repositorio digital, apuntando la forma en que los objetos digitales deben ser preparados, enviados a un archivo, almacenados durante largos períodos, conservados y recuperados. El contenido de la información está íntimamente ligado al formato y de la representación del conocimiento, junto a los inconvenientes de la preservación de los equipos y de los programas necesarios para acceder a los datos, que deben ser también conservados (**Abrams, 2003**).

«OpenURL es un url que contiene los metadatos del artículo»

El modelo de referencia *OAIS* lo desarrolló el *Ccsds*, un organismo internacional formado por las principales agencias espaciales, a propuesta de la *International Organization for Standardization (ISO)*. Corresponde al borrador de la norma *ISO reference model for an open archival information system* (**Hirtle, 2001**) y su propósito inicial fue para la gestión de la información aeroespacial, convirtiéndose luego en estándar para el archivo y acceso a la información en general.

El *Ccsds* estableció la terminología y los conceptos necesarios para describir los modelos de datos y la arquitectura, identificando las principales entidades y las relaciones que se dan en el entorno de un archivo. La parte del modelo de re-

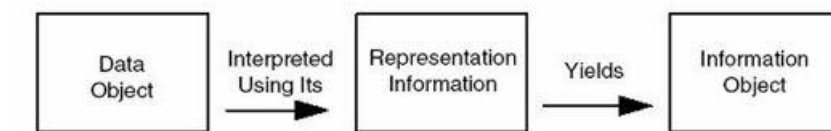


Figura 5. Obtención de la información a partir de los datos (*CCSDS 650.0-B-1*)

ferencia que ha sido relevante para la preservación ha sido el *OAIS information model*, que describe los requerimientos de los metadatos para que un objeto digital perdure (*Ccsds 650.0-B-1*).

En *OAIS* “información” se define como: “cualquier clase de conocimiento que puede ser intercambiado” y “es siempre expresada por algún tipo de datos” (*Ccsds 650.0-B-1*). La información puede existir de dos maneras: como un objeto físico en papel o como un objeto digital. En los dos casos se llama *data object* (objeto de datos), y puede ser el propio material que debe preservarse, o los metadatos asociados al mismo. El objeto de datos debe ser interpretado (figura 5) usando la información sobre su representación (*representation information*), lo que produce un objeto de información (*information object*).

Un “paquete de información” (*information package*) es un contenedor conceptual formado por dos componentes: la información sobre el contenido (*content information*) y la información sobre la preservación (*preservation description information, PDI*). El paquete de información tiene tres variantes: paquete de información de envío, de archivo y de difusión.

“Archivo”, se refiere a “una gran variedad de sistemas de almacenamiento y preservación”. Preservan el contenido y la autenticidad de la información, haciendo accesibles a la comunidad los registros y gestionando su propiedad, y aplicando criterios de aceptación de la información —que deben incluir, entre otros, la materia tratada, fuente, originalidad, técnicas utilizadas, formato, etc.—. Se deben tener en

cuenta varios aspectos, como las implicaciones de copyright y de propiedad intelectual, así como la capacidad para modificar la representación de la información o los acuerdos con organizaciones externas. El objeto de preservación es la información, que debe ser entendida de forma independiente de la comunidad que la ha producido. Se debe considerar de diferente manera la posesión física del contenido informativo de un documento y los derechos de propiedad intelectual de esa información.

OAIS es un modelo de referencia más que un plan de implementación: es un marco conceptual para un sistema completo de archivos que describe la arquitectura de los repositorios, pero no especifica cómo deberían ser implementados por las organizaciones, lo que le otorga gran flexibilidad. Por ello ha sido utilizado por diversas instituciones para el diseño de repositorios de distintos tipos: *Cedars Project*, *Harvard University’s Digital Repository*, *Library of Congress*, *Massachusetts Institute of Technology (MIT) D-Space/Repository schema-tic*, *OCLC Digital Archive*, etc.

Perspectivas del acceso abierto

La revolución digital conlleva cambios en el modo en que se lleva a cabo la investigación científica y en como se genera el conocimiento. Se redefinen la difusión, puesta en común y publicación de la información (**Grivell, 2003**), así como el ciclo del conocimiento. Se establecen nuevas relaciones entre las bases de datos, las hipótesis y los datos resultantes de la investigación (**Grivell, 2003**). La cadena documental de la información científica

ha sido ampliada, abarcando más allá de la entrada, tratamiento y salida de la información, debido a que las propias operaciones de “entrada”, “tratamiento” y “salida” son ahora muchas y muy distintas; los procesos de selección, análisis documental y difusión quedan trastocados por el hecho de que la información sea accesible antes de su publicación, que la de varias disciplinas sea accesible al mismo tiempo, y que la recuperación se vea mejorada por herramientas como la búsqueda por proximidad semántica entre documentos, búsqueda en distintas lenguas, formatos, repositorios, bases de datos, etc. Se redefine incluso el “documento”, que abarca información complementaria. La propia arquitectura de la información científica y técnica ha sido replanteada a causa de internet, en concreto, en el caso de las bibliotecas digitales, donde “OAI-PMH está llamado a ser lo que http es hoy al web” (Barrueco et al., 2003).

Se suele hablar de acceso abierto principalmente en campos como la biomedicina o la física, pero en el campo de las humanidades el acceso abierto significará una revolución sólo comparable a la aparición de la imprenta. Mientras que en las ciencias los conocimientos se superan y aparecen en forma de revisiones (*reviews*), en el campo de las humanidades es preciso citar todas las fuentes anteriores, ya que los conocimientos se acumulan, no se quedan obsoletos, e incluso pueden ser más valiosos. Por ello, iniciativas como el proyecto de Google de digitalizar quince millones de libros de bibliotecas, serán impactantes en el campo de las humanidades. Los autores nunca soñaron tener “todo, aquí y ahora”, algo que recuerda aquella biblioteca de Babel de Borges, “cuando se proclamó que la biblioteca abarcaba todos los libros, la primera impresión fue de extravagante felicidad...”.
<http://print.google.com>

En la medida en que el OA proporcione más visibilidad, rapidez, facilidad de uso, interoperabilidad, número de descargas por artículo y análisis de citas (Hubbard, 2005), permitirá dar idea de cuál es el impacto real de un artículo.

«Que OA triunfe no es sólo una cuestión de modelos económicos, ya que el carácter gratuito de esta iniciativa no será capaz por sí mismo de consolidarla»

Que el acceso abierto triunfe no es sólo una cuestión de modelos económicos, ya que el carácter gratuito de esta iniciativa no será capaz por sí mismo de consolidarlo. Depende más bien de que tenga prestigio entre la comunidad académica y entre los organismos que evalúan a dicha comunidad, por lo que la calidad es el componente decisivo. De nada sirven la interoperabilidad, los metadatos, los buscadores en archivos distribuidos, los gestores de enlaces, la tecnología capaz de resumir, de indizar y de procesar lenguaje natural, si no hay calidad en la información.

Suponiendo que el modelo OA llegue a triunfar —que sería lo deseable y, además, parece viable— se llegaría a otro gran problema: la sobreabundancia de información. No basta con que esté “todo” accesible, sino que es preciso que esté accesible “todo lo que necesito y sólo eso” en un determinado momento. Se puede decir que uno de los grandes impedimentos del OA va a ser la sobreabundancia de información, a menos que gran parte de la apuesta por las iniciativas de acceso abierto pase por las mejoras en técnicas de posicionamiento, procesamiento del lenguaje natural (NLP), contextualización, relevancia, análisis y refinamiento de las búsquedas. Procurar información

científica no sólo consiste en dar acceso abierto, sino en dar *acceso abierto pertinente* (*pertinent open access, POA*): éste es el gran reto.

Conclusiones

OAI focaliza su atención en la interoperabilidad y la rápida distribución de la información, en facilitar el acceso a los repositorios de *e-prints* para aumentar la disponibilidad de la información académica; mientras que OASIS centra su actividad en la preservación a largo plazo de la información como garantía de que será accesible en el futuro. A partir de iniciativas como OAI y OASIS es posible crear bibliotecas y archivos digitales que garanticen un acceso integrado, universal y a largo plazo a la información.

Estos avances tecnológicos constituyen un acontecimiento sin precedentes que posibilita la interconexión de las publicaciones científicas en forma de una verdadera tela de araña, donde a través de una publicación se pueden recorrer infinitamente todas las demás, como el *aleph* borgiano, uno de los puntos del espacio que contiene todos los puntos, algo propio de las ficciones hasta no hace mucho tiempo. Cada vez parece más viable la declaración de la IFLA sobre el acceso abierto de que la red mundial de bibliotecas y servicios de información proporciona acceso a las publicaciones académicas y de investigación pasadas, presentes y futuras y asegura su preservación.

<http://www.ifla.org>

Nota

1. No confundir OASIS Open Archival Information System con OASIS (Organization for the Advancement of Structured Information Standards), que es un consorcio internacional que se ocupa del desarrollo y adopción de estándares para el comercio electrónico y su seguridad. En origen se trataba de SGML Open, y estaba dedicado al intercambio de documentos sgml, mientras que en la actualidad está relacionado con el estándar xml.

<http://oasis-open.org>

<http://xml.org>

Bibliografía

Abrams, Stephen; Seaman, David. "Information technology and preservation and conservation workshop". En: *World library and information congress: 69th IFLA General conference and council*, 2003. Consultado en: abril-2005.
http://www.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf

Association of College & Research Libraries (ACRL). *Principles and strategies for the reform of scholarly communication*, 2003, agosto. Consultado en: marzo-2005.
<http://www.ala.org/ala/acrl/acrlpubs/whitepapers/principlesstrategies.htm>

Bailey, Charles. *Open access bibliography: liberating scholarly literature with e-prints and open access journals*. Association of Research Libraries (ARL), 2005. Consultado en: abril-2005.
<http://www.arl.org/pubscat/pubs/openaccess/OABweb.pdf>

Barrueco Cruz, José Manuel; Subirats Coll, Imma; *Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH): descripción, funciones y aplicación de un protocolo*. En *El Profesional de la Información*, 2003, v. 12, n.2, pp. 99-106
<http://www.elprofesionalde la informacion.com>

Brand, Amy. "Publishers joining forces through CrossRef". En: *Serial reviews*, 2004, v. 30, n. 1, pp. 3-9. Consultado en: marzo-2005.
<http://www.sciencedirect.com/science/journal/00987913>

Ccads 650.0-B-1: reference model for an open archival information system (OAIS). Blue book (standard). Issue 1. January 2002. Consultado en: abril-2005.
<http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>

Cedars guide to the distributed digital archiving prototype. Consultado en: abril-2005.
<http://www.leeds.ac.uk/cedars/guideto/cdap/guidetocdap.pdf>

Crow, Raym. *Guide to institutional repository software*. Open Society Institute, 2004. Consultado en: mayo-2005.
http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf

Cundiff, Morgan. "Towards an archival information package for audiovisual materials". Library of Congress. Consultado en: abril-2005.
<http://www.rlg.org/en/downloads/longterm/forum02/cundiff.html>

Fox, Edward. "Open archives initiative". En: *D-lib magazine*, 2000, v. 6, n. 6. Consultado en: abril-2005.
<http://www.dlib.org/dlib/june00/06inbrief.html#FOX>

newsWORKS

¿Recuerdas el artículo de la semana pasada?



¿A quién no le suena esta pregunta? La prensa, impresa o electrónica, es una de las fuentes de información más importantes para conocer los hechos, opiniones y tendencias. **newsWORKS** te ayuda a automatizar el proceso de digitalización, recorte, indexación y entrada de texto completo. Y casi simultáneamente se prepara el dossier de prensa diario: **3 clicks to go!**

Presentado por el socio tecnológico de newsWORKS:
ISIS THOT S.L.
Madrid - Barcelona - San Sebastián - Sevilla - Lima
☎ +34 91 307 76 41 ✉ isis@isis-spain.com



spain-info@ccs-gmbh.com

CCS

www.cca-gmbh.com

Franklin, Jack. "Open access to scientific and technical information: the state of the art". En: *Inist report*, 2002. Consultado en: abril-2005.
http://www.inist.fr/openaccess/en/etat_art.php

Garrett, John. *ISO archiving standards overview*. Consultado en: abril-2005.
<http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html>

Ginsparg, Paul; Luce, Rick; Van de Sompel, Herbert. *The open archives initiative aimed at the further promotion of author self-archived solutions*, 1999. Consultado en: febrero-2005.
<http://www.openarchives.org/meetings/Santa-Fe1999/ups-invitation-ori.htm>

Ginsparg, Paul; Luce, Rick; Van de Sompel, Herbert. *First meeting of the universal preprint service initiative*, 1999. Consultado en: febrero-

2005.
<http://www.openarchives.org/news/ups1-press.htm>

González Jesús; Seoane, Joaquín; Robles, Gregorio. *Introducción al software libre*, Barcelona, UOC, 2003,
<http://www.uoc.edu/masters/softwarelibre/esp/materials/libre.pdf>

Goodman, David. "The criteria for open access". En: *Serials review*, 2004, v. 30, n. 4, pp. 258-270. Consultado en: marzo-2005.
<http://www.sciencedirect.com/science/journal/00987913>

Goodman, David. "The criteria for open access". En: *Serials review*, 2004, v. 30, n. 4, pp. 258-270. Consultado en: marzo-2005.
http://eprints.rclis.org/archive/00003542/02/SR30_4_Goodman.pdf

- Grivell, Les; E-Biosci.** *Semantic networks of biological information*, (EMBO) 2003
<http://eprints.rclis.org/archive/00000912/>
- Hirtle, Peter.** "OAI and OAIS: what's in a name?" En: *D-lib magazine*, 2001, v. 7, n. 4. Consultado en: marzo-2005.
<http://www.dlib.org/dlib/april01/04editorial.html>
- Hitchcock, Steve.** "Metalist of open access e-print archives: the genesis of institutional archives and independent services". En: *ARL bi-monthly report*, 2003, April, n. 227. Consultado en: abril-2005.
<http://www.arl.org/newsltr/227/metalist.html>
- Holdsworth, David; Sergeant, D. M.** *A blueprint for representation information in the OAIS model*, 1999. Consultado en: abril-2005.
<http://www.personal.leeds.ac.uk/~eclhd/cedars/iee00.html>
- Hubbard, William.** *Open access institutional repositories in UK universities*, 2005. Consultado en: mayo-2005.
<http://eprints.rclis.org/archive/00003519/>
- IFLA statement on open access to scholarly literature and research documentation*, February 24, 2004. Consultado en: marzo-2005.
<http://www.ifla.org/V/cdoc/open-access04.html>
- Lagoze, Carl; Van de Sompel, Herbert; Nelson, Michael; Warner, Simeon.** 2002. *The Open archives initiative protocol for metadata harvesting, version 2.0*, 2002, June. Consultado en: marzo-2005.
<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>
- McKiernan, Gerry.** "Open content and access for digital scholarship". En: *WiLSWorld conference*, 2004. Consultado en: marzo-2005.
<http://www.public.iastate.edu/~gerrymck/OpenContent.ppt>
- Méndez, Eva.** *Metadatos y recuperación de información*. Gijón: Trea, 2002, Isbn 84-9704-055-4.
- Nelson, Michael.** *Institutional archives technology overview*, 2003. Consultado en: marzo-2005.
<http://www.cs.odu.edu/~mln/pubs/loc-ianelson.ppt>
- Nixon, William.** "Daedalus: freeing scholarly communication at the University of Glasgow". En: *Ariadne*, 2003, 14 Jan., n. 34. Consultado en: mayo-2005.
<http://www.ariadne.ac.uk/issue34/nixon/intro.html>
- Raymond, Eric.** *La catedral y el bazar*, 1998. Consultado en: abril-2005.
<http://sindominio.net/biblioweb/telematica/catedral.html>
- Ross, S., et al.** "Core technologies for the cultural and scientific heritage sector". *The Digi-CULT technology watch report 3*. Luxemburgo: Office for Official Publications of the European Communities, 2005, 92-828-5189-3. Consultado en: marzo-2005.
<http://www.digicult.info>
- Samuels, Harry.** *OpenURL: a tutorial*. Consultado en: abril-2005.
http://www.endinfosys.com/pdf/openurl4_02.pdf
- Samuels, Harry.** *The state of linking today and endeavor's linking solutions: Linkfinder and Linkfinder plus*, 2001. Consultado en: marzo-2005.
<http://www.endinfosys.com/ENCompass/linkingwhitepapers.pdf>
- Smith, Alastair.** *Citations and links as a measure of effectiveness of online LIS journals*. En: *IFLA journal*, 2005, v. 31, n. 1, pp. 76-84. Consultado en: marzo-2005.
<http://ifla.org/V/iftaj/IFLA-Journal-1-2005.pdf>
- Suber, Peter.** "Open access in 2004". En: *Spare open access newsletter*, 2005a, enero, n. 81. Consultado en: abril-2005.
<http://www.earlham.edu/~peters/fos/newsletter/01-02-05.htm>
- Suber, Peter.** *Open access overview: focusing on open access to peer-reviewed research articles and their preprints*, 2005b, marzo. Consultado en: febrero-2005.
<http://www.earlham.edu/~peters/fos/overview.htm>
- Weibel, Stuart.** "The Dublin Core Metadata Initiative: mission, current activities, and future directions". En: *D-lib magazine*, 2000, v. 6, n. 12. Consultado en: abril-2005.
<http://www.dlib.org/dlib/december00/weibel/12weibel.html>
- Zurita Sánchez, Juan Manuel.** "Software libre y libre acceso a la información: ingredientes para un ciberespacio público". En: *Proceedings XXXIV Jornadas mexicanas de biblioteconomía*, 2003, pp. 408-424. Consultado en: abril-2005.
<http://eprints.rclis.org/archive/00003480/>
- Van de Sompel, Herbert; Nelson, Michael; Lagoze, Carl; Warner, Simeon.** "Resource harvesting within the OAI-PMH framework". En: *D-lib magazine*, 2004, v. 10, n. 12. Consultado en: marzo-2005.
<http://www.dlib.org/dlib/december04/vandesompel/12vandesompel.html>
- Teresa Silió, Laboratorio de Internet, Cindoc-CSIC.**
Joaquín Costa, 22; 28002 Madrid.
tsilio@cindoc.csic.es