

EL PROFESIONAL DE LA INFORMACIÓN

Revista científica y profesional en español  
líder en el área de información, bibliotecas  
y nuevas tecnologías de la información.



## Observatorio

### Las ventajas de la apertura

Jesús M<sup>a</sup> González-Barahona

## Artículos

### Reutilización de tesauros: el documentalista frente al reto de la web semántica

Alberto García-Torres, Antonio Pareja-Lora y  
Daniel Pradana-López

### Evolución de la estructura científica española: *ISI Web of Science* 1990-2005

Benjamin Vargas-Quesada, Félix de Moya-Anegón,  
Zaida Chinchilla-Rodríguez, Elena Corera-Alvarez  
y Vicente Guerrero Bole (Grupo Scimago)

### Análisis métrico de los blogs españoles de biblioteconomía y documentación (2006-2007)

Daniel Torres-Salinas, Álvaro Cabezas-Clavijo y  
Emilio Delgado-López-Cózar

# SOQFT

Software libre  
para bibliotecas



## Análisis

### Software libre en repositorios de e-información

Ricard de la Vega Sivera

### Biblioteca virtual cooperativa descentralizada con software libre *Greenstone*

Dominique Babini, Florencia Vergara-Rossi,  
Flavia Medici y Jessica González

### Hacia una web semántica social

Jorge Morato, Sonia Sánchez-Cuadrado, Anabel  
Fraga y Valentín Moreno-Pelayo

### Ontología para el control y recuperación de información onomástica en televisión

Jorge Caldera-Serrano y Rodrigo Sánchez-Jiménez

## Software

### Isis para bibliotecas

Ricardo Eito Brun



RICHARD STALLMAN (GNU+COPYLEFT)



# El profesional de la información

Revista bimestral fundada en 1992 por Tomàs Baiget y Francisca García-Sicilia

## REDACCIÓN:

El Profesional de la información

Apartado 32.280

08080 Barcelona

epi@elprofesionaldelainformacion.com

## PUBLICIDAD:

Tel.: +34-609 352 954

publici@elprofesionaldelainformacion.com

## SUSCRIPCIONES:

El profesional de la información

Apartado 32.280

08080 Barcelona, España

suscripciones@elprofesionaldelainformacion.com

<http://www.elprofesionaldelainformacion.com/suscripciones.html>

Teléfono de atención al suscriptor  
+34 609 352 954

## SERVICIOS ONLINE:

Catorze.Com

Apartado 41

08272 Sant Fruitós de Bages

Tel. +34-650 839 200

javier@catorze.com

## DISEÑO:

Ignacio Pastor de la Huerta

ignacio@designio.es

## MAQUETACIÓN:

Jorge Liras

Romargraf, S.A.

## PRODUCCIÓN e IMPRESIÓN:

Romargraf, S.A.

Juventut, 55-57

08904 L'Hospitalet de Ll.

Tel. +34-933 345 466

romargraf@romargraf.es

## DISTRIBUCIÓN ONLINE:

MetaPress, Alabama, EUA

<http://elprofesionaldelainformacion.metapress.com>

metapress.com

Depósito legal: B-12303-97

Los trabajos de la sección "Artículos" son aprobados según el sistema tradicional "peer review": al menos dos expertos en el tema, del consejo asesor de la revista y/o externos, deben dar el visto bueno antes de su publicación.

Para conseguir que los trabajos no pierdan su actualidad, la dirección y los evaluadores de esta revista ponen especial esfuerzo en revisar los artículos con gran rapidez, consiguiendo un tiempo medio de aceptación o rechazo de los trabajos de sólo unas pocas semanas.

## Dirección editorial:

Tomàs Baiget <http://www.sarenet.es/baiget>

## Subdirector:

Javier Guallar [jguallar@gmail.com](mailto:jguallar@gmail.com)

## Redactor jefe:

Jesús Castillo Vidal [jesus.jcastillo@gmail.com](mailto:jesus.jcastillo@gmail.com)

## Coordinador editorial:

Carlos Tejada-Artigas [tejada@ccdoc.ucm.es](mailto:tejada@ccdoc.ucm.es)

## Redactores:

Lluís Codina <http://www.lluiscodina.com>

Elea Giménez-Toledo [elea@cindoc.csic.es](mailto:elea@cindoc.csic.es)

José Antonio Ontalba [joonrui@upv.es](mailto:joonrui@upv.es)

## Editor de sección:

Fernanda Peset [mpesetm@upv.es](mailto:mpesetm@upv.es)

## Colaboradores:

Ricardo Eito [reito@gmv.es](mailto:reito@gmv.es)

Jordi Grau-Moracho [jordi@grau.com](mailto:jordi@grau.com)

Javier Leiva-Aguilera <http://www.javierleiva.info>

Roser Lozano [rlozano@gencat.net](mailto:rlozano@gencat.net)

José Antonio Millán <http://jamillan.com>

Jorge Serrano-Cobos [jorgeserrano@gmail.com](mailto:jorgeserrano@gmail.com)

## Revisión de lengua inglesa:

Elaine M. Lilly [elaine@writersfirstaid.com](mailto:elaine@writersfirstaid.com)

## CONSEJO ASESOR

### Ernest Abadal

Facultat de Biblioteconomia i

Documentació.

Universitat de Barcelona. Barcelona.

### Isidro F. Aguillo

Centro de Información y Documentación

Científica (Cindoc). Consejo Superior

de Investigaciones Científicas (Csic).

Madrid.

### Ramon Alberch

Subdirector General de Archivos

Generalitat de Catalunya. Barcelona.

### Adela d'Alòs-Moner

Doc6. Barcelona.

### Ricardo Baeza-Yates

Depto. de Ciencias de la Computación.

Univ. de Chile. Santiago. Chile.

Yahoo! Research, Barcelona.

### Carlos Benito Amat

Servicio de Biblioteca y Documentación

Científica. Instituto de Agroquímica y

Tecnología de Alimentos, Csic.

Burjassot. Valencia.

### Jesús Bustamante

Biblioteca, CEDEFOP, Salónica, Grecia.

### Carlota Bustelo

Inforárea. Madrid.

### Emilio Delgado López-Cózar

Facultad de Biblioteconomía y

Documentación. Universidad de

Granada. Granada.

### Assumpció Estivill

Facultat de Biblioteconomia i

Documentació. Universitat de

Barcelona. Barcelona.

### Fco. Javier García Marco

Depto. de Ciencias de la

Documentación e Historia de la Ciencia.

Universidad de Zaragoza. Zaragoza.

### Paola Gargiulo

Consorzio per le Applicazioni di

Supercalcolo per Università e Ricerca.

(Caspur), Roma, Italia.

### Johannes Keizer

Food and Agriculture Org. (FAO)

United Nations, Roma, Italia

### Thomas Krichel

Palmer School of Libr. & Inform. Sci.

Long Island Univ., New York, USA

### Victoria Manglano

Ovid Technologies, Madrid.

### Charles McCathieNeville

Opera Software, Oslo, Norway

### Joan Roca

Dean of Library Services

Minnesota State University, USA

### Robert Seal

Loyola University Chicago

Evanston, Illinois, USA

### Ernesto Spinak

Consultor, Montevideo, Uruguay.

### Imma Subirats

Food and Agriculture Org. (FAO)

United Nations, Roma, Italia

### Jesús Tramullas

Depto. de Ciencias de la

Documentación e Historia de la Ciencia.

Universidad de Zaragoza. Zaragoza.

## Temas centrales: Software libre para bibliotecas / Ontologías

Nota de los editores:

Debido a la imposibilidad de publicar en EPI v. 16, n. 6 todos los originales aceptados sobre ontologías, una parte se publica ahora en EPI v. 17, n. 1.

### OBSERVATORIO

- 5 Las ventajas de la apertura  
*Por Jesús M<sup>o</sup> González-Barahona*

### ARTÍCULOS

- 8 Reutilización de tesauros: el documentalista frente al reto de la web semántica  
*Por Alberto García-Torres, Antonio Pareja-Lora y Daniel Pradana-López*
- 22 Evolución de la estructura científica española: ISI Web of Science 1990-2005  
*Por Benjamín Vargas-Quesada, Félix de Moya-Anegón, Zaida Chinchilla-Rodríguez, Elena Corera-Álvarez y Vicente Guerrero-Bote (Grupo Scimago)*
- 38 Análisis métrico de los blogs españoles de biblioteconomía y documentación (2006-2007)  
*Por Daniel Torres-Salinas, Álvaro Cabezas-Clavijo y Emilio Delgado-López-Cózar*

### ANÁLISIS

- 49 Software libre en repositorios de e-información  
*Por Ricard de la Vega Sivera*
- 56 Software libre en las bibliotecas de la UPC  
*Por Javier Clavero, Eulàlia Formentí y Toni Prieto*
- 64 Biblioteca virtual cooperativa descentralizada con software libre Greenstone  
*Por Dominique Babini, Florencia Vergara-Rossi, Flavia Medici y Jessica González*
- 71 Laboratorio virtual de software libre para bibliotecas  
*Por Patricia Russo-Gallo, Josep-Manel Rodríguez-Gairín y Andreu Sulé-Duesa*
- 78 Hacia una web semántica social  
*Por Jorge Morato, Sonia Sánchez-Cuadrado, Anabel Fraga y Valentín Moreno-Pelayo*
- 86 Ontología para el control y recuperación de información onomástica en televisión  
*Por Jorge Caldera-Serrano y Rodrigo Sánchez-Jiménez*
- 92 Ontología del patrimonio de Cantabria  
*Por Francisca Hernández-Carrascal*
- 99 Cómo realizar tests de usabilidad con personas ciegas  
*Por Mireia Ribera, Miquel Térmens y Maika García-Martín*
- 106 Administración electrónica y gestión documental. Consideraciones a la luz de la Ley para el Acceso Electrónico de los Ciudadanos a los Servicios Públicos  
*Por Carlota Bustelo y Elisa García-Morales*

### SOFTWARE

- 112 Isis para bibliotecas  
*Por Ricardo Eito Brun*

### RESEÑAS

- 119 E-LIS: una comunidad internacional de profesionales  
*Por Fernanda Peset*
- 120 Scientific publishing in the European research area  
*Por Fernanda Peset*

### AGENDA

### 124 INFORMACIÓN PARA LOS AUTORES

Los contenidos de **El profesional de la información** están referenciados en los siguientes servicios bibliográficos y bases de datos:

- Bedoc**  
<http://www.inforarea.es/bedoc.htm>
- Biblioteca de Andalucía**  
<http://www.juntadeandalucia.es/cultura/b/cgi-bin/abweb/L1/T1/S09>
- Bire**  
<http://gestion.madrid.org/bire/servlet/Servidor?opcion=ConsultarGeneral&tipoBiblioteca=R&itBibliobuses=>
- Bulletin Board for Libraries (Bubl)**  
<http://bubl.ac.uk/archive/journals/epdli/>
- Compludoc**  
<http://www.ucm.es/BUCM/complu>
- ConnectSciences (Pascal)**  
<http://connectsciences.inist.fr>
- Consorti de Biblioteques Universitàries de Catalunya (Cbuc)**  
<http://sumaris.cbuc.es/13866710.htm>
- Datathéke**  
<http://milano.usal.es/dtt.htm>
- Dialnet**  
<http://dialnet.unirioja.es>
- DoIS (Documents in information science)**  
<http://wotan.liu.edu/doi/data/julqichq.html>
- E-LIS (E-prints in library and information science)**  
<http://eprints.rclis.org/perl/search/advanced?=&publication=profesional+informacion>
- Ebscohost Electronic Journals Service**  
<http://ejournals.ebsco.com/direct.asp?JournalID=105302>
- Gobierno de la Región de Murcia**  
<http://www.carm.es/ctra/cendoc/bdatos/revistas/revista.asp?idrevista=290>
- Google Scholar**  
<http://scholar.google.com>
- GVA (Generalitat Valenciana)**  
[http://www.pre.gva.es/argos/va/contenido\\_general/recursos/bolsum/](http://www.pre.gva.es/argos/va/contenido_general/recursos/bolsum/)
- Índice español de ciencias sociales y humanidades (ISOC)**  
<http://bddoc.csic.es:8080/isoc.do>
- Information science and technology abstracts (Ista)**  
<http://www.epnet.com/thisTopic.php?topicID=91&marketID=1>
- Information Services in Physics, Electronics and Computing (Inspec)**  
<http://www.iee.org.uk/publish/inspec>
- IN-Recs**  
<http://ec3.ugr.es/in-recs>
- Library and information science abstracts (Lisa)**  
<http://www.csa.com/factsheets/lisa-set-c.php>
- Library, information science & technology abstracts (Lista)**  
<http://www.libraryresearch.com>
- MetaPress**  
<http://www.metapress.com/link.asp?id=105302>
- Oclc Firstsearch**  
[http://www2.oclc.org/oclc/fseco/topic\\_area.asp?topic=Z](http://www2.oclc.org/oclc/fseco/topic_area.asp?topic=Z)
- Registros Bibliográficos para Bibliotecas Públicas Españolas (Rebeca)**  
<http://www.mcu.es/REBECA/que.html>
- Social science citation index (ISI, Social SCI)**  
<http://go.isiproducs.com/>
- SwetsWise**  
[http://www.swetswise.com/link/access\\_db?issn=1386-6710](http://www.swetswise.com/link/access_db?issn=1386-6710)
- Universidad de Castilla-La Mancha**  
<http://biblioteca2.uclm.es/biblioteca/sumarios/pi.pdf>
- Universidad de Chile**  
<http://www.al-dia.cl/sistema/tablas/listar.asp?r=3199>
- Universidad de Oviedo**  
[http://librivation.uniovi.es/web/sumarios\\_web/Profesional-de-la-Informacion/](http://librivation.uniovi.es/web/sumarios_web/Profesional-de-la-Informacion/)

# MASmedios renueva la imagen de EPI

Gracias a la colaboración entre MASmedios y EPI se está renovando toda la imagen de la publicación. MASmedios pretende de esta forma aportar a EPI sus más de 10 años de experiencia en diseño gráfico y corporativo. La nueva imagen de la revista es el primer fruto de esta cooperación.



## PRÓXIMOS NÚMEROS

Marzo-abril  
Innovación en bibliotecas públicas  
Mayo-junio  
Presente y futuro de la profesión  
Julio-agosto  
Libros electrónicos  
Sept.-oct.  
Información en la empresa  
Nov.-dic.  
Redes sociales

El profesional de la  
**información**

**masmedios**

REDACCIÓN: epi@elprofesionaldelainformacion.com  
SUSCRIPCIONES: suscripciones@elprofesionaldelainformacion.com  
PUBLICIDAD: publici@elprofesionaldelainformacion.com  
Apartado 32.280, 08080, Barcelona Tel. +34-609 352 954

C/ Garcilaso 15 - 46003 Valencia  
Telf: 96 369 41 23  
Fax: 96 369 34 39  
www.masmedios.com

## Las ventajas de la apertura

Por Jesús M. González-Barahona

**Resumen:** Una de las principales consecuencias de publicar un programa como software libre es que su código fuente queda a disposición de quien quiera examinarlo. Además, mucha otra información sobre el proceso de desarrollo suele estar también fácilmente accesible en internet. En este artículo se presentan algunas de las consecuencias de esta “apertura” del proceso de construcción de software libre, y se explora brevemente cómo estos mismos mecanismos de apertura podrían comportarse en otros campos.

**Palabras clave:** Software libre, Software abierto, Código abierto

**Title:** The advantages of openness

**Abstract:** One of the main consequences of publishing a program as libre (free, open source) software is that its source code is made available to anyone who wants to look at it. In addition, much other information on the development process is often easily accessible on the internet. Some of the consequences of “opening up” the process of building libre software are discussed, briefly exploring how these open mechanisms could behave in other areas.

**Keywords:** Free software, Open source software, Open software

**González-Barahona, Jesús M.** “Las ventajas de la apertura”. En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, pp. 5-7.

DOI: 10.3145/epi.2008.ene.01



**Jesús M. González Barahona** es profesor en la Universidad Rey Juan Carlos (URJC) de Madrid. Comenzó a trabajar en la promoción del software libre en 1991 y colabora con varios proyectos de software libre (entre ellos Debian), con asociaciones como HispaLinux y EuroLinux. Asesora a empresas y administraciones públicas en sus estrategias relacionadas con este tema. Imparte asignaturas de doctorado sobre software libre en programas de la Univ. Politécnica de Madrid y la URJC y es miembro del comité científico del máster en software libre de la Univ. Oberta de Catalunya y codirector del master en software libre de la URJC, impartido en colaboración con Igalia y Caixa Nova. Ha participado y participa en proyectos internacionales: Calibre, FlossWorld, Flossdel software libre en 1991 y colabora con varios proyectos: FlossImpact.

### 1. Software libre, código abierto

El software libre lo es porque se puede estudiar, modificar, y también redistribuir como versión modificada<sup>1</sup>. Por lo tanto, el que un programa sea libre o no viene marcado por la licencia bajo la cual ha sido publicado, que es la que otorga (o no) estos permisos a quien reciba una copia.

Una de las principales consecuencias de esta definición es que el código fuente (que viene a corresponder, en líneas generales, con los “planos” detallados que indican cómo está construido el programa) tiene que estar disponible. En otras palabras, cualquier experto puede entender, con todo el detalle, cómo está construido el programa, hasta el punto de que podrá modificarlo sabiendo lo que hace. Naturalmente, este experto podrá también reconstruir completamente el programa a partir de ese código fuente, si así lo desea.

A primera vista, esta característica del software libre parece de interés sólo para los expertos que pueden

entender el código fuente, y saben cómo modificarlo, por ejemplo para corregir un error. Sin embargo es uno de los aspectos del software libre con más impacto en los usuarios finales. Veamos por qué...

### 2. El círculo virtuoso de la innovación distribuida

Cuando un usuario de software libre percibe un problema con un programa (quizás un funcionamiento erróneo, pero también una funcionalidad que echa de menos, por ejemplo), puede hacer algo más que quedarse esperando a ver si la situación mejora. Puede enviar un informe al grupo de desarrollo del programa. Pero también puede contratar a alguien que le arregle su problema. O puede arreglarlo él mismo si sabe trabajar con el código fuente del programa.

El que el código fuente esté disponible es justamente lo que permite que pueda elegir cualquiera de las

dos últimas opciones. O sea: la mejora del programa no depende sólo del grupo que hace el desarrollo, sino que cualquiera que sepa puede mejorar el programa, o si no tiene conocimiento para ello pero tiene recursos puede encargar que terceras partes realicen la mejora.

Y si algo de esto ocurre, quien realiza la mejora está además muy motivado para contribuir con ella de vuelta al grupo de desarrollo del programa, porque así es muy posible que tal mejora sea incluida en nuevas versiones, sin tener que volver a realizar el esfuerzo correspondiente. Y no hay que olvidar que estas nuevas versiones quedan a disposición de todo el mundo, por lo que hay un potente sentimiento de estar ayudando a un esfuerzo colectivo, incluso si el producto lo elabora una empresa o un grupo relativamente cerrado.

Así, gracias a que el código fuente está disponible, cualquiera puede innovar sobre el programa. Si el grupo de gente interesado en él es suficientemente grande, este mecanismo asegura un flujo constante y muy activo de mejoras e innovaciones. Y cualquier usuario del programa se beneficia de ellas (incluso si no sabe realizar ninguna de ellas).

Naturalmente, si el código fuente no estuviera disponible, nada de esto sería posible, y tendríamos el modelo de innovación centralizada clásico en la industria del software privativo: sólo quien produce un programa puede mejorarlo.

### 3. No sólo el código

El mecanismo descrito anteriormente se ve muy potenciado si no sólo el código, sino toda la información relacionada con el desarrollo del programa, está disponible para terceros. Los debates de diseño, los informes de error, la historia de modificaciones del programa, y otras fuentes de información permiten que cualquiera que se aproxime al proyecto pueda tener con relativa rapidez una visión muy detallada y completa de lo que está ocurriendo en él. Esto no sólo permite hacer evaluaciones detalladas (y muy exactas) de la “salud” de la comunidad que está manteniendo un programa, sino que también permite obtener fácilmente una gran cantidad de conocimiento sobre cómo se desarrolla.

Estos efectos son bien conocidos en la comunidad del software libre, y por ello la idea de “apertura” normalmente se extiende no sólo al código, sino a muchísimos otros aspectos de la producción del programa. Es importante darse cuenta de que, aunque el caso del código venía obligado por la licencia (si el software es libre, el código fuente ha de estar disponible), el publicar o no más información es algo que los proyectos pueden hacer o no, sin por eso dejar de ser software libre.

Los proyectos de software libre suelen tomar esta actitud abierta hacia la información hasta extremos poco concebibles en otras disciplinas, porque consideran que las ventajas de hacerlo son muchas. Son conscientes de que eso está ayudando a que terceros tengan más información sobre las características, y eso potencia el que se animen a realizar innovaciones, y a que esas innovaciones sean útiles, y en línea con las necesidades y requisitos del proyecto.

Cuanto más abierto es el proyecto, más fácil es captar recursos externos para tareas de mejora e innovación, así como atraer personas potencialmente interesadas que pueden acabar incorporándose al grupo de desarrollo, y atraer talento externo que estudie el programa, destaque fallos y problemas, y quizá incluso proponga soluciones.

Así, la cultura de la mayor parte de los proyectos de software libre incluye esta máxima apertura: “publicamos toda la información que podamos, porque eso nos beneficia”.

### 4. Aprender de lo que ya se sabe

Pero esto no es tan extraño. Los beneficios del escrutinio público son bien conocidos en otros dominios. Uno de los que primero viene a la mente es el ámbito científico. Se espera que los investigadores publiquen todos los detalles posibles de cualquier avance que hayan logrado, con la idea de que otros lo puedan reproducir, criticar, y si es posible, lo mejoren. Casi exactamente con los mismos efectos que en el mundo del software libre: mejora de la calidad, innovación distribuida, detección temprana de errores y problemas, etc.

Sin embargo, en uno de los dominios donde podría esperarse más aplicación, se encuentra todavía muy poca. Entre los productores de obras culturales libres (música, documentos, películas) es todavía muy poco habitual publicar información que no sea la propia obra liberada.

Algunos grandes proyectos, como la *Wikipedia*, sí se han preocupado (y se están preocupando cada vez más) de que toda la información relacionada con el proceso de creación colectiva quede documentado públicamente. Las páginas de discusión o las listas de correo públicas, por ejemplo, son grandes pasos en este sentido.

Pero la *Wikipedia* es hoy día más una excepción que una norma. Para muchos documentos libres que circulan por la Red es hasta difícil encontrar una versión editable (un formato que permita cambiarlo de forma simple). Es casi imposible conseguir los materiales separados (distintas pistas, por ejemplo) con los que

se ha editado una obra musical. Y no hablemos de los materiales fuente (planos, bandas de sonido, imágenes estáticas, etc.) para una película.

Hasta que todo esto esté disponible será difícil que la innovación por terceros ocurra sobre esas obras. Y si además, el proceso de desarrollo no está muy documentado públicamente (por ejemplo, las discusiones del equipo de creadores, las pruebas que dieron lugar a una cierta versión de una banda sonora, los borradores de los guiones de una película), más difícil será atraer talento externo que colabore en la mejora (o simplemente, en la evolución) de la obra.

En el caso de obras relativamente utilitaristas, como es el caso de la *Wikipedia*, quizá esto se está reconociendo antes, y por eso el camino está más avanzado. En las obras más puramente artísticas queda un largo camino por andar.

En estos aspectos, como en muchos otros, el software libre tiene ya una historia que merece repasar, para aprender ella. Y aprovecharse de lo aprendido allí...

### Notas

1. Esta definición rápida y poco rigurosa es una simplificación de la de “free software” de la *Free Software Foundation*, y de la de “open source software” de la *Open Source Initiative*, que son mucho más detalladas, y en términos generales definen el mismo concepto al que en este artículo me referiré como “software libre”.

2. Copyright 2008 Jesús M. González Barahona. Algunos derechos reservados. Este artículo se distribuye bajo la licencia “Reconocimiento-CompartirIgual 2.5 España” de *Creative Commons*, disponible en: <http://creativecommons.org/licenses/by-sa/2.5/es/deed.es> Este artículo y otros del mismo autor está disponible en: <http://sinetgy.org/jgb>

**Jesús M. González-Barahona**  
[jesus.gonzalez.barahona@urjc.es](mailto:jesus.gonzalez.barahona@urjc.es)



**DocuMenea** es un sistema de noticias sobre Biblioteconomía, Documentación, Archivística, Tratamiento de la Información, Periodismo, Internet y Nuevas tecnologías basado en el software de Menéame.

No estés ni un día más sin leer las novedades y votar las que consideres importantes para hacerlas más visibles:

<http://www.documenea.com>



The screenshot shows the DocuMenea website interface. At the top, there is a search bar and navigation links. Below, the main content area features a section titled 'últimas noticias' (latest news) with several news items, each with a date, a title, a short description, and a 'ver noticia' (view news) button. On the right side, there are two sidebars: one for voting on the latest news ('Vota las últimas noticias recibidas') and another for categories ('categorias') with a list of topics like 'Lecturas', 'Archivos', 'Bibliotecas', etc.

## Reutilización de tesauros: el documentalista frente al reto de la web semántica

Por Alberto García-Torres, Antonio Pareja-Lora y Daniel Pradana-López

**Resumen:** Se presenta una aproximación práctica a la construcción de ontologías a partir de tesauros inspirada en la metodología Methontology, que permite hacer uso de los lenguajes propios de la definición de ontologías y anotación semántica basada en las mismas. Asimismo es una alternativa a la publicación de tesauros en la web semántica mediante SKOS-Core, que es la propuesta del World Wide Web Consortium (W3C), y ofrece más semántica que el tesoro de partida y que su implementación mediante SKOS-Core. También se detalla una arquitectura genérica para la representación de dominios con dos niveles de descripción: una ontología de representación del conocimiento, basada en las facetas de Ranganathan (1967), para representar las relaciones de alto nivel del dominio, más una ontología de dominio, que proporciona la descripción de un campo particular del dominio.



**Alberto García Torres** es licenciado en biblioteconomía y documentación por la Universidad Complutense de Madrid. En la actualidad compagina los estudios de doctorado en ciencias y técnicas historiográficas con un proyecto de investigación en documentación histórica en el Tribunal Supremo. Sus líneas de investigación actuales van dirigidas hacia la archivística, la informatización de los archivos y la diplomática moderna.



**Antonio Pareja Lora** es licenciado en informática por la Univ. Politécnica de Madrid (UPM), donde cursa el doctorado en CC de la computación e inteligencia artificial. Es especialista en procesamiento del lenguaje natural y participa en el Ontological engineering group (OEG) de la UPM. Es profesor colaborador en el Depto. de Sist. Informáticos y Computación de la UPM. Sus líneas de investigación se encaminan a la representación del conocimiento lingüístico mediante ontologías y su aplicación a la notación semántica de textos.



**Daniel Pradana López** es licenciado en documentación por la Universidad Complutense de Madrid y diplomado en óptica y optometría por la misma universidad. En la actualidad está realizando estudios de doctorado en documentación mientras trabaja como especialista en anomalías visuales en su propia óptica. Sus líneas de investigación van dirigidas en estos momentos a la creación de tesauros y ontologías relacionadas con la optometría.

**Palabras clave:** Ontologías, Tesauros, Metodologías de desarrollo, Methontology, Arquitecturas de software, Web semántica.

**Title: Reusing thesauri: documentalists face the semantic web challenge**

**Abstract:** A practical approach to the development of ontologies from thesauri is presented, using the Methontology methodology, which permits the use of any language developed so far for the implementation of ontologies, as well as semantic web annotations based on these ontologies. This approach is an alternative to SKOS-Core, the World Wide Web Consortium (W3C) approach to publication of thesauri within the semantic web. The advantage of our approach is that it provides significantly more semantic information than a thesaurus (and, hence, than does SKOS-Core). A two-level ontological architecture for domain representation is also described, consisting of a knowledge representation ontology [based on Ranganathan's (1967) facets] for the representation of high-level relationships in the domain and a domain ontology for the description of a particular field within the domain.

**Keywords:** Ontologies, Thesauri, Software methodologies, Methontology, Software architectures, Semantic web.

**García Torres, Alberto; Pareja Lora, Antonio; Pradana López, Daniel.** "Reutilización de tesauros: el documentalista frente al reto de la web semántica". En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, pp. 8-21.

DOI: 10.3145/epi.2008.ene.02

## Introducción

La *www* ha sido y es una gran fuente de información accesible mediante nuestros ordenadores pero hasta ahora sólo comprensible para los seres humanos. En un principio los documentos de la Red nacieron orientados al intercambio informativo entre personas. Estos contenidos incluyen grandes cantidades de información (multimedia, en muchos casos) incomprensible para una computadora, siendo el usuario el principal encargado de encontrar, extraer e interpretar la información relevante.

Actualmente, dado el asombroso crecimiento de la web, resulta imposible que un único usuario realice esas tareas en un tiempo aceptable. Por otra parte, aunque se está facilitando el procesamiento automático de la información, los ordenadores todavía no pueden llevar a cabo con una precisión aceptable las tareas ya mencionadas de acceso, extracción e interpretación de la información relevante sin ayuda, pues no alcanzan el nivel de comprensión del lenguaje que poseen los seres humanos. Precisamente por ello, conseguir que los ordenadores entiendan el significado (la semántica) de los textos escritos y de las páginas web es el motor principal que impulsa el desarrollo de la web semántica (Berners-Lee, et al., 1999).

Existen dos elementos clave para lograr que un ordenador entienda el significado de un documento de la web. Por un lado, las ontologías (especificaciones formales y explícitas de una conceptualización compartida –Gruber, 1993; Borst, 1997; Studer, et al., 1998–), que establecen los términos relevantes del dominio (conceptos, atributos, relaciones, instancias, etc.); por otro lado, la anotación semántica de textos (Benjamins, et al., 1999; Motta, et al., 1999; Luke, et al., 2000; Staab, et al., 2000; Aguado, et al., 2002 (a; b; c); Benjamins, et al., 2003; Wilcock, et al., 2004), que hace explícito para un ordenador el significado de estos términos cuando aparecen en un texto, enlazán-

dolos con sus asociados dentro de las ontologías consideradas en la anotación, mediante los mecanismos de referencia apropiados de cualquiera de los lenguajes propios de la web semántica: *xml*, *rdfs*, *OWL*, etc. (Gómez Pérez; Corcho, 2002; Dean; Schreiber, 2003; Peis, et al., 2003).

### “La reutilización de tesauros es un interesante punto de partida para la construcción de una ontología”

Por lo tanto, cada vez más son necesarias ontologías que representen el conocimiento de los distintos dominios que componen la realidad, y que sirvan de base para la anotación de documentos y recursos de la web semántica. Puesto que partir de cero en la construcción de semejante red ontológica es a priori inviable, dado el alto coste que supone el desarrollo de una ontología, se hace necesario reutilizar todos los recursos disponibles. La reutilización de tesauros es pues un interesante punto de partida para la construcción de una ontología, que reduciría al menos los costes del proceso de educación de conocimientos, así como gran parte del esfuerzo de consenso que requiere una ontología. De hecho, según la clasificación de Lassila y McGuinness (2001), un tesoro se puede considerar ya como una ontología ligera (*lightweight ontology*, en inglés). Esta clasificación se recoge en la figura 1, tomada y traducida de Gómez Pérez, Fernández López y Corcho (2004), donde se ven claramente las características que deben añadirse a un tesoro para convertirlo en una ontología de peso (*heavyweight ontology*).

En el presente artículo presentamos una aproximación práctica a la construcción de ontologías a partir de tesauros inspirada en la metodología *Methontology* (Gómez Pérez; Fernández López; Corcho, 2004), que permite hacer uso de los lenguajes propios de la definición de ontologías y anotación semántica basada en las mismas, como alternativa a la publicación de tesauros en la web semántica mediante la propuesta del W3C, que es *SKOS-Core*. Como se muestra más adelante, nuestra propuesta proporciona más semántica que el tesoro de partida y que su implementación mediante *SKOS-Core*. Al final del artículo se proporciona el resultado de la aplicación de dicha

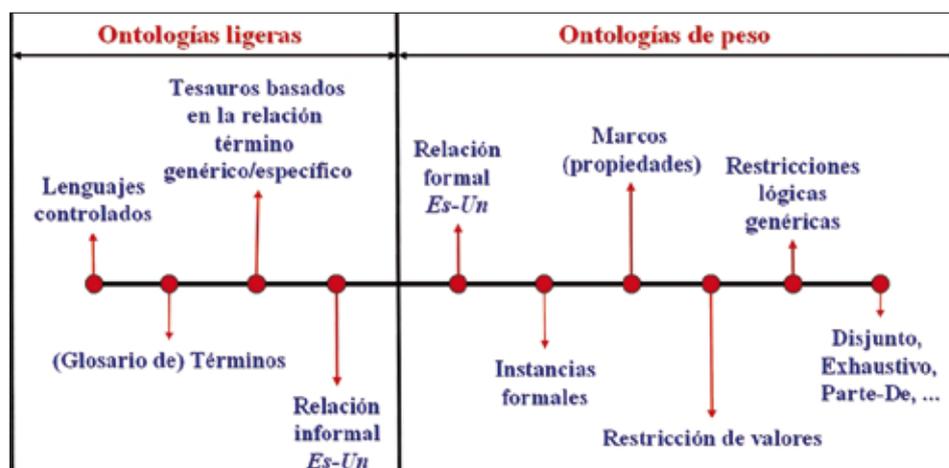


Figura 1. Clasificación de Lassila y McGuinness

metodología a un subdominio de la óptica. Asimismo, partiendo de un tesoro sencillo de dicho dominio, se detalla una arquitectura genérica en dos niveles de descripción: una ontología de representación del conocimiento para identificar dominios, basada en las facetas de **Ranganathan** (1967) –véase también **Gil Urdiciain** (1994)-, instanciada a modo de ejemplo para el dominio de la óptica, más una ontología que proporciona la descripción del tesoro en el campo de las anomalías oculares.

El resto del artículo está organizado como sigue: primeramente se presenta un breve estudio del estado de la cuestión en metodologías, herramientas y lenguajes de desarrollo de ontologías, así como de las propuestas formuladas hasta la fecha para la reutilización de tesoros en la web semántica; a continuación, se lleva a cabo un estudio comparativo de tesoros y ontologías, con sus diferencias y similitudes, así como las ventajas de las unas frente a los otros; seguidamente, se describe cómo aplicar la metodología *Methontology* al desarrollo de ontologías a partir de tesoros; posteriormente, se propone la arquitectura prototipo para la transformación de tesoros de un dominio en ontologías. Finalmente, se aportan unas conclusiones junto con las referencias bibliográficas empleadas.

### Estado de la cuestión

En el desarrollo de ontologías es de gran importancia la elección de un lenguaje apropiado con el que implementarlas (**Gómez Pérez; Corcho**, 2002; **Dean; Schreiber**, 2003; **Peis**, et al., 2003), pero éste es tan sólo uno de los pasos de un proceso complejo, que requiere de la definición y aplicación de una metodología propia de desarrollo, para evitar malgastar tiempo y esfuerzos.

Una metodología de desarrollo de ontologías es un sistema detallado de pasos que deben seguirse durante el proceso de construcción de una ontología, en un determinado orden, desde el momento de su concepción, a partir de las fuentes de las que se dispone para su construcción, hasta que queda finalmente implementada; idealmente debe contemplar también cómo se va a actualizar o evolucionar con el tiempo. Existen distintos tipos de metodologías de este estilo, en función de las fuentes de partida (**Gómez Pérez; Fernández López; Corcho**, 2004): las que describen cómo construir una ontología desde cero, como la ideada por **Grüniger y Fox**, *On-To-Knowledge*, *Methontology*, etc.; otras basadas en procesos de re-ingeniería de otras ontologías implementadas previamente, para adaptarlas a un nuevo entorno de aplicación (*Methontology*, por ejemplo, ha sido extendida para incluir esta posibilidad); teniendo en cuenta técnicas de aprendizaje automático (*ontology learning*); y también asentadas en la fusión de otras

ya existentes para el dominio (similares, pero distintas) en una unificada, que implica la determinación de la correspondencia de los términos de todas las consideradas (*ontology mapping*). Estos últimos puntos mencionados (el aprendizaje automático y la comparación de ontologías) centran gran parte de la atención hoy en día en la investigación en ingeniería ontológica, junto con la evaluación de ontologías, que también es parte del proceso de desarrollo de las mismas.

---

**“Una metodología de desarrollo de ontologías es un sistema detallado de pasos que deben seguirse en un determinado orden, desde el momento de su concepción hasta que queda finalmente implementada”**

---

Otra de las decisiones importantes que deben tomarse en el desarrollo de una ontología es si se desea utilizar una herramienta de las existentes para este fin o no. En caso negativo, esto requeriría generar a mano por completo el código xml, *Rdfs*, *OWL*, etc. de la ontología. En caso afirmativo, la herramienta se encargará de generar el código de implementación a partir de una representación gráfica y abstracta de la ontología, lo cual es bastante más cómodo y robusto. Entre las aplicaciones existentes, algunas restringen el lenguaje de implementación en el que quedará escrita la ontología (es decir, son dependientes del lenguaje), como *OntoSaurus*, *WebOnto* y *OilEd*, y otras permiten crear una ontología independientemente del lenguaje final de implementación, como *Protégé*, *OntoEdit*, *KAON* y *WebODE*. Esta última (*WebODE*) implementa la metodología *Methontology*, lo cual es una de sus grandes ventajas (**Gómez Pérez; Fernández López; Corcho**, 2004). En la figura 2 puede contemplarse un resumen gráfico de lo hasta aquí expuesto.

<http://webode.dia.fi.upm.es>

La principal alternativa a la creación de ontologías para la estructuración del conocimiento en la web semántica mediante anotaciones, es la iniciativa *SKOS* (*Simple Knowledge Organisation System*) del *W3C*. Es una concreción en lenguaje *Rdfs* de las etiquetas y relaciones que caracterizan un tesoro, útil para la publicación de los ya desarrollados, tras su implementación en el lenguaje *SKOS-Core* y/o haciendo uso de su extensión, *SKOS-Extended*. Para realizar esta codificación automáticamente existen al menos dos herramientas disponibles: el importador *SKOS* (**Pérez Agüera**, 2004), que permite convertir tesoros codificados en texto plano mediante marcados estándar (-, >, =>, etc.) en *SKOS-Core*; y la aplicación *Tematres* (**Ferreira**,

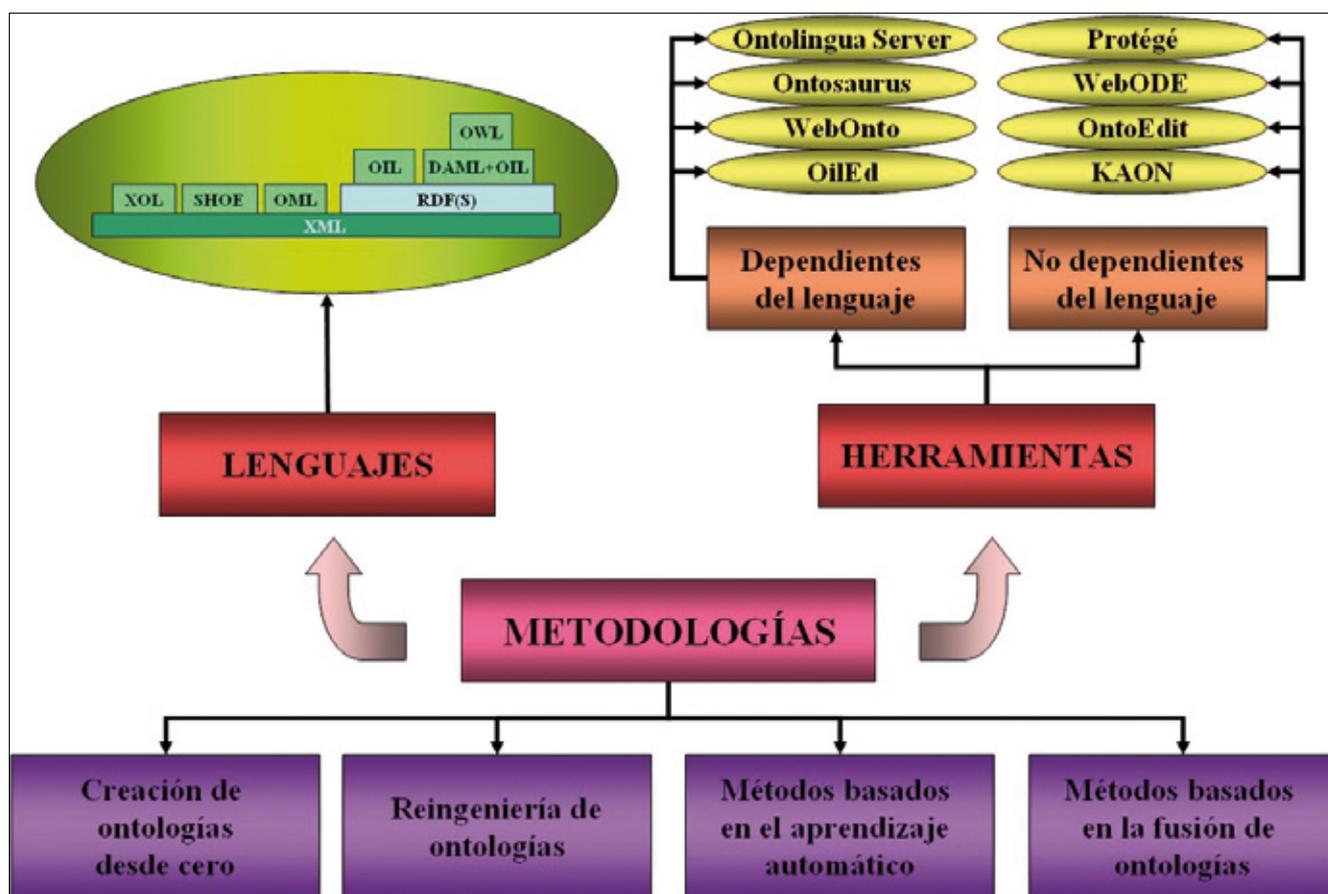


Figura 2. Desarrollo de ontologías. Estado de la cuestión

2007), que proporciona una implementación en *SKOS-Core* del tesoro al completo. La tarea de generar *SKOS* se puede llevar a cabo de forma automática si desarrollamos el tesoro con la propia aplicación, aunque también se puede importar en html, xml u otros formatos a su base de datos y, a partir de ella, generar el *SKOS-Core*. Sin embargo, el nivel de descripción semántica de un tesoro (aun implementado en *SKOS*), comparado con el de una ontología, es muy limitado, por lo que la formalización del conocimiento como ontología y no como tesoro aporta una serie de beneficios adicionales reseñados en el estudio comparativo de la sección siguiente.

<http://www.w3.org/TR/swbp-skos-core-guide>

<http://www.w3.org/2004/02/skos/extensions.rdf>

Es por eso por lo que en este artículo se presenta una concreción de *Methontology* para la construcción de ontologías a partir de tesauros, de forma que puedan ser reutilizados en el contexto de la web semántica.

**“El nivel de descripción semántica de un tesoro, comparado con el de una ontología, es muy limitado”**

Las ventajas son evidentes: se aprovecha el esfuerzo y el tiempo empleados en el desarrollo del tesoro; la mayor expresividad semántica de las ontologías deja abiertas las puertas para un uso más extensivo; hay ya disponibles herramientas de desarrollo (incluyendo su implementación) de ontologías, lo que facilita esta engorrosa tarea; y sólo hay que adaptar alguna metodología de elaboración de ontologías existente cuando el punto de partida sea un tesoro, para convertirla en una tarea sistemática y robusta.

### **Tesauros y ontologías: un estudio comparativo**

Las similitudes existentes entre ambos sistemas hacen que la conversión de unos a otros sea sencilla, siempre y cuando se tengan en cuenta las diferencias existentes entre ellos (véase la tabla 1).

En primer lugar, las instancias, que son cada uno de los objetos particulares que se incluyen dentro del dominio que estamos modelando, son elementos básicos en ontologías que no poseen correspondencia en los tesauros. El resto de elementos de una ontología pueden hallarse especificados unas veces, e implícitos (camuflados) otras, entre los distintos componentes de un tesoro.

Las siguientes entradas de la tabla 1 son los descriptores y los no descriptores. En *Methontology* (y, por ende, en *WebODE*) la información de cada concepto en una ontología incluye los campos: nombre del concepto, sus sinónimos, acrónimos, instancias, atributos de

clase y atributos de instancia, más las relaciones en las que participa. Los descriptores (o términos preferentes en *SKOS-Core*) son implementados dentro de una ontología, en general, como el nombre del concepto o de la clase. En el caso excepcional de que el descriptor fuera

Tesauros	Definición	Ejemplos	Ontologías	Definición	
-	-	-	Instancia	Objeto particular del dominio que se está formalizando	
Descriptor (SKOS: término preferente)	Término representativo de un concepto	Globo ocular	Nombre del concepto (clase)	Identificador del concepto o clase dentro de la ontología	
			Acrónimo (en caso de que se decida elegir como descriptor el desarrollo de unas siglas)	Siglas del identificador del concepto en caso de existir	
No descriptor (SKOS: término no preferente)	Término no representativo de un concepto (sinónimos, cuasi-sinónimos, etc.)	Ojo	Sinónimo	Término alternativo para referirse al concepto	
			Nombre del concepto (si el descriptor es un acrónimo)	Véase más arriba	
Nota aclaratoria (NA)	Nota utilizada para la contextualización de cada término	Anomalía ocular NA Úsese en caso de referirse a aspectos relacionados con anomalías de refracción o estructurales	Descripción	Nota utilizada para la contextualización de cada término (definición en lenguaje natural, etc.)	
			Nombre del concepto (desarrollo de un acrónimo, en el caso de que se elija éste como descriptor)	Véase más arriba	
			Atributo	de clase	Propiedad que adopta un mismo valor para todas las instancias de una misma clase
				de instancia	Propiedad particular para cada una de las instancias incluidas en una determinada clase
			Definición del tipo de datos de atributos	El tipo de datos de atributos es el conjunto de valores que puede adoptar una propiedad (atributo) de una clase o concepto (entero, real, booleano, cadena de caracteres, etc.)	
			Restricción del rango de valores de atributos	El rango de valores que puede adoptar una propiedad puede limitarse en función del dominio	
			Regla	Afirmación de la lógica de primer orden del tipo $P \rightarrow Q$ que debe entenderse como: si se verifica P, entonces se verifica Q; o bien, regla de producción de la forma $P_1 \dots P_n \Rightarrow A_1, \dots, A_n$ donde las premisas $P_i$ son condiciones que deben verificarse simultáneamente y las $A_i$ son acciones que se realizan si se verifican dichas premisas	
Axioma formal	Afirmación siempre cierta dentro del dominio, expresada como una fórmula de la lógica de primer orden				

Tabla 1. Comparativa entre los elementos de un tesauro y de una ontología

Tesauros		Definición	Ejemplos	Ontologías	Definición
Relación de equivalencia	Usado por (UP)	La relación usado por se establece desde el descriptor al no descriptor mientras que use se establece desde el no descriptor al descriptor	Globo ocular	Methontology: representaciones intermedias (tareas 1, 6, 7, 9 y 10)	
	Use (USE)		UP ojo Ojo USE Globo ocular		
Relación jerárquica	Término genérico (TG)	Hiperónimo de otros conceptos	CÓRNEA TG Globo ocular	Relación IsA	El concepto C1 es una subclase de otro concepto C2 si y sólo si todas las instancias de C1 lo son también de C2
			Miopía magna TG Miopía	Descomposición disjunta	División de un concepto C en subclases, de forma que éstas no tienen instancias en común; puede haber instancia de C que no pertenezcan a ninguno de los conceptos de la descomposición
	Término específico (TE)	Cada uno de los hipónimos de otro concepto	Globo ocular TE Córnea	Descomposición exhaustiva	División de un concepto C en subclases, de forma que todas las instancias deben pertenecer al menos a una de las subclases de C, pero puede haber instancias que pertenezcan a más de una subclase
			Miopía TE Miopía magna	Partición	División de un concepto C en subclases, de forma que todas las instancias deben pertenecer a una y sólo una subclase de C
			Relación PartOf	Relación metonímica del todo a cada una de sus partes a un todo	
Relación asociativa	Término relacionado (TR)	Relacionan descriptores que no suponen sinonimia ni jerarquía; se trata de una relación de afinidad semántica	Miopía TR Globo ocular	Relaciones ad hoc	Relaciones que se establecen entre conceptos, pero que no se ajustan a ninguno de los tipos de relaciones jerárquicas anteriores
				Methontology: representaciones intermedias (tareas 1, 6, 7, 9 y 10)	

Tabla 1 (continuación). Comparativa entre los elementos de un tesoro y de una ontología

un acrónimo será reflejado en la ontología en el campo acrónimo de un concepto o de una clase.

Los no descriptores (o términos no preferentes en *SKOS-Core*) se implementarán en las ontologías dentro de los campos de descripción citados en la tabla 1. Es decir, en el caso de que el no descriptor fuera sinónimo de un descriptor, el primero será incluido en el campo de sinónimos del concepto que representa al descriptor en la ontología. En el caso de que el no descriptor fuera el desarrollo de un acrónimo, éste se verá reflejado en la ontología como el nombre del concepto.

Las notas aclaratorias utilizadas en los tesauros pueden implementarse en las ontologías de diferentes formas, atendiendo a las funciones que cumplen dentro del tesoro:

1. Si es utilizada para contextualizar el descriptor a través de una definición en lenguaje natural, quedará

reflejada en la ontología dentro de la descripción asociada al concepto (que puede verse en el glosario de términos de la misma).

2. En el caso de que sea el desarrollo de un acrónimo, como ya vimos más arriba, se incluirá en la ontología precisamente como el nombre del concepto.

3. Si se refiere a las restricciones de aplicación o de uso de un descriptor, se plasmará en la ontología a través de:

3.a. La definición de un atributo propio de un concepto. Los atributos son las propiedades o características que identifican cada concepto o instancia. Existen dos tipos de atributos: si un valor del atributo caracteriza a un grupo de instancias (lo que conocemos como clase), se trata de un atributo de clase; por el contrario, si el valor del atributo es propio de cada instancia, se trata de un atributo de instancia.

3.b. La definición del tipo de un(os) cierto(s) atributo(s) y/o,

3.c. La restricción del rango de valores determinada por el tipo de esos u otros atributos.

4. Si la nota se refiere a relaciones entre valores de atributos de uno o varios conceptos, por ejemplo, se verán implementadas en una ontología como reglas. Por otro lado, es deseable que una nota aclaratoria que contiene una definición en lenguaje natural del término se traduzca en la medida de lo posible en un axioma, pues en ellos y en las reglas reside en gran parte la potencia razonadora de los mecanismos de inferencia basados en ontologías.

Las relaciones de equivalencia que existen en el tesoro entre los descriptores y los no descriptores no ven reflejadas directamente en las ontologías, sino que se recogen en el glosario inicial de la metodología de construcción *Methontology*, por ejemplo, donde se establecen sinónimos y acrónimos de cada uno de los términos admitidos. También pueden recogerse en el resto de las representaciones intermedias en las que se dan los atributos de clase, de instancia y en la definición de los propios axiomas y reglas especiales.

Un caso mucho más interesante son las relaciones jerárquicas del tesoro, que se establecen entre los términos más generales y los más específicos del dominio. En lo que respecta a las ontologías, existen cinco tipos de relaciones para concretar estas relaciones, en función de la semántica con la que se usen dentro del tesoro: la relación *IsA* (o *SubClaseDe*), la descomposición exhaustiva, la descomposición disjunta, la partición y la relación *PartOf* (*Parte-Todo*), todas ellas explicadas más adelante, en el apartado dedicado a la descripción de la metodología.

Por último, en los tesoros tenemos relaciones asociativas entre descriptores que expresan cierta afinidad semántica entre conceptos distintos. En el caso de las ontologías se concretarían como relaciones ad hoc entre términos afines. No obstante, algunas de las funciones de las relaciones asociativas de **Aitchison, Gilchrist** y **Bawden** (1997) como, por ejemplo, aquella que se recoge en una ontología como la relación entre un concepto y sus atributos, se verán reflejadas en el glosario inicial y en las representaciones intermedias resultado de las tareas en las que se detallan los atributos de clase y de instancia, los axiomas y las reglas (explicadas asimismo en la metodología).

A partir de todo lo expuesto anteriormente, puede deducirse fácilmente que la formalización del conocimiento como una ontología y no como tesoro aporta los seis beneficios adicionales siguientes:

– En la creación de una ontología se utilizan modelos matemáticos, lo que la dota de mayor precisión frente al tesoro.

– A diferencia de los tesoros, se pueden añadir elementos concretos denominados instancias, que aportan una información esencial para la extracción de información (por ejemplo, las personas concretas que realizan una acción, con su lugar y fecha de realización).

– Se obtiene un desarrollo semántico más profundo (**Ding; Foo**, 2002; **García-Jiménez**, 2004), pues los términos son desarrollados en toda su extensión, proveyéndolos de atributos (propiedades) y también se formalizan y detallan más aún las relaciones jerárquicas o asociativas de un tesoro, lo que provoca que un concepto esté mucho más definido dentro de la globalidad del vocabulario.

– El conocimiento formalizado puede reutilizarse dentro de un dominio, por definición, al ser previamente consensuado, y se posibilita su incorporación en distintas aplicaciones. Es decir, al describir objetos del mundo reales, y no sólo conceptos o ideas, éstos pueden ser incluidos en diversos sistemas a la vez (**Qin; Paling**, 2002; **García-Jiménez**, 2004).

– Por ser consensuada, una ontología favorece una representación normalizada de la información (**Ding; Foo**, 2002; **García-Jiménez**, 2004).

– La utilización de las ontologías permite la comunicación entre humanos y máquinas (así como entre las propias máquinas o agentes entre sí –**Ding; Foo**, 2002; **García-Jiménez**, 2004-); además, la inclusión de axiomas y reglas allana el camino para la concreción y aplicación de mecanismos de inferencia y razonamiento de la web semántica basados en ontologías, lo cual a su vez permitirá mejorar la precisión de los mecanismos de extracción y recuperación de información.

### **Aplicación de *Methontology* al desarrollo de ontologías a partir de tesoros**

El proceso de construcción de una ontología puede verse simplificado mediante la utilización de la metodología apropiada. Existen varias aplicables, pero reco-

---

**“La utilización de ontologías permite la comunicación entre humanos y máquinas, así como entre las propias máquinas o agentes entre sí”**

mendamos tanto por su sencillez como por su probada eficacia el uso de *Methontology* (Gómez-Pérez, Fernández-López, Corcho, 2004). Consta de once pasos (o tareas) que se han de realizar de forma secuencial, y que son descritos a continuación ilustrándolos mediante ejemplos tomados de un tesauro en el dominio de la óptica y la optometría actualmente en construcción. Este tesauro hace un uso abundante de las notas aclaratorias con toda su capacidad expresiva, según Aitchison, Gilchrist y Bawden (1997), de forma que todo el conocimiento que no es capaz de captar quede en cierta manera ya recogido en él informalmente, pues el objetivo final es la construcción de una ontología de ese dominio.

### Tarea 1. Construcción de un glosario de términos

Se desarrollan los términos relevantes propios del dominio, indicando su nombre, sinónimos, acrónimos, descripción (en lenguaje natural) y el tipo de término que constituye dentro de una ontología (si es un concepto, una instancia, un atributo, una relación entre conceptos, una constante, etc.).

Si disponemos de un tesauro como punto de partida, en este glosario se incluirá cada descriptor como el nombre de un nuevo concepto; los no descriptores, bien como sinónimos o bien como nombre de un concepto, en el caso de que se eligiera un acrónimo como descriptor (en cuyo caso, dentro de la ontología, el descriptor se transformaría en un acrónimo del concepto cuyo nombre es el desarrollo del acrónimo), algunas de las notas aclaratorias (que se transcribirán como una descripción del concepto en la ontología, si contienen una definición, como un atributo si identifican una pro-

#### Ejemplo

A partir del descriptor "Lente de contacto" y de los no descriptores "lentilla" y "LC", y del descriptor "Miopía magna" y su nota de aplicación:

Miopía magna

NA Úsese para referirse a aquella miopía con una graduación superior a 6,00 dioptrías.

podemos consignar dentro del glosario de la ontología el nombre de los conceptos "Lente de contacto" y "Miopía magna", que cada uno de ellos es un "Concepto" (en el campo de "Tipo" de término), que "Lentilla" es un "Sinónimo" de "Lente de contacto", que las siglas "LC" son un "Acrónimo" del concepto al que se refieren ("Lente de contacto"), y que la "Descripción" de una "Miopía magna" es: "Aquella miopía con una graduación superior a 6,00 dioptrías".

iedad asociada a un concepto o a las instancias del mismo) y las relaciones asociativas, a las que habrá que dar primeramente un nombre a partir de su semántica y de los términos relacionados en el tesauro.

### Tarea 2. Construcción de taxonomías

Las cuatro relaciones taxonómicas que se emplean para formalizar una ontología de peso son:

- Subclase de (*IsA*): diremos que el concepto  $C_1$  es una subclase de otro  $C_2$  si, y sólo si, todas las instancias de  $C_1$  lo son también de  $C_2$ . Ejemplo: una opacidad ocular es una subclase de anomalías estructurales.

- Descomposición disjunta: es una división de un concepto  $C$  en subclases, de forma que éstas no tienen instancias en común; puede haber instancias de  $C$  que no pertenezcan a ninguno de los conceptos de la descomposición. Ejemplo: la miopía puede dividirse en miopía normal ( $0 \leq \text{graduación} < 3$  dioptrías), de riesgo ( $3 \leq \text{graduación} < 6$  dioptrías) y magna (graduación  $\geq 6$  dioptrías). Evidentemente, un ojo miope no puede pertenecer a la vez a dos subclases de miopía.

- Descomposición exhaustiva: es una división de un concepto  $C$  en subclases, de forma que todas las instancias deben pertenecer al menos a una de las subclases de  $C$ , pero puede haber instancias de  $C$  que pertenezcan a más de una subclase. Ejemplo: el caso anterior también constituye una descomposición exhaustiva pues, por su forma de definición, no puede haber más subclases de miopía que la normal, la de riesgo y la magna.

- Partición: se define como una división de un concepto  $C$  en subclases, de forma que todas las instancias deben pertenecer a una y sólo una subclase de  $C$ . Las particiones son, por lo tanto, descomposiciones exhaustivas y disjuntas a la vez. Ejemplo: puesto que los tipos de miopía constituyen una descomposición exhaustiva y disjunta, también forman una partición del concepto miopía.

Además de estas relaciones taxonómicas, conviene distinguir, asimismo, las metonímicas, o de tipo "Parte De" (*PartOf*), que se establecen entre cada una de las partes y el todo. Pueden ser de dos tipos: transitivas o intransitivas. Las primeras son las que cumplen la afirmación: "si un concepto  $C_1$  es parte de otro concepto,  $C_2$ , y éste lo es de otro concepto,  $C_3$ , entonces  $C_1$  es parte de  $C_3$ "; e intransitivas las que no la cumplen. Ejemplo: el iris es parte del ojo; el ojo es parte de mi cabeza; como además se verifica que el iris es parte de la cabeza, se llega a la conclusión de que estas relaciones "Parte De" son transitivas.

Si ya disponemos de un tesauro, las relaciones jerárquicas existentes nos ayudarán a obtener las distintas relaciones taxonómicas dentro de la ontología.

### Ejemplo

Si contamos con un tesoro con los siguientes descriptores y relaciones jerárquicas:

	Miopía
TE	Miopía magna
TE	Miopía normal
TE	Miopía de riesgo

y sabemos que sólo existen estos tipos de miopía y que son incompatibles entre sí, por las notas aclaratorias de cada tipo de miopía que restringen la graduación de cada tipo, a partir de estas relaciones jerárquicas sabemos que hay que incluir una partición para representarla dentro de la ontología.

### Tarea 3. Construcción de diagramas de relaciones binarias ad hoc

Se entiende por relación ad hoc cualquiera que no pertenezca a las especificadas en el apartado anterior. Puede establecerse entre conceptos de la misma o de distinta taxonomía de conceptos. En el caso de disponer de un tesoro, vendrán desarrolladas a partir de las relaciones asociativas de las que disponemos.

### Ejemplo

	Gafas
TR	Anomalía de refracción

En nuestra ontología podremos incorporar la relación ad hoc "Compensa", que se establece entre el concepto "Gafa" y "Anomalía de refracción"; su relación inversa sería "Compensada por", que se establece entre el concepto "Anomalía de refracción" y "Gafa".

### Tarea 4. Construcción de un diccionario de conceptos

En él especificaremos cuáles son las propiedades (los atributos) y las relaciones que describen cada concepto, así como sus instancias características. Es de vital importancia explicar el nombre del concepto, sus instancias, los atributos de clase (aquellos cuyo valor caracteriza a los individuos de la misma clase), los atributos de instancia (distintos y distintivos de cada individuo de la clase) y las relaciones (de las que son origen).

Por el proceso seguido anteriormente, en este paso se incluyen ya de una manera u otra prácticamente la

mayoría de los elementos de un tesoro; es una etapa de resumen y de clasificación de la información obtenida del tesoro en etapas anteriores.

### Tarea 5. Definición detallada de las relaciones binarias ad hoc

A partir del diccionario de conceptos creado en la tarea anterior, especificaremos el nombre de la relación, el concepto de origen (denominado también fuente), la cardinalidad (el número mínimo y máximo de instancias del concepto destino con las que se puede corresponder, mediante una misma relación, una instancia del concepto origen), el concepto de destino, las propiedades matemáticas que se cumplen -reflexiva, simétrica, transitiva, etc.; véase su definición, por ejemplo, en: **Caballero Roldán**, et al., (2007)- y la relación inversa a que da lugar.

La cardinalidad de una relación muy difícilmente podrá ser extraída de un tesoro, por lo que habrá que recurrir a las fuentes de elaboración del mismo para determinarla. Las propiedades matemáticas se pueden deducir de la semántica de la relación asociativa del tesoro de la que proviene la relación ad hoc de la ontología.

### Tarea 6. Definición detallada de los atributos de instancia

Asimismo, a partir del diccionario de conceptos, incluiremos los atributos de la instancia que consideremos relevantes en el dominio, identificando para cada uno de ellos su nombre, clase a la que se asocia, el tipo de valores que adopta (cadena de caracteres, número entero, número real, booleano, fecha, etc.), las unidades

### Ejemplo 1

	Dioptría
TR	Miopía

### Ejemplo 2

	Presbicia
NA	Útese en casos de anomalías en la visión próxima padecidas por personas mayores de 40 años de edad

Del ejemplo 1 podemos deducir que la "Dioptría" es la "Unidad de medida" de la "Miopía"; del ejemplo 2 se infiere: el nombre del atributo ("Edad"), el tipo de valor del atributo (entero), así como que "Años" es la unidad en la que se mide una "Edad". En este caso, existe una restricción sobre el rango de valores viene dada por la semántica propia del concepto "Edad".

de medida, la precisión que se observa con cada variación de valor y el rango de valores (en el caso de números), el valor por defecto, las cardinalidades máxima y mínima (se admite la posibilidad de que haya atributos multivaluados) y otros datos que consideremos necesarios (como por ejemplo, qué otros atributos y constantes se usan al inferir el valor de este atributo).

En el caso de partir de un tesoro, las unidades de medida suelen venir reflejadas como descriptores y a la vez como términos relacionados dentro del tesoro, o a veces aparecen dentro de una nota aclaratoria, asociadas a un valor del atributo que califican y que dará a su vez el tipo de valores del mismo.

### Tarea 7. Definición detallada de los atributos de clase

Especificaremos su nombre, el concepto en el que el atributo está definido, el tipo de valor que posee, su valor (pues caracteriza a la clase, lo que lo distingue de los atributos de instancia), las unidades de medida, la precisión, el rango de valores, su cardinalidad y los atributos de instancia cuyos valores se pueden inferir con los valores de este atributo de clase. Si se parte de un tesoro para construir la ontología, las consideraciones son análogas al caso anterior.

#### Ejemplo

Miopía magna

NA Úsese en el caso de aquellas miopías con una graduación mayor de 6,00 dioptrías.

Del ejemplo podemos deducir el nombre del atributo ("Graduación"), el tipo de valor del atributo (real) así como que "Dioptría" es la "Unidad" con la que se mide una "Graduación" y que existe una restricción sobre el rango de valores (> 6,00) para dicho atributo.

### Tarea 8. Definición detallada de las constantes

Gracias al glosario desarrollado en la tarea 1 y al diccionario de conceptos del paso 4 podremos listar las constantes que necesitamos en nuestra ontología. Para cada una de ellas es importante mencionar su nombre, el tipo de valor, su valor, las unidades de medida y los atributos que se pueden inferir mediante su uso. Difícilmente en un tesoro se ven definiciones de constantes, por lo que para la realización de esta tarea habrá que recurrir a las fuentes de la elaboración del mismo.

### Tarea 9. Modelización de los axiomas

Un axioma se puede definir como una "afirmación siempre cierta en el ámbito de la ontología que se está

formalizando". Es necesario, pues, la identificación y descripción de los axiomas del dominio que se está conceptualizando, indicando para cada uno de ellos su nombre, descripción y la expresión lógica que lo representa, así como los conceptos, atributos, relaciones ad hoc, y variables utilizadas en dicha expresión. En el caso de contar con un tesoro como punto de partida, los axiomas vendrán definidos a partir de las notas de aplicación.

#### Ejemplo

Miopía magna

NA Úsese en el caso de aquellas miopías con una graduación mayor de 6,00 dioptrías.

De la nota de aplicación anterior podemos extraer también el siguiente axioma en lógica de primer orden, según el escritor de axiomas de WebODE:

Forall (?X, ?Y)

([Miopía Magna] (?X) and [Graduación] (?X, ?Y) -> [Mayor] (?Y, 6.00))

en el que estarían implicados el concepto "Miopía magna", el atributo "Graduación" y las variables ?X e ?Y.

### Tarea 10. Definición de las reglas

Son un tipo especial de axioma por el que calculamos, por ejemplo, el valor de un atributo en función de otros ya especificados o de valores de constantes. Han de formalizarse especificando el nombre, su descripción en lenguaje natural y la expresión lógica que representa la regla, así como los conceptos, atributos, relaciones ad hoc y variables usados.

En el caso de contar con un tesoro, también las reglas vendrán definidas a partir de las notas de aplicación. La diferencia entre implementar una nota aclaratoria como una regla o como un axioma es contextual: si, por ejemplo, un atributo recibe un valor determinado según la misma, se implementará mediante una regla; en caso contrario, a través de un axioma.

### Tarea 11. Definición de las instancias

La última tarea pasa por indicar las características de las instancias incluidas en el diccionario de conceptos creado en la tarea 4. Indicaremos, para cada una de ellas, su nombre, el del concepto al que pertenece, los atributos de instancia que posee y el valor de dichos atributos. Todo este conocimiento, dado que las instancias no figuran en forma alguna en un tesoro, deberá obtenerse de las fuentes de partida para la construcción del tesoro.

### Una arquitectura prototipo para la transformación de tesauros en ontologías

Existen varios tipos de ontologías, las cuales pueden clasificarse en atención a múltiples criterios, (Gómez Pérez; Fernández López; Corcho, 2004). De todos ellos, nos fijamos ahora en aquél que clasifica una ontología en función del objeto conceptualizado. De acuerdo con esta clasificación, las clases más sobresalientes son:

- De dominio: son reutilizables dentro de una determinada área del saber, y aportan el vocabulario para referirse a los conceptos y relaciones propios de dicho dominio.
- Genéricas o comunes: representan conocimiento de sentido común, reutilizable en varios dominios.
- Ontologías para la representación del conocimiento: situadas en un nivel superior de abstracción, encontramos las que recogen las primitivas que se deben utilizar para representar conocimientos (de acuerdo con un determinado paradigma de representación del conocimiento) y las de nivel superior (*top-level ontologies*), que describen conceptos muy generales, a los cuales deberían referenciar como términos más genéricos los conceptos del resto de ontologías existentes.

En la figura 3 se muestra la representación gráfica, tomada de *WebODE*, de una ontología de peso para la representación del conocimiento asociada a tesauros facetados, realizada siguiendo *Methontology*.

<http://webode.dia.fi.upm.es>

En la tabla 2 se muestra el diccionario de conceptos de la ontología de la figura anterior, instanciada para el dominio de la óptica (tabla extraída de *WebODE*). Estas instancias serán las que se referencien desde la ontología de dominio (de peso) construida para el campo de las anomalías oculares siguiendo asimismo *Methontology*, cuya representación gráfica (también extraída de *WebODE*) se muestra en la figura 4.

La división de la formalización del dominio en estas dos ontologías de peso no es casual; en primer lugar, es modular y extensible; además, proporciona un marco (mediante la ontología para la representación del conocimiento) altamente útil y reutilizable: tan sólo hacen falta detallar en dicha ontología las instancias propias del dominio concreto. De esta forma, además, se obtienen los términos clave de dicho dominio, que servirían como conjunto de arranque (*bootstrapping*) para realizar el aprendizaje automático de las ontologías de dominio propias de cada campo identificado en las distintas facetas instanciadas.

### Conclusiones

Se ha presentado una aproximación práctica a la construcción de ontologías a partir de tesauros inspirada en la metodología *Methontology*, alternativa a la publicación de tesauros en la web semántica mediante *SKOS-Core*. La utilización conjunta de una metodología con una herramienta de desarrollo de ontologías proporciona el código de la ontología en un lenguaje estándar del *W3C*, lo que facilita en gran medida esta

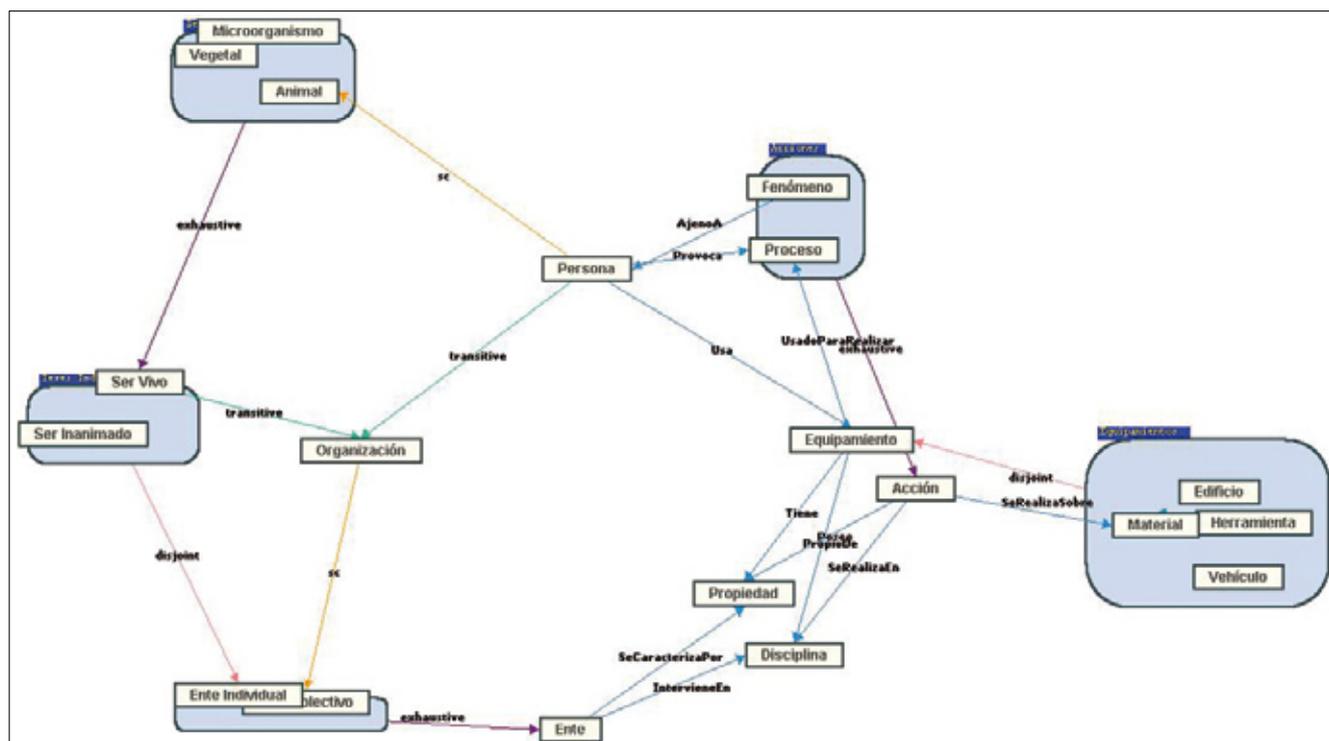


Figura 3. Representación gráfica de una ontología que conceptualiza tesauros facetados

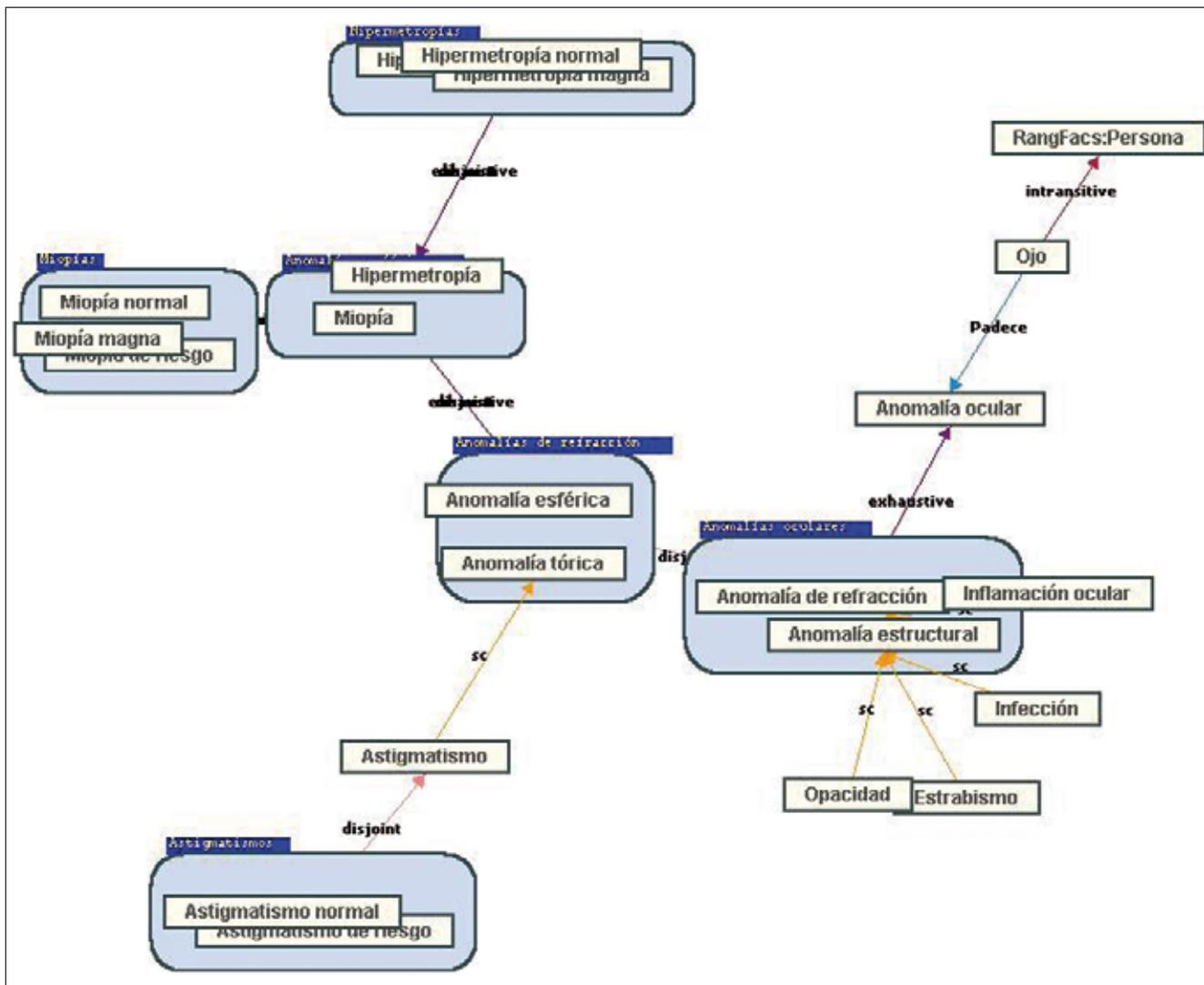


Figura 4. Representación gráfica de una ontología de dominio para anomalías oculares, derivada de un tesoro

tarea, además de sistematizarla y robustecerla. La formalización del conocimiento como una ontología y no como un tesoro aporta los beneficios adicionales mencionados al final del estudio comparativo entre ambas herramientas.

Por otro lado, y como se puede observar en el apartado anterior, se pueden articular varias ontologías, como si de un puzzle se tratara, para obtener una arquitectura ontológica que representa globalmente un dominio, lo cual es impensable en el caso de tesauros, ya que operan en un campo muy restringido del conocimiento. Éste es un aspecto en el que se espera que la investigación sea muy fructífera en el futuro.

El inconveniente que presenta nuestra propuesta para un documentalista es que deberá conocer y manejar una serie de conceptos nuevos propios de la lógica (lógica de predicados, descriptiva, etc.), la matemática discreta (propiedades de relaciones: reflexiva, simétrica, transitiva, etc.), la ingeniería del conocimiento (técnicas de educción y formalismos de representación del conocimiento, etc.), aparte de aquellos propios de su

profesión, como los relacionados con las bases de datos (formatos de representación de la información: números reales, enteros, booleanos, etc.).

**“Se pueden articular varias ontologías, como si de un puzzle se tratara, para obtener una arquitectura de ontologías que representan globalmente un dominio”**

En cualquier caso, el principal problema (sobre todo en la web semántica) es la adopción de ontologías concretas por parte de un grupo tan heterogéneo de usuarios como es internet para un dominio determinado. Las ontologías necesitan ser aceptadas masivamente por toda la red para conseguir sus objetivos: hacer explícitos conocimientos y servir de base para la inferencia sobre los mismos (en caso contrario, harían falta mecanismos de traducción entre ontologías similares a los existentes para la traducción entre distintas len-

Concept name	Synonyms	Acronyms	Instances	Class attributes	Instance attributes	Relations
Acción	-	-	-	Natural	-	SeRealizaEn SeRealizaSobre Posee
Animal	-	-	-	Microscópico Racional SeTraslada	-	-
Disciplina	-	-	Optometría	-	-	-
Edificio	-	-	Óptica	-	-	-
Ente	-	-	-	Individual	-	Intervienen SeCaracterizaPor
Ente Colectivo	-	-	-	Individual	-	-
Ente Individual	-	-	-	Animado Individual	-	-
Equipamiento	-	-	-	-	-	PropioDe UsadoParaRealizar Tiene
Fenómeno	-	-	Infección Inflamación ocular Opacidad del ojo Refracción de la luz	Natural	-	AjenoA
Herramienta	-	-	Queratómetro Frontofocómetro Gafa Autorrefractómetro Destornillador	-	-	OperaSobre
Material	-	-	Ojo Lente	-	-	-
Microorganismo	-	-	-	Microscópico SeTraslada	-	-
Organización	-	-	Colegio Nacional de Ópticos Optometristas	-	-	-
Persona	Ser Humano	-	Optometrista Óptico Paciente	Racional	-	Usa Provoca
Proceso	-	-	Compra de herramientas Compra de materiales Montaje de gafa Graduación	Natural	-	-
Propiedad	-	-	Apellidos Nombre Dirección Magnitud de refracción Signo de refracción	-	-	-
Ser Inanimado	-	-	-	Animado	-	-
Ser Vivo	-	-	-	Animado Microscópico SeTraslada	-	-
Vegetal	Planta	-	-	Microscópico SeTraslada	-	-
Vehículo	-	-	-	-	-	-

Tabla 2: Diccionario de conceptos de la ontología de facetas de Ranganathan

guas); pero es difícil coordinar y organizar la web para alcanzar esta aceptación masiva, pues no hay forma de imponer el uso de una ontología concreta a semejante volumen de usuarios implicados.

## Bibliografía

- Aguado, G.; Álvarez-de-Mon I.; Gómez-Pérez, A.; Pareja-Lora, A.; Plaza-Arteche, R.** *A semantic web page linguistic annotation model. Semantic web meets language resources. Technical report WS-02-16.* American Association for Artificial Intelligence. Menlo Park, California, EUA: AAAI Press, 2002.
- Aguado-de-Cea, G.; Álvarez-de-Mon, I.; Pareja-Lora, A.; Plaza-Arteche, R.** "OntoTag: a semantic web page linguistic annotation model". En: *Proceedings of the ECAI 2002 workshop on semantic authoring, annotation and knowledge markup*, 2002.
- Aguado-de-Cea, G.; Álvarez-de-Mon, I.; Pareja-Lora, A.; Plaza-Arteche, R.** "Rdf(S)/xml linguistic annotation of semantic web pages". En: *Proceedings of the Coling 2002, 2<sup>nd</sup> workshop on nlp and xml (Nlpxml-2002)*, 2002.
- Aitchison, J.; Gilchrist, A.; Bawden, D.** *Thesaurus construction and use: a practical manual*. 3ª ed. Londres: Aslib, 1997.
- Benjamins, V. R.; Fensel, D.; Decker, S.; Gómez-Pérez, A.** "(KA)<sup>2</sup>: building ontologies for the internet: a mid term report". En: *International journal of human computer studies (Ijhc)*, 1999, n. 51, pp. 687-712.
- Benjamins, R.; Contreras, J.; Martín, F.; Navarrete, B.; Aguado-de-Cea, G.; Álvarez-de-Mon, I.; Pareja-Lora, A.; Plaza-Arteche, R.** *State of the art on annotation tools and services*. Esperanto Services (IST-2001-34373), deliverable D3.1. 2003. <http://www.esperanto.net>
- Berners-Lee, T.; Fischetti, M.** *Weaving the web: the original design and ultimate destiny of the world wide web by its inventor*. San Francisco: Harper, 1999.
- Borst, W. N.** *Construction of engineering ontologies for knowledge sharing and reuse. PhD Thesis*. University of Twente, The Netherlands, Enschede, 1997.
- Caballero-Roldán, R.; Hortalá-González, T.; Martí-Oliet, N.; Nieva-Soto, S.; Pareja-Lora, A.; Rodríguez-Artalejo, M.** *Matemática discreta para informáticos: ejercicios resueltos*. Madrid: Pearson Educación SA, 2007.
- Dean, M.; Schreiber, G.** *OWL web ontology language reference. W3C Recommendation*. Edición 2004. Consultado en: 28-09-07. <http://www.w3.org/TR/owl-ref/>
- Ding, Y.; Foo, S.** "Ontology research and development. Part 1—a review of ontology generation". En: *Journal of information science*, 2002, v. 28, n. 2, pp. 123-136.
- Ding, Y.; Foo, S.** "Ontology research and development. Part 2—a review of ontology mapping and evolving". En: *Journal of information science*, 2002, v. 28, n. 5, pp. 375-388.
- Ferreira, D.** *TemaTres: software libre para gestión de tesauros*. Consultado en: 12-09-07. <http://www.r020.com.ar/tematres/>
- García-Jiménez, A.** "Instrumentos de representación del conocimiento: tesauros versus ontologías". En: *Anales de documentación*, 2004, n. 7, pp. 79-95.
- Gil-Urdiciain, B.** "La 'Colon Classification'". En: *Revista general de información y documentación*, 1994, v. 4, n. 1, pp. 111-121.
- Gómez-Pérez, A.; Corcho, O.** "Ontology specification languages for the semantic web". En: *IEEE intelligent systems*, 2002, enero, v. 17, n. 1, pp. 54-60.
- Gómez-Pérez, A.; Fernández-López, M.; Corcho, O.** *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*. Londres; Nueva York: Springer-Verlag, 2003.
- Gruber, R.** "A translation approach to portable ontology specification". En: *Knowledge acquisition*, 1993, n. 5, pp. 199-220.
- Lassila, O.; McGuinness, D.** *The role of frame-based representation on the semantic web. KSL tech report number KSL-01-02*, enero de 2001. <http://www.ksl.stanford.edu/people/dlm/etai/lassila-mcguinness-fbr-sw.html>
- Luke S.; Hefflin J.** *SHOE 1.01. Proposed Specification. SHOE Project*. Consultado en: 14-09-07. <http://www.cs.umd.edu/projects/plus/SHOE/spec.html>
- Motta, Enrico; Buckingham-Shum, Simon; Domingue, John.** "Case studies in ontology-driven document enrichment". En: *Proceedings of the 12<sup>th</sup> Banff knowledge acquisition workshop*, Alberta, 1999.
- OntoMat*. Consultado en: 12-09-07. <http://annotation.semanticweb.org/ontomat/index.html>
- OntoWeb*. Consultado en: 12-09-07. [http://ontoweb.aifb.uni-karlsruhe.de/About/Deliverables/D13\\_v1-0.zip](http://ontoweb.aifb.uni-karlsruhe.de/About/Deliverables/D13_v1-0.zip)
- Peis, E.; Herrera-Viedma, E.; Hassan, Y.; Herrera, J. C.** "Análisis de la web semántica: estado actual y requisitos futuros". En: *El profesional de la información*, 2003, septiembre-octubre, v. 12, n. 15.
- Pérez-Agüera, J. R.** "L'automatització de tesauros i la seva utilització en el web semàntic". En: *BiD, textos universitaris de biblioteconomia i documentació*, 2004, diciembre, n. 13. [http://www2.uib.es/bid/consulta\\_articulos.php?fichero=13perez1.htm](http://www2.uib.es/bid/consulta_articulos.php?fichero=13perez1.htm)
- Qin, J.; Paling, S.** "Converting a controlled vocabulary into an ontology: the case of GEM". En: *Information research*, 2001, v. 6, n. 2. <http://informationr.net/ir/6-2/paper94.html>
- Ranganathan, S. R.** *Prolegomena to library classification*, 3ª ed. New Delhi: Asia Publishing House, 1967.
- Staab, S.; Angele, J.; Decker, S.; Erdmann, M.; Hotho, A.; Mädche, A.; Schnurr, H. P.; Studer, R.** "Semantic community web portals". En: *Proceedings of the 9<sup>th</sup> international world wide web conference (www9)*, 2000, v. 33, n. 1-6 (número especial), pp. 473-491.
- Studer, R.; Benjamins, R.; Fensel, D.** "Knowledge engineering: principles and methods". En: *Data and knowledge engineering (DKE)*, 1998, v. 25, n. 1-2, pp. 161-197.
- WebODE*. Consultado en: 01-07-07. <http://webode.dia.fi.upm.es>
- Wilcock, G.; Buitelaar, P.; Pareja-Lora, A.; Bryant, B.; Lin, J.; Ide, N.** "The roles of natural language and xml in the semantic web". En: **Huang, C. R.; Lenders, W.** (eds.). *Computational linguistics and beyond: perspectives at the beginning of the 21<sup>st</sup> century*. Frontiers in linguistics series, Academia Sinica, Taiwan, 2004.

**Alberto García-Torres, Tribunal Supremo, Plaza Villa de París, s/n, 28071 Madrid.**  
[torres.al@gmail.com](mailto:torres.al@gmail.com)

**Antonio Pareja-Lora, Dpto. Sistemas Informáticos y Computación, Universidad Complutense de Madrid, C/ Profesor García Santesmases, s/n, 28040 Madrid.**  
[apareja@sip.ucm.es](mailto:apareja@sip.ucm.es)

**Daniel Pradana-López, Óptica Gasset, C/ José Ortega y Gasset 96, 28006 Madrid.**  
[opticagasset@yahoo.es](mailto:opticagasset@yahoo.es)

# Evolución de la estructura científica española: ISI Web of Science 1990-2005

Por Benjamín Vargas-Quesada, Félix de Moya-Anegón, Zaida Chinchilla-Rodríguez, Elena Corera-Álvarez y Vicente Guerrero-Bote (Grupo ScImago)



**Benjamín Vargas-Quesada**, profesor colaborador del área de comunicación y documentación de la Universidad de Granada e investigador del Grupo SCImago.



**Félix de Moya-Anegón**, catedrático del área de comunicación y documentación de la Universidad de Granada y responsable del grupo SCImago.



**Zaida Chinchilla-Rodríguez**, doctora en documentación e información científica. Investigadora del Grupo SCImago.

**Resumen:** Este trabajo muestra un estudio longitudinal de la evolución del dominio científico español ISI en tres intervalos temporales: 1990-1995, 1996-2000 y 2001-2005. El objetivo final es mostrar la estructura básica y columna vertebral de la investigación española, así como estudiar su evolución. Para ello se han utilizado las categorías ISI como unidades de representación, la cocitación como unidad de medida y algoritmos de poda para simplificar las relaciones entre categorías en las visualizaciones. Los resultados muestran que desde un punto de vista macro-estructural, biomedicina y ciencias de los materiales son el núcleo

básico de investigación. Con el tiempo, este núcleo se consolida y enriquece con el área de agricultura y ciencias del suelo. Desde la perspectiva micro-estructural, se observa un incremento de las ciencias interdisciplinarias, que ocupan posiciones centrales, indicando un aumento en la interacción entre las ciencias tradicionalmente relevantes o visibles y las más periféricas u ocultas.

**Palabras clave:** Visualización de información, Visualización de redes, Producción científica, Cocitación, España, Análisis de dominios, Evolución de redes, Pathfinder networks, Pfnnet.

**Title: ISI Web of Science 1990-2005: evolution of the science web structure in Spain**

**Abstract:** This longitudinal study of the development of Spain's ISI scientific domain examines three key time periods: 1990-1995, 1996-2000 and 2001-2005, with the goal of describing the basic structure and backbone of Spanish research, as well as studying its development over time. To this end, ISI categories are used as units of analysis, co-citation as units of measure, and pruning algorithms to simplify the relationship between categories in the visualizations. The results show that from the macrostructure point of view, biomedicine and materials science research provide the basic nucleus at first. Over the years this nucleus is consolidated and expanded with agricultural and soil sciences research. From the microstructure perspective, an increase is observed in the number of interdisciplinary categories, which occupy central positions, indicating increased interaction between disciplines that have traditionally been more important or visible and those on the periphery, or less well known.

**Keywords:** Information visualization, Network visualization, Scientific output, Co-citation, Spain, Domain analysis, Network evolution, Pathfinder networks, Pfnnet.

**Vargas-Quesada, Benjamín; Moya-Anegón, Félix de; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; Guerrero-Bote, Vicente.** "Evolución de la estructura científica española: ISI Web of Science 1990-2005". En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, pp. 22-37.

DOI: 10.3145/epi.2008.ene.03



**Elena Corera-Álvarez**, doctora en documentación e información científica e investigadora del Grupo SCImago.



**Vicente Guerrero-Bote**, profesor titular de biblioteconomía y documentación de la Universidad de Extremadura e investigador del Grupo SCImago.

Artículo recibido el 23-07-07

Aceptación definitiva: 26-09-07

## 1. Introducción

En **Moya-Anegón**, et al. (2006) propusimos una metodología para la visualización y análisis del dominio científico español. Basándonos en la cocitación de categorías *ISI*, en la representación de su estructura mediante redes sociales y en su simplificación a través de algoritmos de poda, mostramos el modelo de articulación de la investigación científica española en el período comprendido entre 1990 y 2005. En ese mismo trabajo, además sugerimos que esas mismas visualizaciones construidas con espacios temporales más cortos, podían ser perfectamente utilizadas para el estudio de la evolución de un determinado dominio científico.

El consenso general tanto en el campo de la visualización de la información como en el de la cartografía geográfica, es que varios mapas de un mismo objeto de estudio son mejores que uno solo. Esto es así porque cada mapa puede proporcionar una perspectiva diferente. El precursor del estudio de la ciencia por medio de sucesiones de mapas fue **Garfield** (1994), introduciendo el concepto de *historiogramas*, que consiste en una serie de representaciones cronológicas secuenciales, a partir de las cuales es posible estudiar la evolución del conocimiento científico.

En este trabajo mostramos y analizamos la evolución de la estructura del dominio científico español por medio de la visualización de tres cienciogramas. Comenzaremos haciendo una breve revisión de los trabajos relacionados con éste, seguiremos con el material y metodología utilizados para la construcción de dichos cienciogramas, continuaremos mostrando los resultados alcanzados junto con su discusión oportuna en cada caso, y finalizaremos con las conclusiones obtenidas en base al análisis de su totalidad.

## 2. Antecedentes

Ya desde principios de los años 60, **Garfield** demuestra que es posible construir mapas históricos basados en la citación de los documentos científicos (**Garfield**, 1963), (**Garfield; Sher; Torpie**, 1964). Poco más tarde **Price** (1965) indica que los patrones de citación de los autores de artículos científicos definen los frentes de investigación y que además pueden ser utilizados para reflejar la estructura científica de un dominio. En la década de los 70, **Small** y **Griffith** (**Small; Griffith**, 1974), (**Griffith**, 1974) representan las especialidades de las ciencias naturales, demostrando que la ciencia es una red de especialidades interconectadas entre sí. Un año más tarde **Aaronson** (1975) radiografía la biomedicina de un período de dos años y muestra su evolución. En los años 80, la generación de visualizaciones de mapas de dominios científicos se hace más prolífica. Aparecen los mapas de bioquímica y biología molecular (**Garfield**, 1981), biotecnología y genética

molecular (**Garfield**, 1984), bioquímica, inmunología y biología animal y vegetal (**Garfield**, 1988) y por último, farmacología (**Seiden; Swanson**, 1989).

Con la llegada de los 90 surgen nuevos métodos y técnicas para la recopilación de la información y el análisis, la visualización y el posicionamiento espacial de la información. Así por ejemplo, desde el *Centre for Science and Technology Studies (CWTS)* se trabaja en la generación de mapas de la ciencia, haciendo hincapié en su estructura y aspectos dinámicos (**Braam; Moed; van Raan**, 1991; **van Raan; Tijssen**, 1993; **Braam; Moed; van Raan**, 1991b; **Brandenburg; Himsolt; Rohrer**, 1995). **White** y **McCain** en 1997 y 1998 proponen la visualización como modelo para el estudio y análisis de una disciplina mediante mapas temporales. **Garfield**, en 1998 aboga por el uso de las nuevas técnicas de visualización para la generación de mapas globales-secuenciales de la ciencia. **Chen** (1998a; 1998b) incorpora *Pathfinder networks (Pfnets)* al campo de la documentación para la poda de enlaces en la visualización de redes sociales, el propósito es obtener estructuras simples al mismo tiempo que informativas.

El nuevo milenio incorpora como reto la generación dinámica de los mapas y su uso como interfaces para el acceso y la recuperación de información. **White** (2000) propone el uso de las redes centradas en un sujeto como interfaces de acceso y recuperación de información bibliográfica. **Noyons, Moed** y **Luwel** (1999), **Buter** y **Noyons** (2001) y luego **Noyons** (2001) en solitario, analizan el uso de los mapas como metáforas de una disciplina científica, su uso como interfaces y sus limitaciones. **Chen** y **Paul** (2001), **Chen, Paul** y **O'keefe** (2001) representan los patrones estructurales de la literatura científica en mapas 3D. **Ding, Chowdhury** y **Foo** (2001) ponen de manifiesto los modelos, patrones y tendencias del campo de la recuperación a partir de su estructura intelectual. **Guerrero-Bote, Moya-Anegón** y **Herrero-Solana**, (2002), utilizan un *self-organizative map (SOM)* para la clasificación, browsing y recuperación de información. **White, Buzydlowski** y **Lin** (2000), presentan *Authorlink*: un sistema dinámico de visualización de la información basado en la cocitación de autores, que permite el browsing y la recuperación de información en tiempo real (**Buzydlowski; White; Lin**, 2002; **Lin; White; Buzydlowski**, 2003). **Small** (2003), teoriza sobre el diseño de una herramienta web capaz de detectar y monitorizar cambios en los frentes de investigación de un área en tiempo real. **Chen** y **Kuljis** (2003) estudian la aparición y evolución de frentes de investigación en el campo de la física. De nuevo **Chen**, pero ahora con **Morris** (**Chen; Morris**, 2003), enfrentan *Minimum Spanning Tree (MST)* con *Pfnets* para la detección e identificación de frentes de investigación, llegando a la conclusión de que el segundo ofrece resultados más estables. **Morris**,

et al. (2003) trabajan en la visualización, detección e identificación de los cambios temporales en los frentes de investigación. El *Grupo SCImago (Moya-Anegón, et al., 2004)*, (*Moya-Anegón, et al., 2005*) propone la cocitación de categorías *ISI* como unidades de análisis y representación para la generación de mapas de grandes dominios científicos y compara tres de ellos. Ese mismo año, **Boyack, Klavans y Börner** (2005) desarrollan una metodología que permite la visualización a “vista de pájaro” del mapa de la ciencia mundial para el año 2000. Un año más tarde, **Samoylenko, et al.** (2006) propone un método para la visualización de la estructura científica mundial y su evolución por medio de *MSTs*. Finalmente **Leydesdorff** (2007) combina el *Journal Citation Report* del *SCI* y del *SSCI*, ambos de 2004, y a partir de los mapas de revistas que obtiene, muestra sus especialidades.

### 3. Material y métodos

#### 3.1. Recopilación de información

Con fines estrictamente investigadores, hemos descargado del *ISI Web of Science (The Thomson Corporation, 2007)*, en concreto de las bases de datos *SCI-Expanded*, *SSCI* y *A&HCI*, todos los registros de la producción científica española publicados entre los años 1990 y 2005. Para ello se han recuperado todos los documentos que en ese período de tiempo contenían la palabra *spain* en el campo “Address”. También hemos descargado los *Journal Citation Reports (JCRs)*, tanto en su edición *Science* como *Social Sciences*, para los años que se pretende cubrir. Los registros se han exportado a una base de datos relacional que contiene un total de 370.047 documentos fuente (*articles, biographical items, book reviews, corrections, editorial materials, letters, meeting abstracts, news items y reviews*) que se han publicado en 7.585 revistas *ISI* clasificadas en 241 categorías y que citan a 9.550.959 documentos *ISI* anteriormente publicados.

Algunos autores (entre los que se encuentra un evaluador de este trabajo) consideran que es posible que exista un cierto sesgo metodológico al introducir un mecanismo de evaluación (las bases de datos *ISI*) manifiestamente centrado en la biomedicina. En el *Grupo SCImago (SCImago Group, 2007)* no estamos de acuerdo con esta opinión. La historia de las bases de datos *ISI* ha estado plagada de críticas relacionadas con el sesgo en la cobertura de sus revistas en términos de disciplinariedad y nacionalidad. No obstante, estudios que comparan la cobertura del *SCI* con la del *Ulrich's International Periodicals Directory (Braun; Glänzel; Schubert, 2000)*, demuestran que esto no es así. El conjunto de revistas *SCI* presenta un balance equilibrado con respecto al *Ulrich's* a nivel macro, por lo que afecta al menos a países y disciplinas. En contra de la creencia general, este estudio demuestra que no existe un sesgo en estas bases de datos a favor de EUA o de la biomedicina. En algunos casos,

incluso existe una infra-representación. Las excepciones en cuanto a cobertura por disciplinas se centran en Alemania y en concreto en la agricultura, y en lo referente a editores, destaca *Elsevier*. En general, existe una sobre-representación de los principales editores en el *SCI* y, en cualquier caso, este fenómeno no afecta a los objetivos de este trabajo.

#### 3.2. Unidades de análisis

Como hemos indicado anteriormente, recurrimos a las categorías *ISI* como unidades de análisis y representación. Cada categoría aglutina las revistas que han sido categorizadas bajo ese nombre y a su vez, a los documentos que se han publicado en esas publicaciones. No utilizamos la categoría *Multidisciplinary sciences*, por lo que todos los documentos adscritos ella se han recategorizado. Para ello, hemos seguido el método propuesto por **Glänzel** y sus colegas (**Glänzel; Schubert; Czerwon, 1999; Glänzel, et al., 1999**), que consiste en volver a asignar a cada uno de esos documentos la categoría desde la que son más referenciados.

---

**“Tomando como patrón el directorio de revistas Ulrich, Braun, Glänzel y Schubert demostraron en 2000 que no existe un sesgo en las bases de datos ISI a favor de EUA o de la biomedicina”**

---

#### 3.3. Unidades de medida

De acuerdo con nuestra experiencia (**Vargas-Quesada; Moya-Anegón, 2007**), utilizaremos la cocitación como unidad de medida para cuantificar el grado de relación existente entre cada una de las categorías del *JCR*. No aplicaremos ninguna medida de similaridad entre categorías. Al igual que **Leydesdorff y Vaughan** (2006) y **Ortega y Aguillo** (2006), hemos comprobado que la normalización de los valores de cocitación en matrices simétricas, provoca distorsiones. En nuestro caso hace que las relaciones entre categorías no sean tan coherentes como cuando se utiliza *raw data*.

#### 3.4. Reducción de la dimensión

La visualización de un dominio científico implica la representación de un número elevado de unidades de análisis que muestran sus interacciones mediante enlaces o conexiones. Entre las distintas técnicas que existen para la creación de visualizaciones bidimensionales, *Pfnets* basado en valores de cocitación sin normalizar y sustentado en algoritmos de distribución espacial del tipo *spring embedders*, parece ofrecer numerosas ventajas desde un punto de vista estético y de análisis (**White; Buzdowski; Lin, 2000**).

Aunque *Pfnets* se ha utilizado en los campos de la bibliometría, informetría y ciencimetría desde 1990 (Fowler; Dearhold, 1990), su introducción en el campo de la citación vino de la mano de Chen (1998a; 1999). Se basa en dos elementos: en la *distancia de Minkowski* ( $r$ ), y en una extensión del principio de desigual del triángulo aplicada a la longitud de los *paths* o caminos que conectan dos nodos ( $q$ ). De esta forma, todos los caminos que violen la desigualdad triangular y que tengan una distancia menor que otro camino entre los mismos puntos compuesto de hasta  $q$  enlaces, calculando las distancias mediante la ecuación paramétrica de *Minkowski* con el parámetro  $r$ , serán eliminados. Desde nuestro punto de vista, *Pfnets* con valores de  $r = \infty$ , y  $q = n - 1$  es la mejor opción para capturar la estructura intelectual de un dominio, por medio de los enlaces entre categorías más prominentes.

### 3.5. Distribución espacial de la información

Existen muchos métodos para la generación automática de grafos. De todos ellos, los insertadores de muelles (*spring embedders*, son los más utilizados en el área de la documentación. De este tipo de algoritmos, el más utilizado en documentación es el de Kamada y Kawai (1989). Su comportamiento ante los mínimos locales, el intento de minimizar las diferencias con las distancias teóricas en todo el gráfico, sus buenos tiempos de computación, etc., hacen que sea uno de los más utilizados para la visualización de la información.

La combinación de valores de cocitación sin normalizar, *Pfnets*, y **Kamada-Kawai**, provoca que las categorías más interdisciplinarias de un dominio científico tiendan a situarse en el centro, como consecuencia de su mayor grado o número de enlaces (Chen, 1998a). Esto les da un valor añadido a las visualizaciones obtenidas, pues además de mostrar la estructura científica de un dominio, informa de una manera intuitiva sobre cuáles son las categorías o áreas temáticas que más participan en su desarrollo.

## 4. Resultados y discusión

El siguiente análisis se centra en la evolución temporal del dominio geográfico español a través de tres cienciogramas comprendidos entre el período 1990 y 2005. Estos cienciogramas tienen una peculiaridad que no aparecía en los mostrados en Moya-Anegón, et al. (2006): presentan algunas conexiones o enlaces muy débiles entre categorías, provocando lo que podríamos denominar como conexiones extrañas o enlaces “dudosos” y que para Granoveter (1973), son los responsables de la interconexión de nodos de diferentes clusters. Para Small (1999), identifican conexiones inestables que se producen entre agrupaciones de documentos que representan el punto interdisciplinar entre dos áreas de conocimiento. Para el caso que nos ocupa, identifican

conexiones tangenciales entre categorías, y son consecuencia de la falta de información referencial que las bases de datos *ISI* tienen en algunos dominios científicos, como por ejemplo España, en las áreas de ciencias sociales y humanidades. Sirva como ejemplo de estos enlaces débiles o dudosos la conexión que se produce entre la categoría *Literature, African, Australian, Canadian y Ecology* (zona central-derecha) en la ilustración 1.

En este artículo y para el período de tiempo que aquí se estudia, consideraremos que los enlaces con un valor de cocitación inferior a cinco, pueden ser responsables de conexiones dudosas entre categorías. Para hacerlo saber a todo aquel que observe estos cienciogramas, los colorearemos de rojo.

### 4.1. Cienciogramas

Las ilustraciones 1, 2 y 3, muestran los cienciogramas de la producción científica española durante los períodos: 1990-1995, 1996-2000 y 2001-2005, que suponen la esquematización gráfica de 90.223, 121.996 y 157.828 documentos, agrupados en sus respectivas categorías *JCR*.

El cienciograma del primer período se compone de 219 categorías, dos de las cuales: *Poetry*, y *Literature German, Netherlandic or Scandinavian* se muestran aisladas, es decir, cuentan con artículos recogidos por las bases de datos del *ISI*, pero no muestran relación con ninguna otra categoría, salvo con ellas mismas. El segundo período está constituido por 223 categorías, no mostrándose ninguna de ellas como aislada. El último intervalo consta de 240, no apareciendo tampoco ninguna de ellas aislada.

Los tres cienciogramas presentan la ya típica estructura vista en los trabajos en los que aplicamos esta metodología (Moya-Anegón, et al., 2006; Vargas-Quesada; Moya-Anegón, 2007): una neurona con un gran núcleo central del que surgen diversas prolongaciones o dendritas. A simple vista, se puede observar a lo largo de los tres cienciogramas, que cuanto más actual es el período de estudio, mayor es el núcleo central y más uniforme y reducido es el número de dendritas que parten de él. Es decir, parece verse una evolución consistente en el aumento de las categorías que constituyen el centro de la investigación del dominio español, al tiempo que se aprecia una mayor definición y unificación de las principales líneas de investigación de dicho dominio.

Para un estudio más detallado de la evolución del dominio, realizamos análisis factorial (AF) sobre la matriz de cocitación original de cada período. Posteriormente trasladamos los resultados a cada uno de sus correspondientes cienciogramas, dando lugar a lo que hemos dado en llamar cienciogramas factoriales (CF).

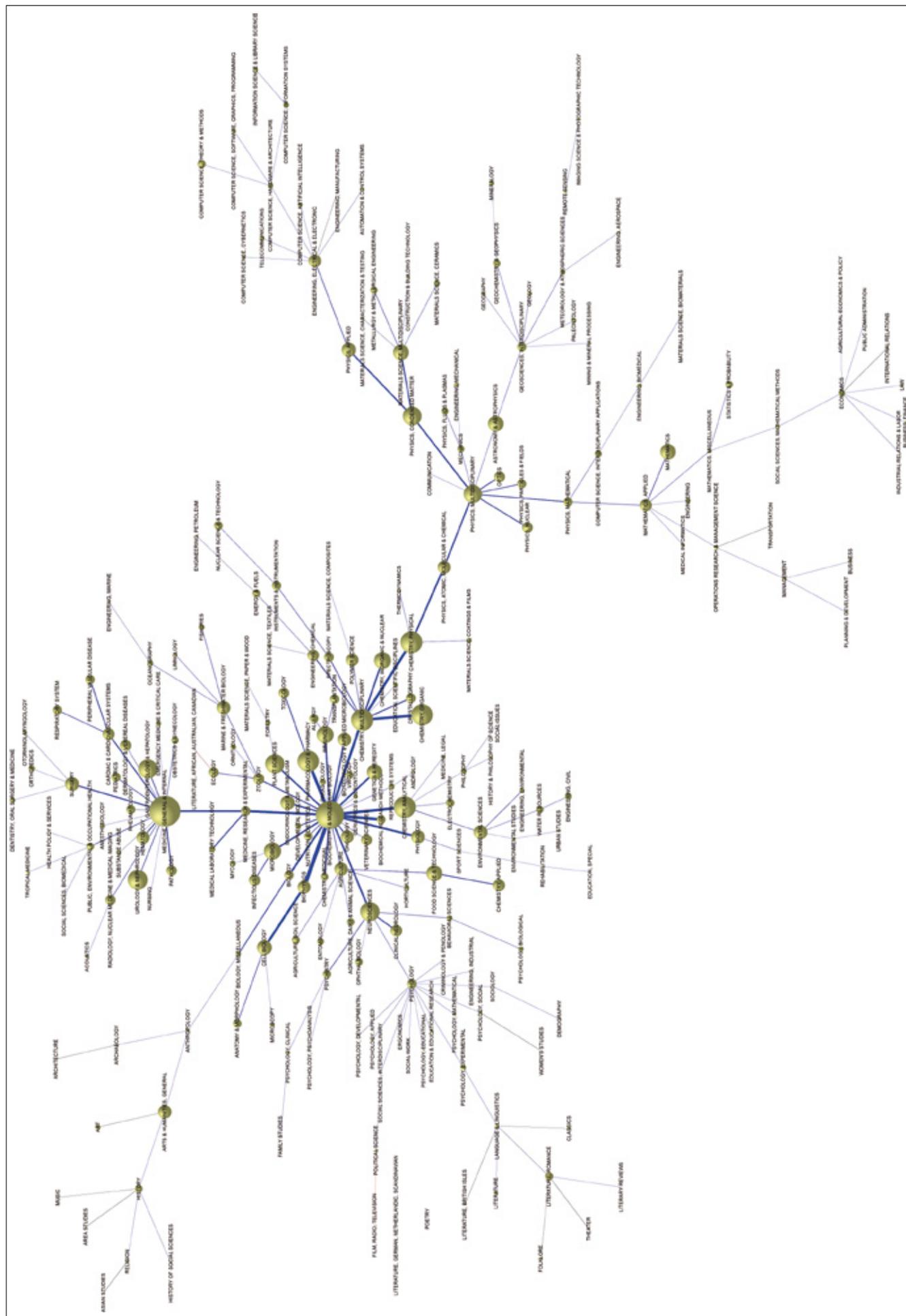
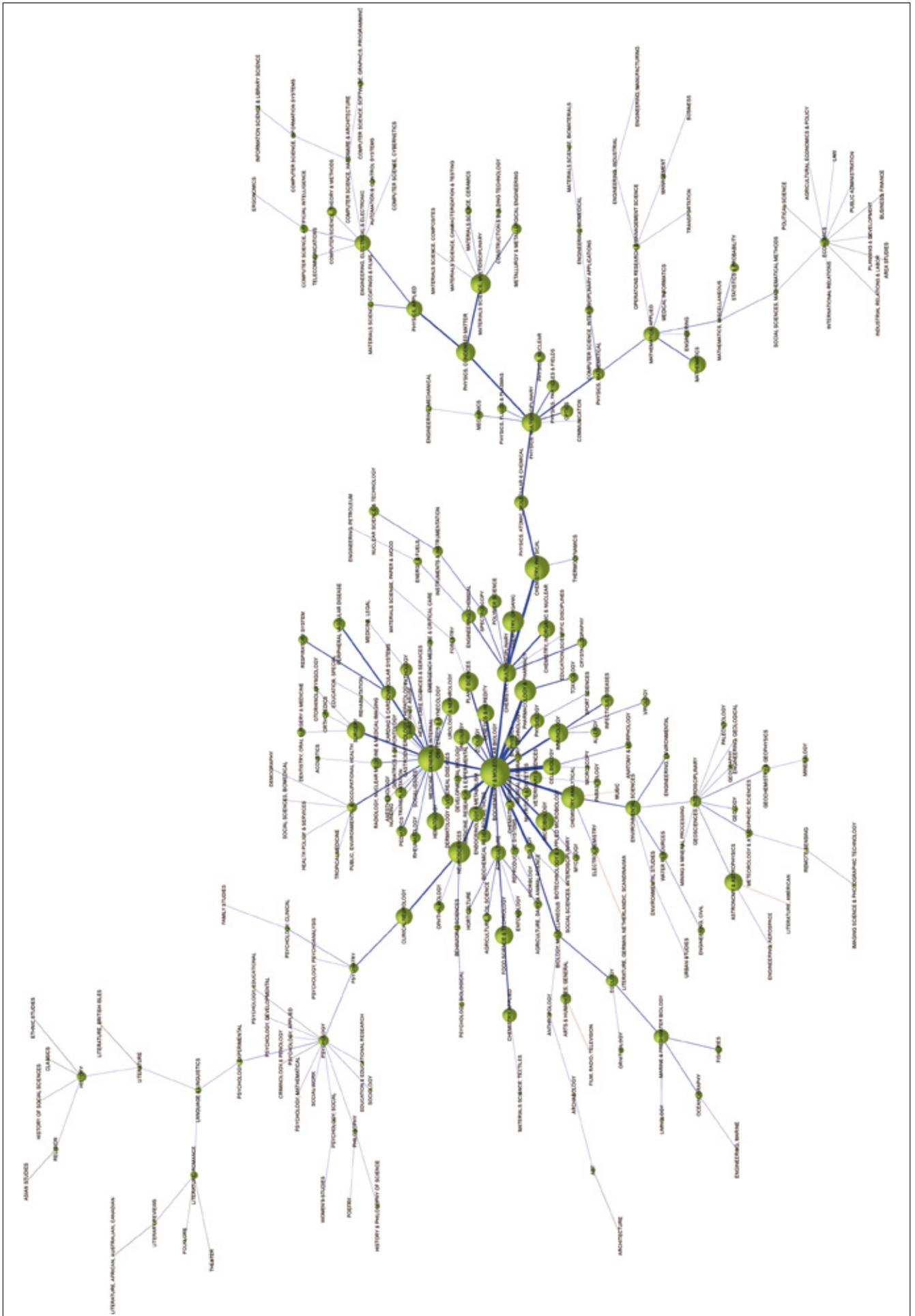


Ilustración 1. Cienciograma del dominio geográfico español, 1990-1995

Ilustración 2. Cienciograma del dominio geográfico español, 1996-2000



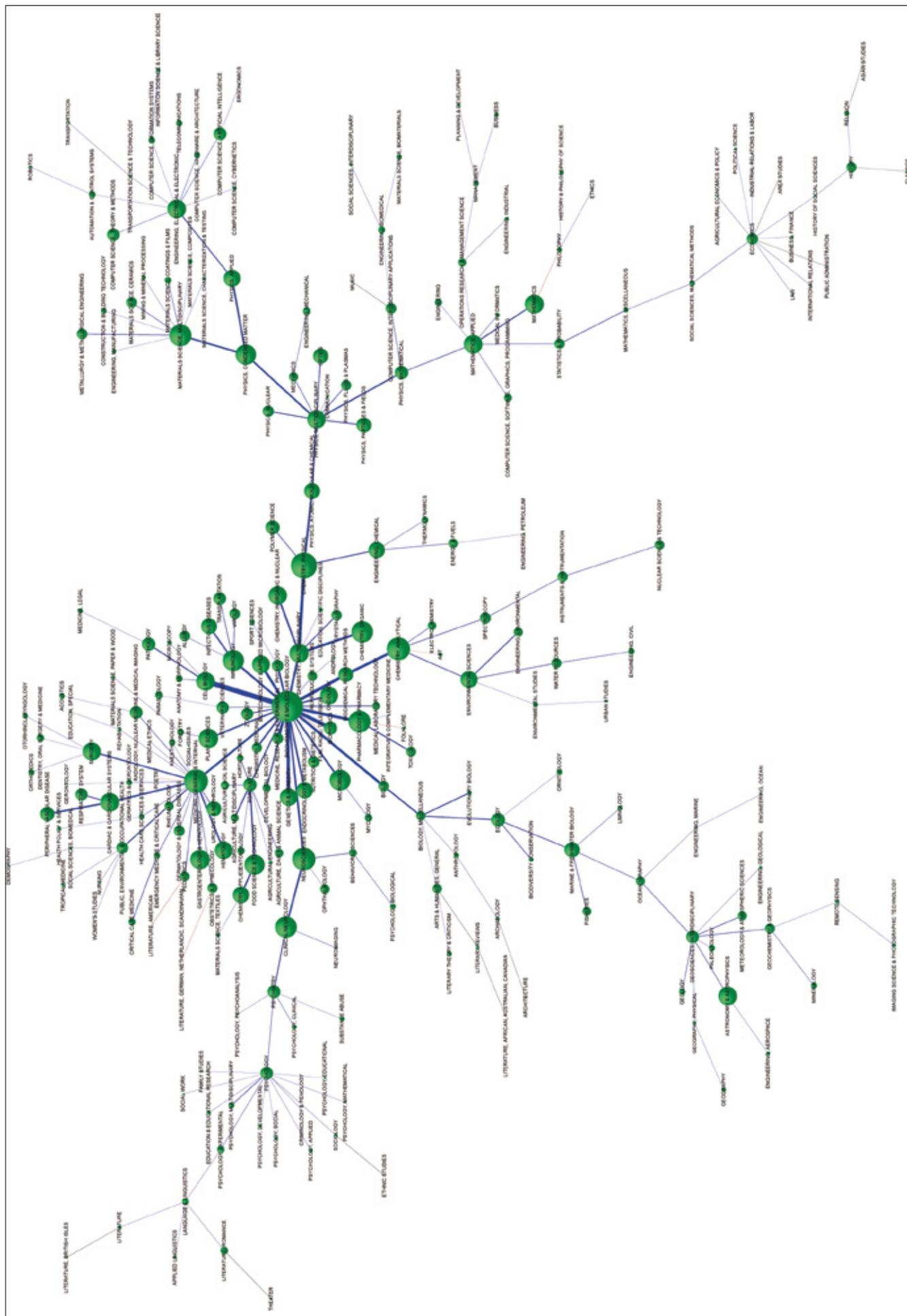


Ilustración 3. Cienciograma del dominio geográfico español, 2001-2005

## 4.2. Cienciogramas factoriales

De acuerdo con el procedimiento explicado en **Moya-Anegón, et al. (2006)**, extraemos y categorizamos diez factores del período 1990-1995, otros diez en el de 1996-2000 y once, en el de 2001-2005. El nombre de cada factor, junto con su varianza individual y acumulada en cada caso, es el que se muestra en las tablas 1, 2 y 3.

Comparando las tres tablas anteriores, podemos ver que los dos primeros períodos son prácticamente iguales. Pues cuando más, difieren levemente en la denominación de un factor, así como en la varianza explicada por alguno uno de ellos. Las diferencias del tercer período con respecto a los dos anteriores son algo mayores. En primer lugar aparece un nuevo factor como es *Política sanitaria y servicios médicos*, que se desgaja de *Biomedicina*. Además, es posible observar un movimiento descendente, y por tanto de pérdida de correlación entre sus categorías, en los factores relacionados con la agricultura y las humanidades. Por el contrario, se aprecia un movimiento ascendente, y por tanto de correlación de sus componentes, en los factores relacionados con las ciencias del medioambiente. Esto es debido a una mayor profusión en la utilización de fuentes externas a sus propios factores, en el primer caso, y lo contrario en el segundo.

El cienciograma factorial (CF) del período 1990-1995 deja veintiséis categorías sin factorizar (de color gris, el de 1996-2000 treinta y seis, y el de 2001-2005, 29). No obstante, de las veintiséis primeras categorías del período 1990-1995, sólo siete siguen sin factorizar entre 2001-2005. Esto quiere decir que de las veintiséis categorías que intervenían muy poco en la investigación española en el período 1990-1995, como pone de manifiesto el hecho de no ser factorizadas por su bajo peso, sólo siete continúan haciéndolo hoy día. El resto, como consecuencia de la evolución de la investigación, han ido ganando peso a lo largo del tiempo. No obstante, y precisamente a raíz del desarrollo de la investigación científica, surgen nuevas categorías, que de momento en el dominio español tienen poca relevancia. Esta es la razón del alto número de categorías sin factorizar que se recogen en el período 2001-2005.

## 4.3. Macro-estructura

Los cienciogramas de los tres períodos muestran la típica distribución bibliométrica de pocas áreas temáticas de gran tamaño y muchas de dimensión reducida. Este tamaño diferencial refleja la naturaleza hiperbólica esencial de todas las distribuciones bibliométricas ya explicada por **Small y Garfield (1985)**. Al mismo tiempo, existe una gran área temática central que hace de punto de interconexión de otras más pequeñas que la rodean. En la medida en que avanzamos temporalmente en los CFs

1990-1995			
Factor	Nombre de factor	% varianza	% de varianza acumulada
1	Biomedicina	23,1	23,1
2	Ciencias de los materiales y química aplicada	12,9	36,0
3	Psicología	7,8	43,8
4	Agricultura y ciencias del suelo	6,3	50,1
5	Informática y telecomunicaciones	4,1	54,2
6	Ciencias de la tierra y del espacio	3,1	57,3
7	Gestión, derecho y economía	3,3	60,6
8	Arte y humanidades	2,4	63,0
9	Biología animal y ecología	2,4	65,4
10	Física nuclear y de partículas	1,5	66,9

Tabla 1. Factores extraídos del dominio geográfico español 1990-1995

se confirma la observación a la que hacíamos referencia anteriormente en los cienciogramas no factoriales: el aumento del número de categorías que ocupan posiciones centrales. Esto no se debe tanto a un incremento de las pertenecientes al área temática central, pues su número sólo aumenta en cinco a lo largo de todo el período, sino más bien a que categorías de otras áreas como *Biología animal y ecología*, o *Agricultura y ciencias del suelo*, pasan a ocupar posiciones relevantes o centrales, a raíz de compartir fuentes, fundamentalmente, con otras que ya ocupaban esta posición. De esta forma consiguen mejorar su universalidad e interés en la investigación del dominio, provocando a su vez el engrosamiento del núcleo.

Las posiciones relativas que ocupan las áreas temáticas, son prácticamente las mismas en los tres períodos. Sólo se diferencian en que durante 1990-1995, las *Ciencias de la Tierra y del espacio* se sitúan en la zona derecha, mientras que en los otros dos, aparecen en la parte central. No obstante, sí que hay diferencias en la forma en que éstas se interconectan. En 1990-1995, la *Psicología* parece tener una orientación más dedicada al estudio de la mente, de ahí su conexión con la *Biomedicina: Biochemistry & molecular biology* ↔ *Neurosciences* ↔ *Psychology*, mientras que en los intervalos 1996-2000 y 2001-2005, la investigación en esta área se reorienta a los estudios clínicos y patológicos como muestra el cambio de su ruta: *Biochemistry & molecular biology* ↔ *Neurosciences* ↔ *Clinical*

1996-2000			
Factor	Nombre de factor	% varianza	% de varianza acumulada
1	Biomedicina	23,1	23,1
2	Ciencias de los materiales y física aplicada	13,2	36,3
3	Psicología	8,2	44,5
4	Agricultura y ciencias del suelo	6,8	51,3
5	Informática y telecomunicaciones	4,4	55,7
6	Ciencias de la tierra y del espacio	4,0	59,7
7	Gestión, derecho y economía	3,4	63,1
8	Humanidades	3,1	66,2
9	Biología animal y ecología	2,3	68,5
10	Física nuclear y de partículas	1,4	69,9

Tabla 2. Factores extraídos del dominio geográfico español 1996-2000

2001-2005			
Factor	Nombre de factor	% varianza	% de varianza acumulada
1	Biomedicina	22,1	22,1
2	Ciencias de los materiales y física aplicada	12,3	34,4
3	Psicología	7,7	42,1
4	Ciencias de la tierra y del espacio	6,9	49,0
5	Informática y telecomunicaciones	4,4	53,4
6	Agricultura y ciencias del suelo	4,1	57,5
7	Gestión, derecho y economía	3,4	60,9
8	Biología animal y ecología	2,7	63,6
9	Política sanitaria y servicios médicos	2,1	65,7
10	Humanidades	1,9	67,6
11	Física nuclear y de partículas	1,7	69,3

Tabla 3. Factores extraídos del dominio geográfico español 2001-2005

**“Los cienciogramas de los tres períodos muestran la típica distribución bibliométrica de pocas áreas temáticas de gran tamaño y muchas de dimensión reducida”**

*neurology ↔ Psychiatry ↔ Psychology*. En el primer período, las *Humanidades* aparecen conectadas con la *Biología animal y la ecología* a través de la *Antropología: History ↔ Arts & humanities ↔ Anthropology ↔ Biology miscellaneus ↔ Biology*, mientras que en los dos restantes se muestran unidas a la *Psicología* por medio de la categoría *Language & linguistic*, indicando así un cambio en la investigación y en las fuentes utilizadas en esta área. El área temática de las *Ciencias de la Tierra y del espacio* también sufre cambios a lo largo de este período de estudio. Entre 1990 y 1995, aparece vinculada a *Ciencias de los materiales y física aplicada: Geosciences interdisciplinary ↔ Astronomy & astrophysics ↔ Physics multidisciplinary*. En el espacio comprendido entre 1996 y 2000 su conexión más fuerte la establece con la *Biomedicina: geosciences interdisciplinary ↔ Environmental sciences ↔ Chemistry analytical ↔ Biochemistry & molecular biology*. Finalmente, en el período 2001-2005, su investigación básica está compartida con el área temática de la *Biolo-*

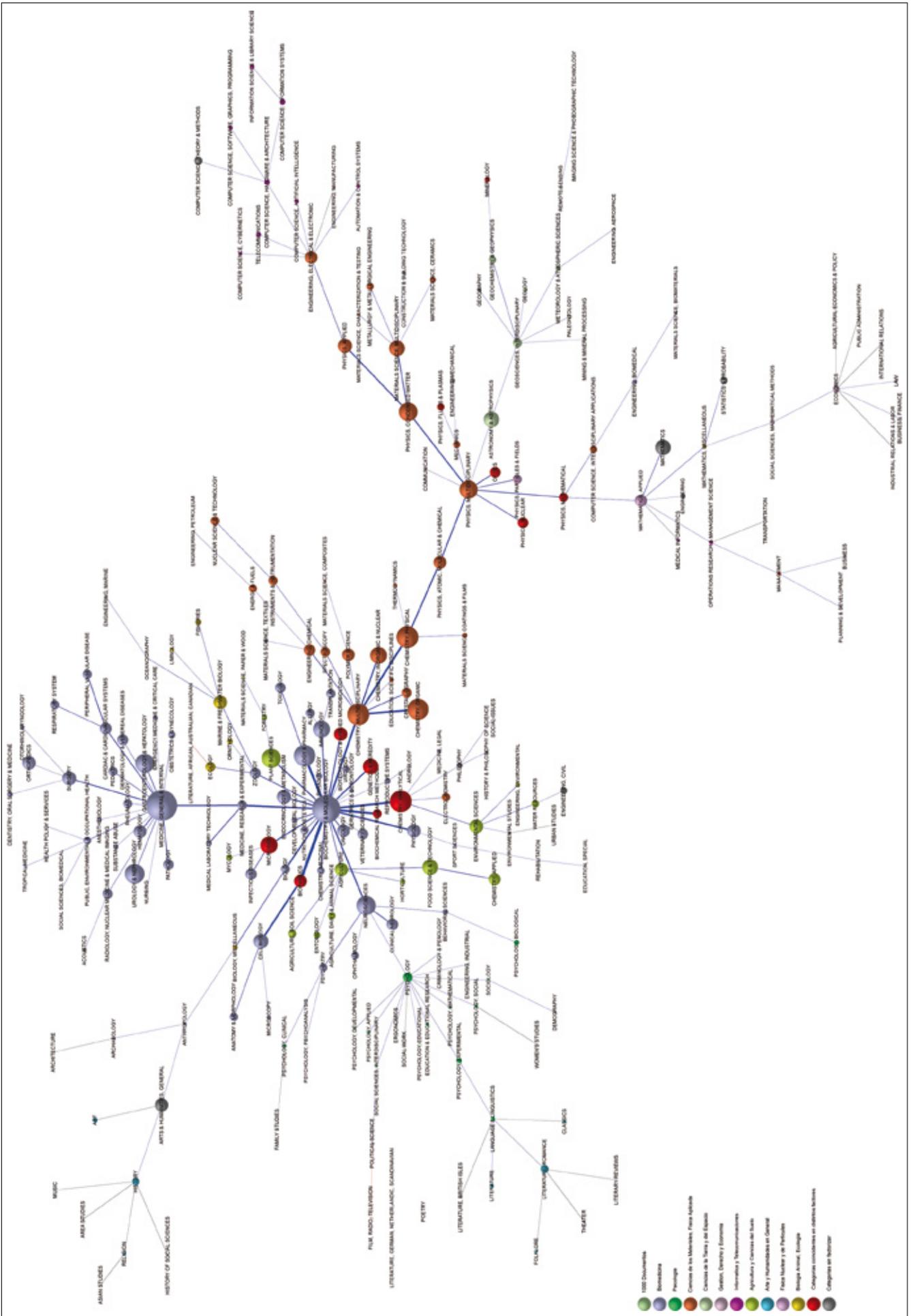
*gía animal y ecología: Geosciences ↔ Oceanography ↔ Marine & freshwater biology ↔ Ecology ↔ Biology miscellaneus ↔ Biology*.

A simple vista, el área que ocupa una posición más central en los tres períodos es *Biomedicina*. Esto nos indica que se trata del área temática más universal, la que más fuentes comparten el resto y por tanto, la que más conocimiento aporta a las demás. En definitiva, la que más interviene en el desarrollo del dominio científico español. Esta posición nuclear de la *Biomedicina*, se corresponde con la detectada por **Boyack, Klavans y Börner** (2005) en el mundo y por **Vargas-Quesada y Moya-Anegón** (2007) en EUA y Europa. El grado de universalidad/exclusividad del resto de áreas temáticas, está en función de su distancia al centro. Permaneciendo este invariable en los tres períodos, con la ligera excepción del último, al contar este con una nueva área temática desgajada de la *Biomedicina*.

Las categorías de color rojo de los cienciogramas indican los puntos de interacción o intercambio de información entre las distintas áreas temáticas. Partiendo de las categorías objeto de interacción del primer período, podemos estudiar su evolución a lo largo del tiempo.

Entre 1990 y 1995, como consecuencia de las inquietudes científicas de los investigadores en ese momento, son quince las categorías que muestran una ac-

Ilustración 4. Cienciograma factorial del dominio español 1990-1995  
 Disponible para el lector en tamaño real en la web. <http://www.scimago.es/benjamin/sp90-95.jpg>



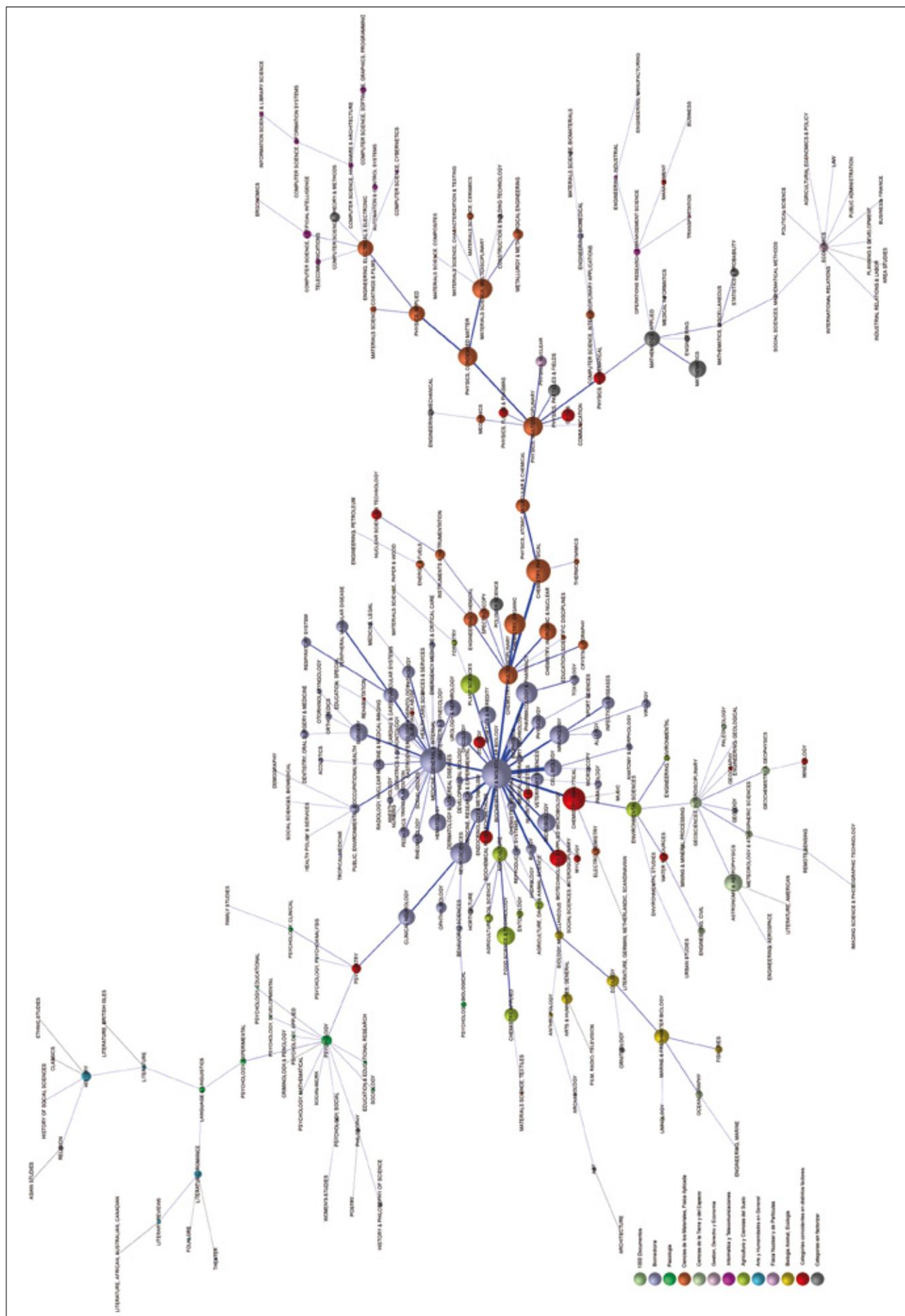
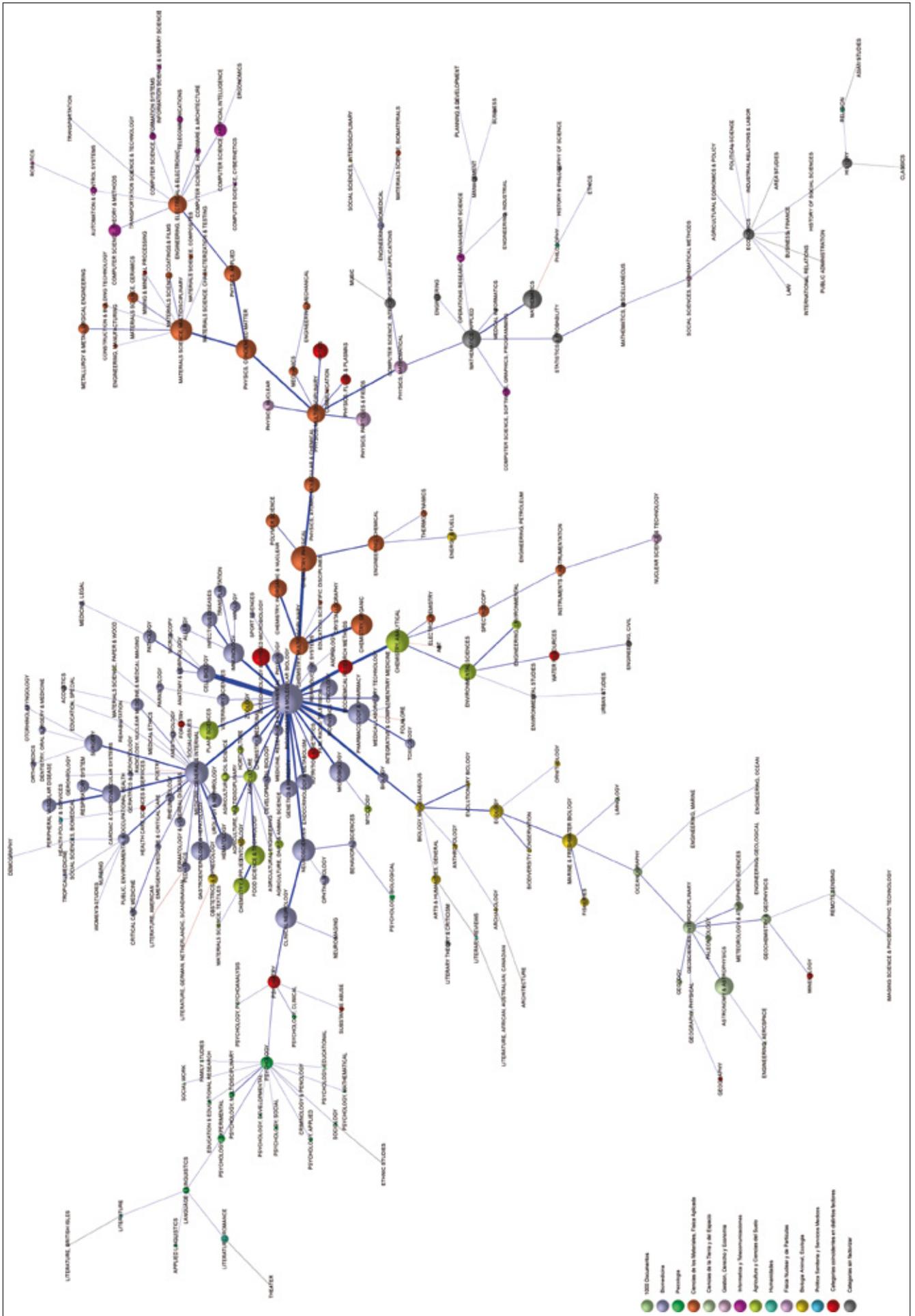


Ilustración 5. Cienciagrama factorial del dominio español 1996-2000 Disponible para el lector en tamaño real en la web. <http://www.scimago.es/benjamin.in/sp96-00.jpg>

Ilustración 6. Cienciograma factorial del dominio español 2001-2005  
 Disponible para el lector en tamaño real en la web. <http://www.scimago.es/benjamin/sp01-05.jpg>



tividad de intercambio de fuentes entre distintas áreas temáticas. Fundamentalmente son dos los emparejamientos entre áreas temáticas que se reflejan a raíz de esta actitud innovadora: *Biomedicina con Agricultura y ciencias del suelo*, y *Ciencias de los materiales y física aplicada con Física nuclear y de partículas*.

En el período 1996-2000, son diecinueve las categorías objeto de intercambio de información. Ocho de ellas, en negrita, con sus correspondientes emparejamientos de áreas temáticas, coinciden con el período 1990-1995. De las once nuevas categorías que aparecen, algunas son la continuación de los intercambios que se producían en el período anterior, pero otras ponen de manifiesto una nueva línea de investigación mediante el emparejamiento de las áreas: *Biomedicina y Psicología*.

Por último, en el intervalo 2001-2005, encontramos dieciséis categorías responsables del flujo de información entre distintas áreas temáticas. Cinco de ellas, en negrita, son comunes a los dos períodos anteriores y ponen en contacto a las mismas áreas temáticas. Seis, en cursiva, se repiten en el período anterior, y también relacionan a las mismas áreas temáticas en ambos períodos. Y finalmente, las cinco últimas, muestran lo que podría considerarse como un nuevo emparejamiento y por tanto el inicio de un nuevo foco de investigación: *Agricultura y ciencias del suelo con Biología animal y ecología*.

#### 4.4. Microestructura

A nivel de categorías, los tres cienciogramas son un claro ejemplo de la naturaleza hiperbólica de las distribuciones bibliométricas. De igual modo, en los tres casos, se produce una mayor concentración de categorías en la zona central y derecha que en la izquierda, presentando al mismo tiempo la ya típica estructura centro-periferia de este tipo de cienciogramas.

Tanto a simple vista como por medio de la centralidad de grado (Wasserman; Faust, 1998), detectamos que *Biochemistry & molecular biology* es la categoría más central con un grado nodal de 26, en los tres cienciogramas. Esto hace que dicha categoría sea la más universal de todas, la que más fuentes comparte con el resto y, en consecuencia, la que más interviene en el desarrollo científico de cada período temporal. Al obtener el mismo grado en los tres casos, no podemos decir que destaque sobre el resto en ningún período, sino que en todos lo hace por igual. El grado de universalidad y participación que el resto de categorías del cienciograma tienen en el desarrollo científico de cada período, es directamente proporcional a su distancia geodésica con respecto a esta categoría. Si tenemos en cuenta que las redes de tipo centro-periferia presentan estructuras muy estables (García Muñoz; Ramos Carvajal, 2004), es decir, son resistentes a los cambios, la desaparición y/o aparición de nuevas categorías en el núcleo, no hace más que mostrarnos la evolución de las distintas líneas de investigación del

dominio científico español. Por ejemplo, la aparición de *Film, radio, televisión* como una de las categorías que constituyen el núcleo de la *Biomedicina* en el tercer período, nos indica el inicio y consolidación del vídeo-diagnóstico y de la vídeo-cirugía en este último.

## 5. Columna vertebral

La columna vertebral de un dominio se construye a partir del valor más alto del enlace que une dos áreas temáticas y eliminando todos aquellos que quedan por debajo, junto con las categorías que resultan desconectadas (Moya-Anegón, et al., 2006). Desde la perspectiva de su evolución, podemos decir que la columna vertebral de la investigación del dominio español, se afianza y se desarrolla con el paso del tiempo.

En el período 1990-1995 (ilustración 7), son dos las áreas temáticas que componen la columna vertebral de la ciencia española: *Biomedicina* y *ciencias de los materiales*, y está integrada por siete categorías. No obstante, se puede observar una tercera área incipiente: *Agricultura y ciencias del suelo*, pero que no está del todo consolidada, pues todas las categorías que la componen son de tipo interdisciplinar. Es decir, pertenecen también al área de *Biomedicina*, de ahí su color rojo.

En el intervalo 1996-2000 (ilustración 8) la situación es muy parecida a la anterior, donde la columna vertebral de la investigación española está compuesta por dos áreas temáticas consolidadas y una emergente. No obstante, se produce un aumento en el número de categorías de *Biomedicina*: (tres), al tiempo que la reducción a sólo una, en el área de la *Agricultura y ciencias del suelo*. Por lo que el número total de categorías aumenta con respecto al período anterior: diez frente a siete.

En el último período (ilustración 9), las tres áreas temáticas aparecen consolidadas, pues al menos una de las categorías que conforman *Agricultura y ciencias del suelo*, ha dejado de ser interdisciplinar, de ahí su color verde. El número de categorías que integran estas tres áreas, también crece respecto al período anterior: trece frente a diez.

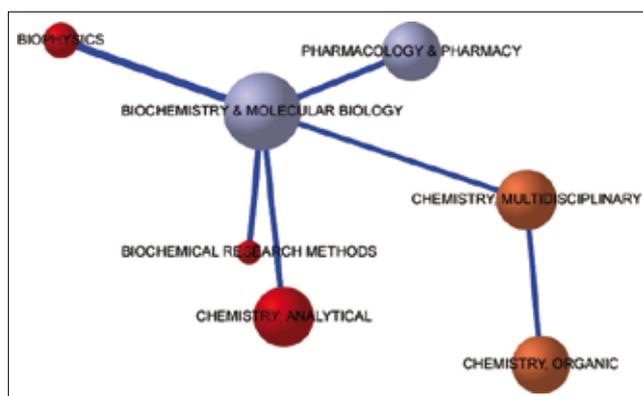


Ilustración 7. Columna vertebral de la investigación española 1990-1995

En esencia, la columna vertebral de la investigación española no cambia en los tres períodos estudiados. Sólo el último refleja la consolidación de un área incipiente como es *Agricultura y ciencias del suelo*. No obstante, desde un punto de vista microestructural sí que hay cambios. En la medida en que se avanza en el tiempo, se observa un aumento del número de categorías en las áreas de investigación ya consolidadas: *Biomedicina* y *Ciencias de los materiales*. De igual modo, se producen cambios de adscripción temática en las categorías de cada área. Esto da como resultado una estructura intelectual de la ciencia más acorde a la observada en los cienciogramas de otros países como EUA, Francia, Alemania o Inglaterra.

## 6. Conclusiones

En este artículo hemos mostrado la evolución del dominio científico español a partir de la producción científica recogida en las bases de datos del *ISI* desde 1990 hasta 2005. Se han identificado sus principales áreas temáticas, categorías y esquema de vertebración para posteriormente mostrar su desarrollo por medio de su comparación. A nivel general, podemos decir que la estructura científica española se asemeja bastante a la de otros grandes dominios como el de EUA y Europa, vistos en **Vargas-Quesada** y **Moya-Anegón** (2007). La reciente inclusión por parte del *ISI* de nuevas revistas españolas en sus bases de datos, especialmente en *Social Science Citation Index* y *Arts & Humanities Citation Index*, incrementarán esta similitud, además

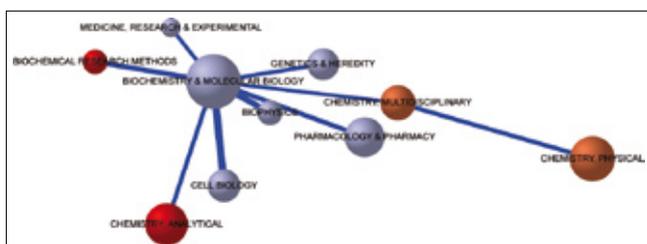


Ilustración 8. Columna vertebral de la investigación española 1996-2000

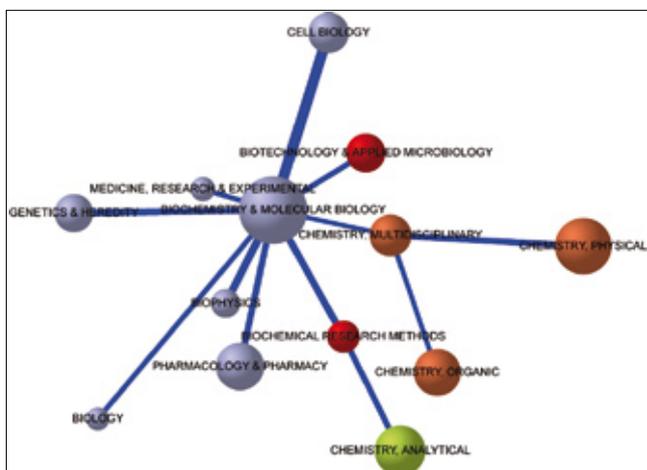


Ilustración 9. Columna vertebral de la investigación española 2001-2005

## “La columna vertebral de la investigación española no cambia en los tres períodos estudiados. Sólo el último refleja la consolidación de un área incipiente como es Agricultura y ciencias del suelo”

de atenuar los problemas de falta de información referencial ya comentados.

Desde un principio, el área temática de *Biomedicina* se constituye como el núcleo de la estructura intelectual de la investigación española. Esto se debe a que es la que más recursos o documentos fuente exporta al resto, lo cual la convierte en la más interdisciplinar o central por su mayor grado o número de enlaces. Por otro lado, también es aquella con mayor producción científica, como corroboran los trabajos de: **Gómez**, et al. (2002); **Butler**, (2004); **Moya-Anegón**, et al. (2004); **Gómez**, et al. (2004); **Moya-Anegón**, et al. (2005); **Chinchilla-Rodríguez**; **Moya-Anegón**, (2007). Según hemos podido constatar en otros cienciogramas, esta posición central de *Biomedicina* es algo propio de países desarrollados como: EUA, Francia, Reino Unido, Alemania, Japón (**Moya-Anegón**, et al., 2005); y no tanto de otros en vías de desarrollo como: China, Cuba y Venezuela (**Vargas-Quesada**, et al., 2006); (*SCImago Group*, 2007). En nuestra opinión, esto se debe al esfuerzo que los países desarrollados dedican al sostenimiento e incremento de los servicios de la sociedad del bienestar, entre los que se encuentran la salud y la sanidad (Comisión Europea, 2003).

El aumento del número de categorías que se producen en el área de *Biomedicina*, tanto a nivel macro-estructural, como desde el punto de vista del esquema de vertebración, muestra tendencias similares a las observadas en los cienciogramas de otros dominios con una investigación más desarrollada, como son EUA y Europa. *Biochemistry & molecular biology* es la categoría más prominente en esta área y en todos los cienciogramas, poniendo de manifiesto que es la que más fuentes comparte con el resto y por tanto, la que más interviene en el desarrollo científico del dominio.

La segunda área temática más influyente del dominio científico español es *Ciencias de los Materiales*. También es la segunda más productiva: **Gómez**, et al. (2002); **Butler**, (2004); **Moya-Anegón**, et al. (2004); **Gómez**, et al. (2004); **Moya-Anegón**, et al. (2005); **Chinchilla-Rodríguez**; **Moya-Anegón**, (2007), lo cual es tan lógico como necesario, pues es la base para el desarrollo industrial y tecnológico de un dominio. Desde el punto de vista de la columna vertebral de la investigación española, el número de categorías que

pertencen a esta área, se incrementa con el tiempo. No obstante, su crecimiento es lento y debe aumentar más para parecerse al de otros dominios más desarrollados. Su categoría más prominente o central es *Chemistry multidisciplinary*.

*Agricultura y ciencias del suelo*, en dura pugna con *Biología animal y ecología*, termina por constituirse como la tercera área temática más influyente del dominio español. En los cienciogramas de otros dominios desarrollados, como los mencionados anteriormente, ocurre lo contrario. De esta forma, se pone de manifiesto la relevancia que *Agricultura y ciencias del suelo* está empezando adquirir en la investigación y economía española, frente a otros países. No obstante hay que tener en cuenta que se trata de un área joven, consolidada como tal en el último período estudiado, y que la tendencia es imitar los cienciogramas y esquemas de los dominios más desarrollados, por lo que podría verse desplazada por el área de *Biología animal y ecología* en un período posterior. Su categoría más prominente es *Chemistry analytical*.

Agradecimientos: Este trabajo ha sido posible gracias a la financiación pública obtenida mediante el *Plan Nacional I+D+I: SEJ2004-08358-CO2-01/SOC*.

## Bibliografía

**Aaronson, Steve.** "The footness of science". En: *Mosaic*, 1975, March-April, v. 6, pp. 22-27.

**Boyack, Kevin W.; Klavans, Richard; Börner, Katy.** "Mapping the backbone of science". En: *Scientometrics*, 2005, v. 64, n. 3, pp. 351-374.

**Braam, Robert R.; Moed, Henk F.; Van Raan, Anthony F. J.** "Mapping of science by combined co-citation and word analysis. I: structural aspects". En: *Journal of the American Society for Information Science (Jasis)*, 1991a, v. 42, n. 4, pp. 233-251.

**Braam, Robert R.; Moed, Henk F.; Van Raan, Anthony F. J.** "Mapping of science by combined co-citation and word analysis. II: Dynamic aspects". En: *Journal of the American Society for Information Science (Jasis)*, 1991b, v. 42, n. 4, pp. 252-266.

**Brandenburg, Franz Josef; Himsolt, Michael; Rohrer, Christoph.** "An experimental comparison of force-directed and randomized graph drawing algorithms". En: *Lecture notes in computer science*, 1995, v. 1.027, n. 76, pp. 87.

**Braun, Tibor; Glänzel, Wolfgang; Schubert, Andras.** "How balanced is the Science Citation Index's journal coverage? a preliminary overview of macrolevel statistical data". En: **B. Cronin; Atkins, H. B.** (eds.). *The web of knowledge: a festschrift in honor of Eugene Garfield*. New Jersey: Information Today, 2000.

**Buter, Renald K.; Noyons, Ed C. M.** "Improving the functionality of interactive bibliometric science maps". En: *Scientometrics*, 2001, v. 51, n. 1, pp. 55-67.

**Butler, Linda.** *ARC-supported research: the impact of journal publication output 1996-2000*, Australian Research Council. Consultado en: 23-03-07. [http://www.arc.gov.au/pdf/arc\\_supported\\_research.pdf](http://www.arc.gov.au/pdf/arc_supported_research.pdf)

**Buzydowski, Jan; White, Howard D.; Lin, Xia.** "Term co-occurrence analysis as an interface for digital libraries". En: *Lecture notes in computer science series*, 2002, v. 2.539, pp. 133-144.

*Centre for Science and Technology Studies (CWTS)*. Consultado en: 18-02-07. <http://www.cwts.nl>

**Chen, Chaomei.** "Bridging the gap: the use of pathfinder networks in visual navigation". En: *Journal of visual languages and computing*, 1998a, v. 9, pp. 267-286.

**Chen, Chaomei.** "Generalised similarity analysis and pathfinder network scaling". En: *Interacting with computers*, 1998b, v. 10, pp. 107-128.

**Chen, Chaomei.** "Visualising semantic spaces and author co-citation networks in digital libraries". En: *Information processing & management*, 1999, v. 35, pp. 401-420.

**Chen, Chaomei; Kuljis, Jasna.** "The rising landscape: a visual exploration of superstring revolutions in physics". En: *Journal of the American Society for Information Science and Technology (Jasist)*, 2003, v. 54, n. 5, pp. 435-446.

**Chen, Chaomei; Morris, Steven.** "Visualizing evolving networks: Minimum Spanning Trees versus Pathfinder Networks". En: *IEEE Symposium on information visualization*, 2003, pp. 67-74.

**Chen, Chaomei; Paul, R. J.** "Visualizing a knowledge domain's intellectual structure". En: *Computer*, 2001, v. 34, n. 3, pp. 65-71.

**Chen, Chaomei; Paul, Ray J.; O'Keefe, Robert M.** "Fitting the jigsaw of citation: information visualization in domain analysis". En: *Journal of the American Society for Information Science and Technology (Jasist)*, 2001, v. 52, n. 4, pp. 315-330.

*Comisión Europea (UE).* *Third European report on science & technology indicators, 2003: towards a knowledge-based economy*. Luxembourg: Office for Official Publications of the European Communities, 2003.

**Ding, Ying; Chowdhury, Gobinda G.; Foo, Schubert.** "Bibliometric cartography of information retrieval research by using co-word analysis". En: *Information processing & management*, 2001, v. 37, pp. 801-817

**Fowler, R. H.; Dearhold, D. W.** "Information retrieval using path finder networks". En: **Schvaneveldt, R. W.** (ed.). *Pathfinder associative networks; studies in knowledge organization*. Ablex (NJ): Norwood, 1990.

**García-Muñiz, Ana-Salomé; Ramos-Carvajal, Carmen.** "Core/periphery structure models: an alternative methodological proposal". En: *Social networks*, 2006, v. 28, pp. 442-448.

**Garfield, Eugene.** "Citation indexes in sociological and historical research". En: *American documentation*, 1963, v. 14, n. 4, pp. 289-291.

**Garfield, Eugene.** "Introducing the ISI Atlas of Science: biochemistry and molecular biology, 1978-1980". En: *Current contents*, 1981, n. 42, pp. 5-13.

**Garfield, Eugene.** "Introducing the ISI Atlas of Science: biotechnology and molecular genetics, 1981/82 and bibliographic update for 1983-1984". En: *Current contents*, 1984, n. 41, pp. 3-15.

**Garfield, Eugene.** "The encyclopedic ISI-Atlas of Science launches 3 new sections: biochemistry, immunology, and animal and plant sciences". En: *Current contents*, 1988, n. 7, pp. 3-8.

**Garfield, Eugene.** "Scientography: mapping the tracks of science". En: *Current contents: social & behavioral sciences*, 1994, v. 7, n. 45, pp. 5-10.

**Garfield, Eugene.** *Mapping the world of science (at the 150 Anniversary Meeting of the AAAS, Philadelphia, PA)*. Consultado en: 19-02-07. <http://www.garfield.library.upenn.edu/papers/mapsciworld.html>

**Garfield, Eugene; Sher, Irving H.; Torpie, Richard J.** *The use of citation data in writing the history of science*. Philadelphia: Institute for Scientific Information, 1964.

**Glänzel, Wolfgang; Schubert, A.; Czerwon, H. J.** "An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis". En: *Scientometrics*, 1999, v. 44, n. 3, pp. 427-439.

**Glänzel, Wolfgang; Schubert, A.; Czerwon, H. J.; Shoepflin, U.** "An item-by-item subject classification of papers published in journals covered by the SSCI database using reference analysis". En: *Scientometrics*, 1999, v. 46, n. 3, pp. 431-441.

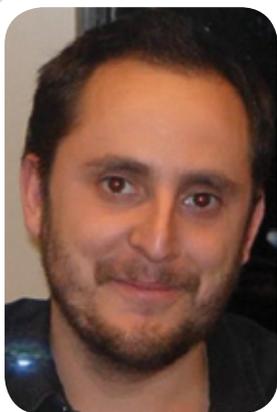
**Gómez, Isabel; Fernández, M. Teresa; Bordons, María; Cabrero, A.; Morillo, Fernanda; Rojo, Raquel.** *La actividad científica del CSIC a través del Science Citation Index, Social Sciences Citation Index y Arts & Humanities Citation Index*. Estudio bibliométrico del período 1997-2000. Madrid: Cindoc, 2002.

- Gómez, Isabel; Fernández, M. Teresa; Bordons, María; Cabrero, A.; Morillo, F.** *Proyecto de obtención de indicadores de producción científica y tecnológica de España*. Madrid, Cindoc, 2004.
- Granoveter, Mark S.** "The strength of weak ties". En: *American journal of sociology*, 1973, v. 78, n. 6, pp. 1.360-1.380.
- Griffith, Bellver C.; Small, Henry; Stonehill, Judy A.; Dey, Sandra.** "The structure of scientific literature, II: toward a macro and microstructure for science". En: *Science studies*, 1974, v. 4, pp. 339-365.
- Guerrero-Bote, Vicente P.; Moya-Anegón, Félix de; Herrero-Solana, Víctor.** "Document organization using Kohonen's algorithm". En: *Information processing & management*, 2002, v. 38, pp. 79-89.
- Kamada, Tomihisa; Kawai, Satoru.** "An algorithm for drawing general undirected graphs". En: *Information processing letters*, 1989, v. 31, n. 1, pp. 7-15.
- Leydesdorff, Loet.** "Mapping interdisciplinarity at the interfaces between the Science citation index and the Social science citation index". En: *Scientometrics*, 2007, v. 71, n. 3.
- Leydesdorff, Loet; Vaughan, Liwen.** "Co-occurrence matrices and their applications in information science: extending ACA to the web environment". En: *Journal of the American Society for Information Science and Technology (Jasist)*, 2006, v. 57, n. 12, pp. 1.616-1.628.
- Lin, Xia; White, Howard D.; Buzdyłowski, Jan.** "Real-time author co-citation mapping for online searching". En: *Information processing & management*, 2003, pp. 689-706.
- Morris, Steven A.; Yen, G.; Wu, Zheng; Asnake, Benyam.** "Time line visualization of research fronts". En: *Journal of the American Society for Information Science and Technology (Jasist)*, 2003, v. 54, n. 5, pp. 413-422.
- Moya-Anegón, Félix de; Vargas-Quesada, Benjamín; Herrero-Solana, Víctor; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; Muñoz-Fernández, Francisco J.** "A new technique for building maps of large scientific domains based on the cocitation of classes and categories". En: *Scientometrics*, 2004, v. 61, n. 1, pp. 129-145.
- Moya-Anegón, Félix de; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; Herrero-Solana, Víctor; Muñoz-Fernández, Francisco J.; Navarrete-Cortés, José; Vargas-Quesada, Benjamín.** *Indicadores bibliométricos de la actividad científica española 1998-2002*. Madrid: Fundación Española para la Ciencia y la Tecnología (Fecyt), 2004.
- Moya-Anegón, Félix de; Vargas-Quesada, Benjamín; Chinchilla-Rodríguez, Zaida; Herrero-Solana, Víctor; Corera-Álvarez, Elena; Muñoz-Fernández, Francisco J.** "Domain analysis and information retrieval through the construction of heliocentric maps based on ISI-JCR category cocitation". En: *Information Processing & Management*, 2005, v. 41, n. 6, pp. 1520-1533.
- Moya-Anegón, Félix de; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; Herrero-Solana, Víctor; Muñoz-Fernández, Francisco J.; Vargas-Quesada, Benjamín.** *Indicadores bibliométricos de la actividad científica española: ISI Web of Science 1995-2003*. Madrid: Fecyt, 2005.
- Moya-Anegón, Félix de; Vargas-Quesada, Benjamín; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; González-Molina, Antonio; Muñoz-Fernández, Francisco; Herrero-Solana, Víctor J.** "Visualización y análisis de la estructura científica española: ISI Web of Science 1990-2005". En: *El profesional de la información*, 2006, v. 15, n. 4, pp. 258-269.
- Noyons, Ed C. M.** "Bibliometric mapping of science in a science policy context". En: *Scientometrics*, 2001, v. 50, n. 1, pp. 83-98.
- Noyons, Ed C. M.; Moed, Henk F.; Luwel, M.** "Combining mapping and citation analysis for evaluative bibliometric purposes: a bibliometric study". En: *Journal of the American Society for Information Science (Jasis)*, 1999, v. 50, n. 2, pp. 115-131.
- Ortega, José-Luis; Aguillo, Isidro F.** "Análisis de co-enlaces: una aproximación teórica". En: *El profesional de la información*, 2006, v. 15, n. 4, pp. 270-277.
- Price, Derek de Solla.** "Networks of scientific papers". En: *Science*, 1965, v. 149, pp. 510-515.
- Samoylenko, I.; Chao, T. C.; Liu, W. C.; Chen, Chaomei M.** "Visualizing the scientific world and its evolution". En: *Journal of the American Society for Information Science and Technology (Jasist)*, 2006, v. 57, n. 11, pp. 1.461-1.469.
- SCImago Group.** "Atlas of science". Consultado en: 23-03-07. <http://www.atlasofscience.net>
- Seiden, L. S.; Swanson, D. R.** "ISI Atlas of Science: pharmacology 1987, vol. 1 - Inst-SCI- Informat". En: *Library quarterly*, 1989, v. 59, n. 1, pp. 72-73.
- Small, Henry.** "A passage through science: crossing disciplinary boundaries". En: *Library trends*, 1999, v. 48, n. 1, pp. 72-108.
- Small, Henry.** "Paradigms, citations and maps of science: a personal history". En: *Journal of the American Society for Information Science and Technology (Jasist)*, 2003, v. 54, n. 5, pp. 394-399.
- Small, Henry; Garfield, Eugene.** "The geography of science: disciplinary and national mappings". En: *Journal of information science*, 1985, v. 11, n. 4, pp. 147-159.
- Small, Henry; Griffith, Bellver C.** "The structure of scientific literature, I: identifying and graphing specialties". En: *Science studies*, 1974, v. 4, pp. 17-40.
- Small, H.; Sweeney, E.** "Clustering the science citation index using co-citations. 1. A comparison of methods". En: *Scientometrics*, 1985, v. 7, n. 3-6, pp. 391-409.
- Small, H.; Sweeney, E.; Greenlee, E.** "Clustering the science citation index using co-citations. 2. Mapping science". En: *Scientometrics*, 1985, v. 8, n. 5-6, pp. 321-340.
- The Thomson Corporation. ISI Web of Knowledge*. Consultado en: 19-02-07. <http://www.isiwebofknowledge.com>
- Van Raan, Anthony F. J.; Tijssen, Robert J. W.** "The neuronal net of neuronal network research. An exercise in bibliometric mapping". En: *Scientometrics*, 1993, v. 26, n. 1, pp. 169-192.
- Vargas-Quesada, Benjamín; Moya-Anegón, Félix de.** *Visualizing the structure of science*. Berlin: Springer, 2007.
- Vargas-Quesada, Benjamín; Moya-Anegón, Félix de; Chinchilla-Rodríguez, Zaida; González-Molina, Antonio.** "Domain analysis by means of the visualization of maps of vast scientific domains". En: *I Conferencia internacional sobre ciencias y tecnologías multidisciplinares de la información*, 2006.
- Wasserman, Stanley; Faust, Katherine.** *Social network analysis: methods and applications*. Cambridge: Cambridge University Press, 1998.
- White, Howard D.** "Toward ego-centered citation analysis". En: **Cronin, B.; Atkins, H. B.** (eds.). *The web of knowledge: a festschrift in honor of Eugene Garfield*. New Jersey: Information Today, 2000.
- White, Howard D.; Buzdyłowski, Jan; Lin, Xia.** "Co-cited author maps as interfaces to digital libraries: designing pathfinder networks in the humanities". En: *IEEE International conference on information visualization*, 2000, pp. 25-30.
- White, Howard D.; McCain, Katherine W.** "Visualization of literatures". En: *Annual review of information systems and technology (Arist)*, 1997, v. 32, pp. 99-168.
- White, Howard D.; McCain, Katherine W.** "Visualizing a discipline: an author co-citation analysis of information science, 1972-1995". En: *Journal of the American Society for Information Science (Jasis)*, 1998, v. 49, n. 4, pp. 327-355.
- Benjamín Vargas-Quesada, Félix de Moya-Anegón, Zaida Chinchilla-Rodríguez, Elena Corera-Álvarez,** *Universidad de Granada.*  
[benjamin@ugr.es](mailto:benjamin@ugr.es), [felix@ugr.es](mailto:felix@ugr.es), [zchinch@ugr.es](mailto:zchinch@ugr.es), [ecorera@ugr.es](mailto:ecorera@ugr.es)
- Vicente Guerrero-Bote,** *Universidad de Extremadura.*  
[vicente@alcazaba.unex.es](mailto:vicente@alcazaba.unex.es)

# Análisis métrico de los blogs españoles de biblioteconomía y documentación (2006-2007)

Por Daniel Torres-Salinas, Álvaro Cabezas-Clavijo y Emilio Delgado-López-Cózar

**Resumen:** Se analizan diversos indicadores de producción y participación en los blogs españoles de biblioteconomía y documentación. A partir de una muestra de 46 blogs de autoría personal y corporativa, se toman los 12 con mayor número de enlaces recibidos. Son los que obtienen más del 70% de los enlaces recogidos por el conjunto de blogs documentales. Se detecta una bajada productiva significativa a lo largo del período de estudio, manteniéndose constante el número de comentarios por entrada. Más del 50% de los comentarios son realizados por los propios autores de bitácoras documentales. Destaca un conjunto de 4 blogs, que presentan indicadores por encima de la media en cuanto a producción y participación: *Deakialli*, *Catorze*, *Documentalista enredado* y *Véase además*.



**Daniel Torres Salinas** es investigador del Centro de Investigación Médica Aplicada de la Universidad de Navarra y miembro del grupo EC3 (Evaluación de la Ciencia y de la Comunicación Científica) de la Universidad de Granada, donde recientemente ha leído su tesis doctoral.



**Álvaro Cabezas Clavijo** es licenciado en documentación y miembro del grupo EC3 (Evaluación de la Ciencia y de la Comunicación Científica) de la Universidad de Granada. Su blog, *Documentación, biblioteconomía e información* es uno de los más visitados de la blogosfera documental hispana.



**Emilio Delgado López-Cózar** es profesor de metodología de la investigación en la Facultad de Biblioteconomía y Documentación y miembro del grupo EC3 (Evaluación de la Ciencia y de la Comunicación Científica) de la Universidad Granada.

**Palabras clave:** Blogs, Blogosfera, Blogmetría, Cibermetría, Biblioteconomía y documentación, España.

**Title: Metric analysis of Spanish library and information science blogs (2006-2007)**

**Abstract:** Several indicators of production and participation in Spanish library and information science blogs are analysed, based on a sample of 46 personal and corporate weblogs. The 12 blogs with the highest number of incoming links logged more than 70% of the total number of links received by the whole group. Our results detect an important decline in production during the study period, keeping constant the comments-per-post indicator. More than 50% of comments are written by the library blog authors and 4 blogs show higher than average indicators of production and participation: *Deakialli*, *Catorze*, *Documentalista enredado* and *Véase además*.

**Keywords:** Blogs, Blogosphere, Blogmetrics, Cybermetrics, Library and information science, Spain.

**Torres-Salinas, Daniel; Cabezas Clavijo, Álvaro; Delgado López-Cózar, Emilio.** "Análisis métrico de los blogs españoles de Biblioteconomía y Documentación (2006-2007)". En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, pp. 38-48.

DOI: 10.3145/epi.2008.ene.04

## Introducción

La popularización y el crecimiento de internet como medio de comunicación ha acabado derivando en los últimos años en la denominada web 2.0. Este nuevo ciberespacio se compone de un conjunto de herramientas complementarias entre sí (wikis, *Flickr*, *Youtube*, *Panoramio*, etc.) que facilitan la generación de contenidos en múltiples formatos por todo tipo de usuarios, y la interacción entre los mismos. Uno de los más

genuinos y reconocidos es la blogosfera, formada por los weblogs, blogs o bitácoras. Estos sitios se caracterizan por presentar una ordenación cronológica inversa de sus contenidos, poseer una constante actualización y contar con frecuentes enlaces. Otros de sus aspectos fundamentales es la oportunidad que tienen los lectores de dejar comentarios en las diferentes entradas, lo que da lugar a fructíferos debates entre el autor y los lectores, por lo que se suele decir que "los blogs son la conversación" (Rojas, et al., 2005).

Artículo recibido el 26-07-07

Aceptación definitiva: 23-10-07

En el ámbito bibliotecario y documental español el fenómeno *blogger* no ha pasado desapercibido y las bitácoras han estado presentes desde 2001 con la creación de *Trucos de Google*. El interés despertado en los círculos académicos ha quedado patente en diferentes acontecimientos acaecidos en los últimos meses, como el número especial que *El profesional de la información* (v. 16, n. 2, 2007) dedicó a dicho fenómeno o las recientes actividades paralelas que, con gran éxito de público, se llevaron a cabo en las últimas *Jornadas españolas de documentación* organizadas por *Fesabid*.

<http://www.fesabid.org/santiago2007/>

Buena prueba también es la variada y diversa bibliografía que se ha generado en torno al tema. A nivel general, distintos autores se han encargado de mostrar las utilidades de los weblogs para los documentalistas (Lledó, 2003; Sorli; Merlo, 2003; López; Martínez, 2005), su capacidad para la redifusión de contenidos mediante la sindicación (Franganillo; Catalán, 2005), caracterizado su terminología específica (Ferrada, 2006b) o señalado sus propias experiencias en la creación y gestión de las bitácoras (López; Penadés, 2007; Becerril, et al., 2007).

Específicamente también se ha estudiado la bibliosfera (bibliogsfera o biblioblogsfera) es decir los weblogs especializados en biblioteconomía y documentación (ByD). Así, nos encontramos trabajos que han evaluado la calidad formal de 30 blogs académicos y de investigación (Jiménez; Salvador, 2007), caracterizado las redes sociales existentes entre la comunidad de blogs y algunos de sus indicadores cibernéticos (Flores; Legerén, 2005; Ros; Mateo, 2005a), realizado una historia de los blogs documentales hispanos (Ros; Mateo, 2005b) o lanzado propuestas de clasificación en función de su contenido (Leiva, 2006; Sedic, 2006).

Sin embargo, hasta el momento, no se ha realizado ningún acercamiento profundo a la blogosfera documental nacional desde una perspectiva puramente métrica, tal y como ha ocurrido en el mundo anglosajón (Bar-Ilan, 2005; Crawford, 2005), que nos permita describir nuestra comunidad. La única aportación que contamos a nivel nacional en este sentido fue el post *¿Se han estancado los blogs documentales?* publicado en el blog *Documentación, biblioteconomía e información* el pasado mes de mayo (Cabezas, 2007).

Concretamente en esta entrada, a través del análisis de producción y participación en los dos primeros trimestres de 2006 y 2007, se evidenció un descenso en la producción y número de comentarios en seis blogs representativos del ámbito documental. Dicho post recibió más de 30 comentarios, en su mayoría de otros *bloggers* reconocidos, que valoraron positivamente el

estudio, sugirieron ideas y líneas de investigación futuras, así como valiosas explicaciones del decaimiento en la producción y participación.

Teniendo en cuenta estos antecedentes, así como el debate y la expectación generada, parece interesante ampliar este breve estudio. El objetivo fundamental que se persigue es determinar si realmente se está produciendo un descenso en la actividad y si esta tendencia, además, es común a otros blogs de la disciplina. Asimismo introducimos nuevos indicadores que nos ayuden a describir otros aspectos como la longitud de las entradas, la frecuencia de publicación, los autores más célebres, los posts más comentados o el número de entradas que no han generado participación. En definitiva ofrecemos un análisis métrico, en este caso *blogmétrico*, que nos ayude a desvelar qué ha ocurrido en los últimos meses en la blogosfera documental española.

## Material y métodos

Se realiza un estudio descriptivo longitudinal retrospectivo de 46 blogs personales y corporativos españoles especializados temáticamente en ByD, por lo que la unidad de análisis se circunscribe, por tanto, a la blogosfera documental española. Sin embargo no es fácil delimitar esta comunidad con claridad debido al constate surgimiento de nuevos blogs y la desaparición de otros. Nos encontramos asimismo las propias dificultades conceptuales que comporta la delimitación de un campo multidisciplinar como éste (Jiménez, et al., 2006). Según recientes estimaciones, existen unas 80 bitácoras españolas relativas a ByD, desde bibliotecas públicas, universitarias y especializadas a personales o de asociaciones e instituciones (Sedic, 2006).

Partiendo de este grupo, para el presente trabajo se han considerado y delimitado como unidad de análisis todos los que aparecen recopilados en el esquema clasificatorio propuesto por el equipo del *Sedic blog* (2006) bajo los encabezamientos “Blogs colectivos de asociaciones o instituciones”, “Blogs personales” y “Blogs individuales que aportan contenidos muy personales, de carácter heterogéneo” (un total de 46 sitios, como se dijo). Esta clasificación es una de las más recientes y exhaustivas, y una de las aproximaciones más coherentes a una categorización de los blogs de ByD. El marco cronológico corresponde a los meses comprendidos entre enero de 2006 y abril de 2007, abarcándose por tanto un total de 485 días.

Una vez establecida y delimitada la unidad de análisis se ha calculado para cada uno de los 46 blogs su “Autoridad” (*Technorati authority*) un indicador creado y proporcionado por el buscador especializado *Technorati*, que se entiende como el número de blogs citantes en los últimos seis meses (Carroll, 2007). Esta medida posee algunas limitaciones, como que no distingue los

enlaces enviados desde el *blogroll*; no contempla los links recibidos desde sitios web que no sean blogs; y depende de la fecha en la que se realizó la búsqueda. Sin embargo nos ofrece una aproximación válida del impacto de un blog dentro de su propia comunidad, y del reconocimiento que otros autores le otorgan. Una vez calculada la “Autoridad”, se han seleccionado los situados en el primer cuartil del ranking a fin de calcular para cada uno de ellos los indicadores de la tabla 1.

<http://technorati.com/>

Indicador	Abrev.
Número de posts o entradas	<b>P</b>
Número de posts por día	<b>PxD</b>
Número de palabras contenidas en los posts	<b>PR</b>
Número de palabras por post	<b>PRxD</b>
Número de comentarios recibidos	<b>C</b>
Número de comentarios por post	<b>CxP</b>
Número de posts sin comentarios	<b>PSC</b>

Tabla 1. Indicadores empleados en el análisis de los blogs españoles de ByD

Asimismo se ha realizado una caracterización específica de los comentarios clasificándolos en tres tipos diferentes:

- Auto: los realizados por un autor en su propio sitio.
- Internos: comentario recibido por autores de alguno de los 46 blogs de ByD pertenecientes a la unidad de análisis.
- Externos: cuando el comentario no proviene de ninguno de los dos anteriores.

## Resultados y discusión

### 1. Resultados generales

La mayor parte de los blogs presentan algunas características comunes que le otorgan al grupo cierta homogeneidad (tabla 2). En su mayor parte son personales (90%) por lo que no se adscriben a asociaciones o instituciones; suelen ser redactados por una persona encargada tanto de su gestión como del mantenimiento de los contenidos (76%); son escritos en castellano (91%) y su fecha de inicio suele situarse entre 2004-2006 (80%).

Centrando la atención en el indicador de “Autoridad” se constata como 12 blogs acumulan más del 70% de los enlaces. Los más enlazados son en su mayoría de tipo personal y gestionados por uno o dos autores.

Diez de ellos se redactan en castellano, mientras que los dos restantes usan el catalán como lengua principal. La mayor parte de estos 12 blogs comenzaron su andadura entre 2003 y 2004 y son los que seleccionamos finalmente para el cálculo de los indicadores propuesto en la tabla 2.

El número de posts producidos por los 12 blogs más enlazados ha sido de 1.970 lo que supone una media de 123 entradas al mes (tabla 3). Comparados el primer cuatrimestre de 2006 y 2007, existe una diferencia significativa ya que en los cuatros primeros meses de 2006 el número de posts fue de 610 mientras que en el mismo marco, para 2007, esta cifra sólo alcanza 418. Este descenso se refleja asimismo en el número de posts por día. El valor general para PxD es de 4, siendo los meses de enero y febrero de 2006 los únicos en los que se alcanzaron 5 o más diarios, estas cantidades ya no se vuelven a alcanzar y en el último mes, abril de 2007, se llegó al mínimo con 2,8 entradas diarias.

---

**“La mayor parte de los blogs son personales (90%) y su fecha de inicio suele situarse entre 2004-2006 (80%)”**

---

En cuanto a la longitud de los posts (número de palabras), el conjunto de blogs ha escrito casi medio millón de palabras. Este indicador muestra un carácter decreciente, como es lógico: a un menor número de posts le equivale una menor cantidad de texto. La longitud estándar de los posts se situó en 253 palabras, valor que se mantiene bastante constante a lo largo del período, lo que revela que aunque se escriban menos posts suelen ser casi siempre del mismo tamaño.

Otro de los aspectos interesantes son los comentarios, denotativos, en principio, de la capacidad de un autor para generar debate en la blogosfera así como de la actividad de sus lectores. El número de ellos en total fue de 5.729, con tendencia a disminuir, ya que de los más de 2.000 que se realizaron en los cuatro primeros meses de 2006 se ha pasado en 2007 a no superar los 1.500. Sin embargo, este descenso va unido de nuevo al número de posts redactados. Si tomamos el indicador CxP vemos que este valor se sitúa en 2,9 y a lo largo del período no se identifica ninguna tendencia. Otro de los aspectos destacados es el número de posts que no han recibido ningún comentario (PSC), que en total son 822 (42%). Este valor es bastante estable, y tan sólo en enero y febrero de 2006 se mostraron porcentajes muy por debajo de la media.

Rank	NOMBRE	URL	Auto- ridad	% Auto- ridad	% Acumu- lado	Tipolo- gía	Nº Auto- res	Lengua	Año de Inicio
1	<i>Catorze.blog</i>	<a href="http://www.catorze.com/blog">http://www.catorze.com/blog</a>	<b>164</b>	13,19	13,19	Personal	1	Castellano	2003
2	<i>Deakialli docu- Mental</i>	<a href="http://www.deakialli.com/">http://www.deakialli.com/</a>	<b>150</b>	12,07	25,26	Personal	2	Castellano	2003
3	<i>El documentalista enredado</i>	<a href="http://www.documentalistaenredado.net/">http://www.documentalistaenredado.net/</a>	<b>138</b>	11,10	36,36	Personal	2	Castellano	2004
4	<i>Úsalo. Usabilidad para todos</i>	<a href="http://usalo.es/">http://usalo.es/</a>	<b>95</b>	7,64	44,00	Personal	2	Castellano	2004
5	<i>Véase además</i>	<a href="http://www.veaseademas.com">http://www.veaseademas.com</a>	<b>84</b>	6,76	50,76	Personal	4	Castellano	2004 <sup>(=)</sup>
6	<i>Documentación, biblioteconomía e información</i>	<a href="http://www.lacoctelera.com/documentacion">http://www.lacoctelera.com/documentacion</a>	<b>48</b>	3,86	54,62	Personal	1	Castellano	2005
7	<i>Recuperación de información en la web</i>	<a href="http://irsweb.blogspot.com/">http://irsweb.blogspot.com/</a>	<b>48</b>	3,86	58,48	Personal	1	Castellano	2004
8	<i>Sedic blog</i>	<a href="http://blog.sedic.es">http://blog.sedic.es</a>	<b>35</b>	2,82	61,30	Asoc. E Inst.	6 <sup>(+)</sup>	Castellano	2006
9	<i>Gamoia. bloc d'una bibliotecària</i>	<a href="http://gamoia.bloc.cat/">http://gamoia.bloc.cat/</a>	<b>33</b>	2,65	63,95	Personal	1	Catalán	2005
10	<i>Bibliometría</i>	<a href="http://www.bibliometria.com/">http://www.bibliometria.com/</a>	<b>32</b>	2,57	66,53	Personal	1 <sup>(+)</sup>	Castellano	2005 <sup>(=)</sup>
11	<i>Bauen blog</i>	<a href="http://bauenblog.info/">http://bauenblog.info/</a>	<b>30*</b>	2,41	68,94	Personal	2	Catalán	2004
12	<i>Métodos de busca</i>	<a href="http://www.metodosdebusca.com">http://www.metodosdebusca.com</a>	<b>30</b>	2,41	71,36	Personal	1	Castellano	2001
13	<i>La imagen social del bibliotecario</i>	<a href="http://lisdb.blogspot.com/">http://lisdb.blogspot.com/</a>	<b>25</b>	2,01	73,37	Personal	1	Castellano	2006
14	<i>Alas y balas</i>	<a href="http://www.alasybalas.com/">http://www.alasybalas.com/</a>	<b>24</b>	1,93	75,30	Personal	1	Castellano	2003
15	<i>Biblioblog. bitàcora informativa sobre bibliotecas</i>	<a href="http://www.interreg-eet.info/weblogs/biblio-&lt;br/&gt;blog/index.html">http://www.interreg-eet.info/weblogs/biblio- blog/index.html</a>	<b>23</b>	1,85	77,15	Asoc. E Inst.	6	Castellano	2003
16	<i>Bibliotecosas: silva bibliotecaria de varia lección:</i>	<a href="http://bibliotecosas.blogia.com/">http://bibliotecosas.blogia.com/</a>	<b>23</b>	1,85	79,00	Personal	1	Castellano	2004
17	<i>Anécdota?</i>	<a href="http://www.anecdótica.net/">http://www.anecdótica.net/</a>	<b>21</b>	1,69	80,69	Personal	1	Castellano	2005
18	<i>Open Access</i>	<a href="http://weblogs.madrimasd.org/openaccess/">http://weblogs.madrimasd.org/openaccess/</a>	<b>19</b>	1,53	82,22	Personal	1	Castellano	2006
19	<i>@rchivista: una mirada al mundo de los archivos</i>	<a href="http://archivista.wordpress.com/">http://archivista.wordpress.com/</a>	<b>17</b>	1,37	83,58	Personal	1	Castellano	2006
20	<i>A propósito de</i>	<a href="http://www_apropositode.blogspot.com">http://www_apropositode.blogspot.com</a>	<b>17</b>	1,37	84,95	Personal	1	Castellano	2005
21	<i>TekaTeka</i>	<a href="http://tekateka.blogspot.com/">http://tekateka.blogspot.com/</a>	<b>16</b>	1,29	86,24	Personal	1	Castellano	2004
22	<i>Archivística.net</i>	<a href="http://www.archivística.blogspot.com">http://www.archivística.blogspot.com</a>	<b>15</b>	1,21	87,45	Personal	1	Castellano	2003
23	<i>Interneteando</i>	<a href="http://interneteando-lara.blogspot.com/index.&lt;br/&gt;html">http://interneteando-lara.blogspot.com/index. html</a>	<b>15</b>	1,21	88,65	Personal	1	Castellano	2006
24	<i>El documentalista audaz</i>	<a href="http://sirmijail.blogspot.com/">http://sirmijail.blogspot.com/</a>	<b>14</b>	1,13	89,78	Personal	1	Castellano	2005
25	<i>Geekteca</i>	<a href="http://geekteca.blogspot.com/">http://geekteca.blogspot.com/</a>	<b>13</b>	1,05	90,82	Personal	1	Castellano	2004
26	<i>biBLOGteca. bBlog de Asnabi</i>	<a href="http://asnabi.blogspot.com">http://asnabi.blogspot.com</a>	<b>12</b>	0,97	91,79	Asoc. E Inst.	Sin dato	Castellano	2006
27	<i>Outbook: diseño web, usabilidad y documentación</i>	<a href="http://outbook.blogspot.com/">http://outbook.blogspot.com/</a>	<b>11</b>	0,88	92,68	Personal	1	Castellano	2004
28	<i>Expediente, web, blog y otros documentos del montón</i>	<a href="http://documentalqueridowatson.blogspot.com">http://documentalqueridowatson.blogspot.com</a>	<b>10*</b>	0,80	93,48	Personal	1	Castellano	2005
29	<i>Servicio de alerta</i>	<a href="http://docdigital.typepad.com/">http://docdigital.typepad.com/</a>	<b>10</b>	0,80	94,28	Personal	3	Castellano	2004

Tabla 2. Características de los blogs españoles de ByD con mención del indicador de "Autoridad" elaborado por Technorati

- Búsqueda en Technorati realizada el 18 de mayo de 2007.
- En gris: blogs seleccionados para el cálculo de los indicadores propuestos en la tabla 1.
- \* Cálculo manual.
- \*\* Cambio url.
- \*\*\* No actualizado en 2007.
- (+) Colaboraciones frecuentes de otros autores.
- (=) En la blogosfera, con blogs de contenido muy similar desde 2003.

Rank	NOMBRE	URL	Autoridad	% Autoridad	% Acumulado	Tipología	Nº Autores	Lengua	Año de Inicio
30	Ranganatha2: atrévete a pensar (Think ahead)	<a href="http://www.espacioblog.com/ranganatha2">http://www.espacioblog.com/ranganatha2</a>	10	0,80	95,09	Personal	2	Castellano	2006
31	Tentándole	<a href="http://tentandole.blogspot.com/">http://tentandole.blogspot.com/</a>	10	0,80	95,89	Personal	1	Castellano	2006
32	Baiget info notas	<a href="http://baiget.blogspot.com/">http://baiget.blogspot.com/</a>	9	0,72	96,62	Personal	1	Castellano	2004
33	EC3noticias	<a href="http://ec3noticias.blogspot.com/">http://ec3noticias.blogspot.com/</a>	7	0,56	97,18	Asoc. E Inst.	1(+)	Castellano	2005
34	Bibliotecaescolar.net	<a href="http://antoniotejero.blogspot.com/">http://antoniotejero.blogspot.com/</a>	7	0,56	97,74	Personal	1	Castellano	2005
35	Mixobitácora	<a href="http://mixobitacora.blogspot.com/">http://mixobitacora.blogspot.com/</a>	5	0,40	98,15	Personal	1	Castellano	2005
36	Vida de una men-ganita	<a href="http://menganita.blogspot.com/">http://menganita.blogspot.com/</a>	5	0,40	98,55	Personal	1	Castellano	Sin dato
37	Bibliotecas digitales*	<a href="http://bibliotecas-digitales.blogspot.com/">http://bibliotecas-digitales.blogspot.com/</a>	5	0,40	98,95	Personal	1	Castellano	2006
38	Biblio's	<a href="http://libiblios.blogspot.com/">http://libiblios.blogspot.com/</a>	4	0,32	99,27	Personal	1	Catalán	2004
39	Documentamania	<a href="http://documentamania.blogspot.com/">http://documentamania.blogspot.com/</a>	3	0,24	99,51	Personal	1	Castellano	2006
40	Archiveros Españoles en la Función Pública	<a href="http://archiverosefp.blogspot.com/">http://archiverosefp.blogspot.com/</a>	2	0,16	99,67	Asoc. E Inst.	Sin dato	Castellano	2006
41	Bibliosalud	<a href="http://www.lacoctelera.com/bibliosalud">http://www.lacoctelera.com/bibliosalud</a>	2	0,16	99,84	Personal	1	Castellano	2006
42	Joana Albret bibliotekonómia mintegia	<a href="http://www.eibar.org/blogak/joana-albret">http://www.eibar.org/blogak/joana-albret</a>	1	0,08	99,92	Personal	1	Euskera	2007
43	La mirilla digital	<a href="http://mirilladigital.blogspot.com/">http://mirilladigital.blogspot.com/</a>	1	0,08	100,00	Personal	1	Castellano	2006
44	De todo un poko	<a href="http://dtnpk.blogspot.com">http://dtnpk.blogspot.com</a>	0	0,00	100,00	Personal	1	Castellano	2005
45	Alfin: avances de la alfabetización Informacional	<a href="http://alfin.blogspirit.com">http://alfin.blogspirit.com</a>	X**	0,00	100,00	Personal	1	Castellano	2005
46	El hombre irrazonable	<a href="http://victormacias.blogia.com/">http://victormacias.blogia.com/</a>	X***	0,00	100,00	Personal	1	Castellano	2006

Tabla 2 (continuación). Características de los blogs españoles de ByD con mención del indicador de "Autoridad" elaborado por Technorati

- Búsqueda en Technorati realizada el 18 de mayo de 2007.
- En gris: blogs seleccionados para el cálculo de los indicadores propuestos en la tabla 1.
- \* Cálculo manual.
- \*\* Cambio url.
- \*\*\* No actualizado en 2007.
- (+) Colaboraciones frecuentes de otros autores.
- (=) En la blogosfera, con blogs de contenido muy similar desde 2003.

Mes	P	PxD	PR	PRxP	C	CxP	PSC	%PSC
1. Enero	190	6,1	47.501	250	658	3,5	65	34
2. Febrero	141	5,0	37.887	269	698	5,0	35	25
3. Marzo	140	4,5	34.849	249	365	2,6	57	41
4. Abril	139	4,6	31.357	226	370	2,7	54	39
5. Mayo	126	4,1	28.758	228	235	1,9	63	50
6. Junio	117	3,9	34.321	293	330	2,8	53	45
7. Julio	96	3,1	24.858	259	261	2,7	42	44
8. Agosto	83	2,7	24.357	293	309	3,7	31	37
9. Septiembre	109	3,6	24.633	226	171	1,6	60	55
10. Octubre	136	4,4	42.937	316	302	2,2	55	40
11. Noviembre	143	4,8	37.551	263	349	2,4	62	43
12. Diciembre	132	4,3	28.484	216	252	1,9	52	39
13. Enero	121	3,9	29.371	243	316	2,6	57	47
14. Febrero	117	4,2	27.584	236	267	2,3	49	42
15. Marzo	96	3,1	21.854	228	599	6,2	51	53
16. Abril	84	2,8	22.034	262	247	2,9	36	43
<b>Total general</b>	<b>1.970</b>	<b>4,1</b>	<b>498.336</b>	<b>253</b>	<b>5.729</b>	<b>2,9</b>	<b>822</b>	<b>42</b>

Tabla 3. Indicadores generales para los blogs de ByD: posts, palabras y comentarios

En la tabla 4 se han identificado y caracterizado temáticamente los posts que han recibido 17 o más comentarios, destacando el *Criterio Hotmail*, de *Catorze*, con 337. Sin embargo esta tabla ha de leerse con ciertas reservas. Si tomamos como botón de muestra los tres casos más comentados, se observa la escasa relación que guardan con la documentación; dos de ellos son de tipo generalista (*Criterio Hotmail* y *¿Qué es la ciencia?*) y otro (*Marina d'Or apesta*) es una breve frase al margen de la documentación. Los usuarios de los blogs que han dejado los comentarios son personas alejadas, en su gran mayoría, de la disciplina y el listado completo dista mucho de un debate serio y fructífero.

No siempre, por tanto, la cantidad de comentarios es una medida de la influencia en el campo; normal-

**“La cantidad de comentarios no siempre es una medida de la influencia en el campo; normalmente los posts desenfadados sobre temas poco especializados son los que más reacciones desencadenan”**

mente los posts con temas poco especializados, generalistas y de carácter desenfadado son los que más reacciones desencadenan. Si atendemos a toda la tabla, observamos que tan sólo un tercio pueden ser considerados de temática profesional. Otro tanto corresponde a

Rank	Blog	Título del post	C	Año	Mes	Tema
1	<i>Catorze</i>	<i>El criterio Hotmail</i>	337	2007	Marzo	No ByD
2	<i>Véase</i>	<i>Marina d'Or apesta</i>	88	2006	Febrero	No ByD
3	<i>Bibliometria</i>	<i>¿Qué es ciencia?</i>	85	2006	Agosto	Profesional
4	<i>Deakialli</i>	<i>4 velas para Deakialli, cuatro años blogueando</i>	38	2007	Marzo	Personal
5	<i>Catorze</i>	<i>Google Co-op y los blogs sobre ByD</i>	30	2006	Octubre	Blogs
6	<i>Catorze</i>	<i>Hotmail</i>	30	2007	Abril	No ByD
7	<i>Deakialli</i>	<i>Tres años de Deakialli</i>	28	2006	Marzo	Personal
8	<i>Deakialli</i>	<i>Se pide opinión. La utilidad de los blogs</i>	27	2006	Julio	Blogs
9	<i>Bibliometria</i>	<i>Una cronología del fraude científico</i>	26	2006	Febrero	Profesional
10	<i>Catorze</i>	<i>Repositorio de comentarios</i>	25	2007	Marzo	No ByD
11	<i>Deakialli</i>	<i>Deakialli.com</i>	24	2006	Noviembre	Personal
12	<i>Véase</i>	<i>Tips para que tu trabajo no interfiera en tu bitácora</i>	24	2006	Febrero	Blogs
13	<i>Véase</i>	<i>Wikis como organizadores personales</i>	24	2006	Febrero	Profesional
14	<i>Catorze</i>	<i>Pasado, presente, futuro</i>	24	2006	Agosto	Personal
15	<i>Úsalo</i>	<i>Fraudes medallistas</i>	23	2006	Abril	Profesional
16	<i>Documentalista Enredado</i>	<i>Carpe diem blogger, carpe diem</i>	20	2006	Enero	Blogs
17	<i>Deakialli</i>	<i>Deakialli en Localía TV</i>	20	2006	Junio	Personal
18	<i>Documentacion</i>	<i>Historias de terror (II)</i>	20	2006	Junio	Profesional
19	<i>Véase</i>	<i>Exhaustividad y precisión en SRIs basados en tagging</i>	20	2006	Julio	Profesional
20	<i>Catorze</i>	<i>ISSN para weblogs: intento fallido</i>	19	2006	Enero	Blogs
21	<i>Documentacion</i>	<i>Fallo de seguridad en El país?</i>	18	2006	Enero	Profesional
22	<i>Úsalo</i>	<i>Usabilidad para principiantes</i>	18	2006	Enero	Profesional
23	<i>Deakialli</i>	<i>Los estudios universitarios de biblioteconomía y documentación existen ¿lo sabías?</i>	18	2007	Marzo	Profesional
24	<i>Deakialli</i>	<i>Blog da asesoría de bibliotecas escolares</i>	18	2007	Abril	Blogs

Tabla 4. Ranking de posts con mayor número de comentarios en los blogs de ByD: número de comentarios y tema

cuestiones sin relación alguna con la documentación o relativas a celebraciones y anuncios que el autor realiza sobre su bitácora. El tercio restante se relaciona directamente con diversos aspectos del universo *blogger*.

En la figura 1 se presenta una caracterización de los 5.729 comentarios recopilados según el origen del comentarista. Predominan con un 43% los realizados por lectores ajenos a los blogs de ByD, mientras que un 27% fueron internos. El 30% de autocomentarios indica el interés que los autores se toman no sólo en escribir sus entradas sino también en responder a los lectores. Si tomamos conjuntamente los autocomentarios y los internos se obtiene un 57%, cifra bastante elevada.

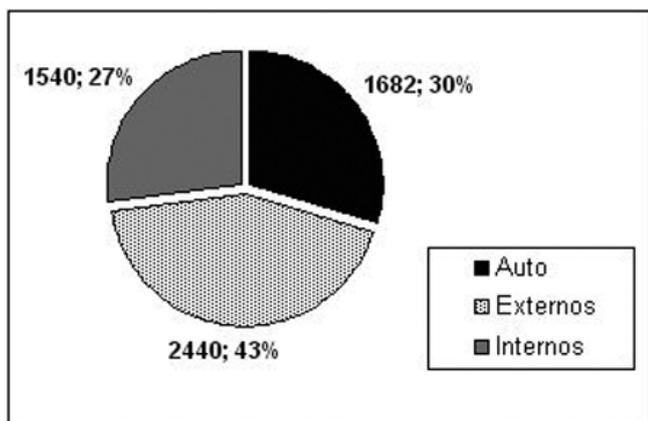


Figura 1. Porcentaje y número de comentarios según su origen (auto, externos e internos) para los blogs de ByD

## 2. Resultados por blog

En la tabla 5 se describen más detalladamente los doce blogs más enlazados presentándose los autores, sus profesiones y otros medios de comunicación empleados por ellos en la difusión de sus ideas. Como se observa, prácticamente todos los *blogger*s están vinculados a alguna profesión relacionada directamente con la ByD y más de la mitad pertenecen a *Exit* (*Directorio de expertos en el tratamiento de la información*). La mayor parte de ellos, además, ha empleado en alguna ocasión otros medios de comunicación al margen de su bitácora, como es el caso las revistas, preferentemente científicas y nacionales, o las comunicaciones a congresos.

<http://www.directorioexit.info/>

En la tabla 6 se muestran los indicadores para cada uno de los blogs. Si atendemos a la actividad produc-

tiva, el más destacado ha sido *Deakialli* con 342 entradas seguido de *Véase además* con 260. Ambos son redactados por más de una persona. Los menos actualizados durante el período han sido *Métodos de busca* y *Úsalo*. Ninguno supera durante el período el post por día publicado, siendo la media de publicación de una entrada cada tres días (0,34). Entre los que publican las más extensas, sin entrar en su calidad, se encuentra el *Documentalista enredado* con una extensión de 602 palabras de media. El blog con una ratio más baja ha sido *Véase además* con 98. La longitud media de los posts se mueve en un arco bastante amplio por lo que no existe homogeneidad para este indicador.

En cuanto a los comentarios, el más destacado, para el indicador *C*, ha sido *Catorze* con 1.133, seguido de *Deakialli* con 1.107. Además, tal y como se reflejó en la tabla 4 son los que tienen mayor capacidad para generar posts con un gran número de comentarios. Respecto a la cantidad de comentarios por post, sólo dos blogs alcanzan o superan los 5 CxP: *Catorze* y *Documentación*, *biblioteconomía e información*. La zona baja vuelve a ocuparla *Métodos de busca*, que prácticamente no ha recibido ninguno. Igualmente hay que señalar la importante brecha existente entre los cuatro blogs que menos comentarios reciben (*Métodos de busca*, *Bauen*, *RI en la web* y *Sedic*) y los ocho restantes, que alcanzan cifras de participación muy estimables. Estas diferencias también quedan reflejadas para el indicador número de posts sin comentarios, pues sólo tres blogs consiguen que menos del 20% de sus entradas se queden sin comentarios (*Úsalo*; *Documentación*, *Biblioteconomía e información*; y el *Documentalista enredado*).

En la figura 2 resumimos cómo se sitúan los blogs en relación con la media general de dos indicadores importantes: Número de posts y Número de comentarios por post. Son cuatro blogs los que dominan la escena, correspondientes a los que ocupan el margen superior

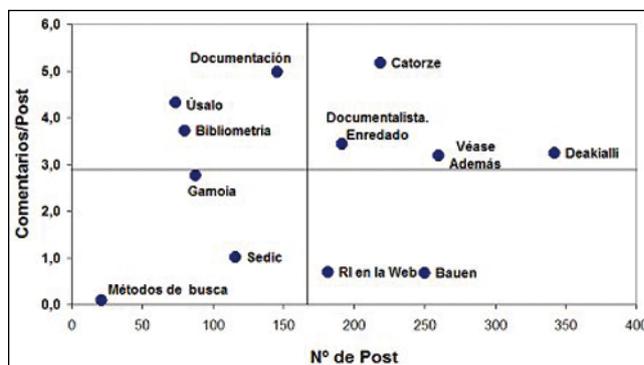


Figura 2. Posición de los blogs de ByD respecto a la media general de los indicadores "Número de posts" (P) y "Número de comentarios por post" (CxP)

\*Línea horizontal paralela a abscisas: media general de Comentarios por post = 2,9

\*Línea vertical paralela a ordenadas: media general de Número de posts = 164

**"Todos los bloggers están vinculados a la ByD y más de la mitad pertenecen a Exit"**

Autor	Blog	Profesión*	Ficha Exit	Libros	Revis-tas	Otros blogs	Otros me-dios	Congre-sos	Cur-sos
<b>Aina Gionés Valls</b>	Bauen	Estudiante	-	-	-	-	-	-	-
<b>Daniel Gil Solés</b>	Bauen	Bibliotecario	X	X	X	X	-	X	
<b>Álvaro Roldan</b>	Bibliometria	Administración pública	X	-	X	X	-	-	-
<b>Javier Leiva Aguilera</b>	Catorze/ Véase además	Profesor universitario	X	-	X	X	X	-	-
<b>Catuxa Seoane García</b>	Deakialli	Bibliotecaria	X	-	-	X	-	X	X
<b>Vanesa Barrero Robledo</b>	Deakialli	Documentalista	-	-	-	-	-	-	-
<b>Álvaro Cabezas Clavijo</b>	Doc., bib. e inf.	Becario investigación	X	-	X	X	-	-	-
<b>Marcos Ros Martín</b>	Doc. enredado	Documentalista	-	-	X	X	X	X	X
<b>Maria Elena Mateo</b>	Doc. enredado	Documentalista	-	-	X	-	-	-	-
<b>Núria Gallart</b>	Gamoia	Bibliotecaria	X	X	X	-	-	X	X
<b>Ricardo Fornas Carrasco</b>	Métodos de busca	Consultor	-	-	X	-	X	X	
<b>Fº Javier Martínez Méndez</b>	RI en la web	Profesor universitario	-	-	X	-	-	X	-
<b>Luis Rodríguez Yunta</b>	<i>Sedic</i>	Documentalista	X	X	X	-	X	X	-
<b>Elisa Prieto Castro</b>	<i>Sedic</i>	Bibliotecaria	X	-	X	-	X	-	-
<b>Nuria Escudero Galán</b>	<i>Sedic</i>	Espec. en informac.	-	-	X	-	-	X	-
<b>María Jesús Butera Fajardo</b>	<i>Sedic</i>	Documentalista	X	-	X	X	-	X	-
<b>Virginia Becerril González</b>	<i>Sedic</i>	Documentalista	X	-	X	-	-	-	-
<b>Julio Igualador Osoro</b>	<i>Sedic</i>	Documentalista	-	-	X	-	-	-	-
<b>Juan Carlos García Gómez</b>	<i>Usalo</i>	Documentalista	X	X	X	-	X	X	X
<b>Tomás Saorín Pérez</b>	<i>Usalo</i>	Administración pública	X	X	X	X	-	X	-
<b>Yusef Hassan Montero</b>	Véase además	Investigador	X	-	X	-	X	X	X
<b>Paola Andrea Ramírez</b>	Véase además	Profesora universitaria	-	-	-	X	X	-	-
<b>Fernando Siles</b>	Véase además	Documentalista	-	-	-	-	-	-	-

Tabla 5. Autores de los blogs de ByD: profesión y medios de comunicación empleados

derecho. Todos ellos, dedicados a diversos ámbitos de la ByD, son los que más producen y los que generan un mayor debate en la bibliogósfera y pueden ser considerados como el núcleo de la disciplina.

En el cuadrante superior izquierdo se localizan los blogs que publican pocos posts pero que reciben un número significativo de comentarios. Encontramos los muy especializados que tratan temáticas documentales muy específicas, como la bibliometría y la usabilidad. En el cuadrante inferior derecho están dos con gran número de

posts publicados pero con poca capacidad para atraer comentarios. Por último, en el cuadrante inferior izquierdo encontramos los peor posicionados, aquellos que producen poco y además reciben pocos comentarios.

---

**“Dentro del grupo de 12 blogs más relevantes hay una tendencia clara hacia la caída en la redacción de posts”**

---

Blog	P	PxD	Pr	PrxP	C	CxP	PSC	%PSC
<i>Bauen</i>	250	0,5	71.898	288	170	0,7	181	72
<i>Bibliometría</i>	80	0,2	20.722	259	297	3,7	29	36
<i>Catorze</i>	219	0,5	24.540	112	1.133	5,2	66	30
<i>Deakialli</i>	342	0,7	59.208	173	1.107	3,2	126	37
<i>Documentacion, bibl. e inf.</i>	146	0,3	32.574	223	726	5,0	26	18
<i>Documentalista enredado</i>	192	0,4	115.563	602	659	3,4	38	20
<i>Gamoia</i>	88	0,2	23.959	272	243	2,8	34	39
<i>Métodos de busca</i>	21	0,0	2.532	121	2	0,1	20	95
<i>RI en la Web</i>	182	0,4	58.360	321	127	0,7	126	69
<i>Sedic</i>	116	0,2	27.034	233	118	1,0	71	61
<i>Úsalo</i>	74	0,2	36.581	494	320	4,3	8	11
<i>Véase además</i>	260	0,5	25.365	98	827	3,2	97	37

Tabla 6. Indicadores generales para los blogs de ByD: posts, palabras y comentarios

## Discusión y conclusiones

A raíz de los resultados y retomando como herramienta de análisis los comentarios efectuados en el post *¿Se han estancado los blogs documentales?* (véase cuadro de texto adjunto) se puede concluir básicamente lo siguiente:

Existe un grupo de 12 blogs que son los más relevantes de la bibliogósfera nacional si consideramos el indicador “Autoridad”. Dentro de este grupo hay una tendencia clara hacia la caída en la redacción de posts. Este descenso ha provocado también que decrezca el número bruto de comentarios por lo que los debates y la actividad de los lectores se ha visto reducida al no tener contenidos para ello, la ratio comentarios/post no indica en ningún momento que se produzca una menor participación. Igual parece ocurrir con la longitud, aunque se publique menos, las entradas mantienen una extensión similar a lo largo del período. Por tanto se constata una menor actividad en la generación de contenidos pero no en la participación.

– Esta tendencia decreciente de la producción es síntoma evidente de un desgaste y de la adquisición de una posición acomodada dentro de la comunidad profesional de blogs documentales. La creciente especialización de algunos de ellos, la falta de tiempo por tenerse que dedicar a un puesto de trabajo (al que muchas veces se ha accedido gracias a la excelente plataforma profesional que significa mantener un blog de calidad) o la búsqueda de canales más apegados a la comunicación científica tradicional (artículos, congresos, etc.) son algunas de las causas aducidas por los propios *bloggers* que, en general, reconocen la bajada de la actividad.

– El análisis de los comentarios ha ilustrado diversos aspectos sobre la participación. Nos encontra-

mos ante una comunidad bastante reducida y cerrada compuesta activamente por unas decenas de personas. Pese a ser la blogósfera un espacio abierto a cualquier lector, sólo se consigue un 43% de comentarios externos, es decir de personas no vinculadas a ningún blog documental. Esta situación es característica de un grupo endogámico con fuertes lazos internos que se reparte la mayor parte de los comentarios entre ellos mismos. Las cifras globales de participación y el perfil de los participantes en cierta medida revelan el escaso interés y el desconocimiento que la comunidad de ByD le presta a los blogs como medio de comunicación para el intercambio de experiencias profesionales y científicas. Este fenómeno queda reflejado también en el número de posts sin comentarios que se situó en el 42% del total.

---

### “Es una comunidad bastante reducida y cerrada compuesta activamente por unas decenas de personas”

---

– Los comentarios tienden a concentrarse en posts poco relevantes que temáticamente guardan una escasa o nula relación con la ByD. En este sentido no se puede afirmar que el número de comentarios se corresponda con la calidad de los contenidos o la influencia de los autores en la disciplina; contrariamente: los temas populares consiguen atraer más comentaristas que poco pueden aportar. Se sitúan también entre los posts más comentados aquellos directamente relacionados con los blogs, un tema recurrente sobre el que, evidentemente,

## Comentarios destacados realizados sobre el post *¿Se han estancado los blogs documentales?* publicado en el blog *Documentación, biblioteconomía e información*.

- “Me parece muy interesante este mini estudio y simboliza perfectamente una cosa: desgaste” (**Marcos Ros**, *El documentalista enredado*).
  - “Creo que la reducción en el número de comentarios obedece a dos razones. Por una parte, el aumento del uso de herramientas de sindicación de contenidos hace que muchos lectores lean los posts sin acceder al blog y que, por tanto, no lleguen a leer los comentarios (a los que rara vez se suscribe alguien). Por otra parte, a mayor especialización de los blogs menor número de comentarios” (**Álvaro Roldán**, *Bibliometria.com*).
  - “Estaría bien hacer un análisis de quién comenta, dónde y con qué frecuencia. Si los debates en una comunidad de blogs se nutren de comentarios procedentes fundamentalmente de miembros de esa misma comunidad, y la actividad de ésta decae en lo que a producción de contenidos se refiere, parece lógico que también se reduzcan sus comentarios. Al fin y al cabo, la actividad bloguera no se limita (generalmente) a escribir en el blog propio, sino que va todo unido (leer-comentar-escribir)”. (**Paco Fernández**, *Archivista*).
  - “La mayoría de autores citados han/habéis “crecido” profesionalmente lo cual os ha abierto las puertas a otras formas de comunicación. Lo mismo ha sucedido con determinados “bloggers” americanos, que han dejado de publicar posts sustanciosos en la misma medida que han aumentado sus artículos en revistas, libros, cursillos, etc.”. (**Dídac Margaix**, *Dospuntocero*).
  - “Echo en falta sobre todo los blogs de investigadores y docentes, que sin duda alguna podrían aprovechar los blogs como medio de comunicación; sin embargo, tengo la sensación de que siguen demasiado enfrascados en los canales tradicionales de comunicación: jornadas, congresos y artículos en publicaciones especializadas”. (**María-José Butera**, *Sedic-blog*).
  - “Ahora hay miles de novedades cada día, por suerte la mayoría hemos evolucionado mucho profesionalmente y estamos muy ocupados (eso aborta muchos posts), el formato ha madurado y nosotros con él, hemos ido cambiando de gustos y seguramente muchos visitamos blogs de temas que antes no visitábamos, hay más gente posteando (aunque pudiera haber aún más, claro), etc. No sé, yo no creo en absoluto que los blogs sobre este sector gocen de mala salud”. (**Javier Leiva**, *Catorze*).
  - “Si se hiciera una clasificación de blogs de temática bibliodocumental debería haber numerosos huecos vacíos o al menos blogs duplicados y no los hay. Faltan temas por cubrir, yo creo. Hemos crecido muy poco en número”. (**Vanesa Barrero**, *Deakialli*).
- <http://www.lacoctelera.com/documentacion/post/2007/05/08/ase-han-estancado-blogs-documentales->

todo *blogger* tiene algo que decir. Esta concentración temática es síntoma de la escasa cobertura de la ByD en general. La discusión de la blogosfera poco tiene que ver con los frentes de investigación en ByD o con las materias clásicas de la diplomatura y la licenciatura.

– No hay un perfil que defina claramente al conjunto de *bloggers*; todos ellos se dedican a diferentes ramas profesionales de la ByD, pero no existen unas características comunes dominantes; el rango profesional abarca desde estudiantes hasta profesores pasando por investigadores y personal de la administración. Estos autores aunque emplean los blogs como medio de

---

**“La discusión de la blogosfera tiene poco que ver con los frentes de investigación en ByD o con las materias clásicas de la diplomatura y la licenciatura”**

---

comunicación preferente no se limitan a los mismos. Los medios tradicionales también son utilizados, sobre todo las revistas científicas y las comunicaciones a congresos. La difusión en estos canales suele ser de

ámbito nacional e incluso regional. En ocasiones colaboran además en otros blogs y medios digitales y algunos transmiten sus experiencias profesionales a través de cursos.

– De los 12 weblogs estudiados resalta un grupo de cuatro que pueden ser considerados los más importantes, en cuanto a producción y comentarios por post. Estos son *Deakialli*, *Véase además*, *Catorze* y *Documentalista enredado*. Son, asimismo, los que más enlaces reciben, y sus autores poseen una experiencia acumulada de entre 3-4 años en la redacción de sus blogs.

## Bibliografía

- Bar-Ilan, J.** "Information hub blogs". En: *Journal of information science*, 2005, v. 31, n. 4, pp. 297-307.
- Becerril, V.; Butera, M. J.; Escudero, N.; Igualador, J.; Prieto, E.; Rodríguez-Yunta, L.** "Experiencia con el blog colectivo de Sedic". En: *El profesional de la información*, 2007, v. 16, n. 2, pp. 134-137.
- Cabezas, Á.** *¿Se han estancado los blogs documentales?*. En: *Documentación, biblioteconomía e información*. Consultado en: 10-07-07. <http://www.lacoctelera.com/documentacion/post/2007/05/08/ase-han-estancado-blogs-documentales>
- Carroll, D.** *Technorati authority and rank*. Consultado en: 08-07-07. <http://technorati.com/weblog/2007/05/354.html>
- Crawford, W.** "Investigating the biblioblogosphere". En: *Cites & insights: crawford at large*, 2005, v. 5, n. 10, pp. 2-13.
- Ferrada, M.** *Biblogsfera, comunidad de bitácoras cuya temática versa total o parcialmente sobre biblioteconomía y documentación*. Consultado en: 08-07-07. [http://eprints.rclis.org/archive/00007179/01/ferrada\\_biblogsfera.pdf](http://eprints.rclis.org/archive/00007179/01/ferrada_biblogsfera.pdf)
- Flores, B.; Legerén E.** "El fenómeno weblog como nuevo medio de comunicación: su incidencia en el campo de la biblioteconomía y la documentación". En: *I Encuentro de estudiantes para la organización del conocimiento, 7º congreso ISKO-España*, 2005.
- Franganillo, J.; Catalán, M. A.** "Bitácoras y sindicación de contenidos: dos herramientas para difundir información". En: *BiD: textos universitarios de biblioteconomía i documentació*, 2005, diciembre, n. 15.
- Jiménez-Contreras, E.; Delgado-López-Cózar, E.; Ruiz-Pérez, R.** "Producción española en biblioteconomía y documentación con visibilidad internacional a través del Web of Science (1995-2004)". En: *El profesional de la información*, 2006, v. 15, n. 5, pp. 373-383.
- Jiménez-Hidalgo, S.; Salvador-Bruna, J.** "Evaluación formal de blogs con contenidos académicos y de investigación en el área de documentación". En: *El profesional de la información*, 2007, v. 16, n. 2, pp. 114-22.
- Leiva, J.** "Blogs: una herramienta de difusión para profesionales de la información". En: *TK*, 2006, n. 18.
- Lledó, M.** *El profesional de la información ante los weblogs*. Consultado en: 08-07-07. [http://eprints.rclis.org/archive/00000498/01/mlledo\\_weblogs.pdf](http://eprints.rclis.org/archive/00000498/01/mlledo_weblogs.pdf)
- López-Carreño, R.; Martínez-Méndez, F. J.** "Revisión de los servicios de internet: su aprovechamiento en unidades de información y documentación". En: *Revista española de documentación científica*, 2005, v. 28, n. 4, pp. 585-609.
- López, F.; Penadés, H.** "365 días de libros: blog para la promoción de la lectura". En: *El profesional de la información*, 2007, v. 16, n. 2, pp. 131-133.
- Merlo-Vega, J. A.; Sorli-Rojo, Á.** "Weblogs: un recurso para los profesionales de la información". En: *Revista española de documentación científica*, 2003, v. 26, n. 2, pp. 227-236.
- Rojas, O.; Alonso, J.; Antúnez, J. L.; Orihuela, J. L.; Varela, J.** *Blogs. La conversación en internet que está revolucionando medios, empresas y ciudadanos*. Madrid: ESIC, 2005. Isbn 84-7356-427-8.
- Ros, M.; Mateo, M. E.** *Análisis de la biblioblogosfera hispana (4). Estudio cibernético: quién y cómo*. En: *El documentalista enredado*. Consultado en: 10-07-07. <http://www.documentalistaenredado.net/?p=285>
- Ros, M.; Mateo, M. E.** *Análisis de la biblioblogosfera hispana (7), cronología*. En: *El documentalista enredado*. Consultado en: 10-07-07. <http://www.documentalistaenredado.net/?p=288>
- Sedic blog equipo de redacción.* "Blogs de biblioteconomía y documentación en España". En: *Revista española de documentación científica*, 2006, v. 29, n. 4, pp. 603-27.

**Daniel Torres-Salinas**, Centro de Investigación Médica Aplicada. Universidad de Navarra  
[torressalinas@gmail.com](mailto:torressalinas@gmail.com)

**Álvaro Cabezas-Clavijo**, Departamento de Biblioteconomía y Documentación. Universidad de Granada.  
[acabezasclavijo@gmail.com](mailto:acabezasclavijo@gmail.com)

**Emilio Delgado-López-Cózar**, Departamento de Biblioteconomía y Documentación. Universidad de Granada.  
[edelgado@ugr.es](mailto:edelgado@ugr.es)

## Suscripción EPI sólo online

Pensando sobre todo en los posibles suscriptores latinoamericanos, ya no es obligatorio pagar la suscripción impresa de EPI para acceder a la online.

EPI se ofrece a instituciones en suscripción "sólo online" a un precio considerablemente más reducido (85 euros/año), puesto que en esta modalidad no hay que cubrir los gastos de imprenta ni de correo postal.

## Software libre en repositorios de e-información

Por Ricard de-la-Vega-Sivera

**Resumen:** Las bibliotecas universitarias no sólo gestionan aquellos recursos electrónicos generados por editoriales comerciales, sino que también, y cada vez más, documentos fruto de actividades docentes e investigadoras de sus propias instituciones. Es función de las bibliotecas difundir estos materiales en forma electrónica, pasando previamente por las fases de almacenamiento y organización, así como garantizar su preservación. La forma de hacerlo es mediante los repositorios de e-información. En este artículo se pretende dar una rápida visión de los diferentes programas de software libre para el desarrollo de repositorios, centrándose en aquellos que se han usado para la elaboración de depósitos cooperativos catalanes. También se repasan algunas características deseables de estos sistemas, como son el uso de estándares abiertos de comunicación entre repositorios, las estrategias de preservación de los contenidos digitales y las comunidades de usuarios.

**Palabras clave:** Repositorios, E-ciencia, Acceso abierto, Software libre, Investigación.

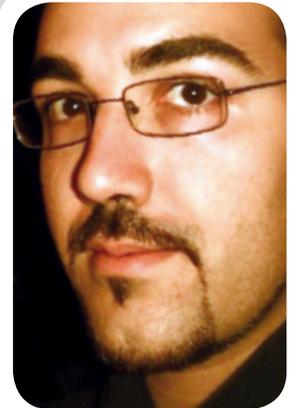
**Title:** Free software in e-information repositories

**Abstract:** University libraries manage not only commercial electronic resources, but also documents generated by research and educational activities at their own institutions. Libraries are committed to disseminating these electronic documents, which usually requires that they compile and store them as well as guaranteeing their preservation. The best way to achieve this objective is e-information repositories. We briefly present an overview of the free software available to develop repositories, with a special focus on the software used to build Catalan cooperative repositories. We also review desirable system characteristics, such as the use of open communication standards between repositories, preservation strategies for digital content, and user communities.

**Keywords:** Repositories, E-science, Open access, Free software, Research.

**De-la-Vega-Sivera, Ricard.** "Software libre en repositorios de e-información". En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, pp. 49-55.

DOI: 10.3145/epi.2008.ene.05



**Ricard de la Vega** es ingeniero técnico en informática de gestión por la Universidad Politécnica de Catalunya y ha realizado un postgrado de tecnologías de la comunicación. Ha sido formador de títulos oficiales de empresas del sector informático y desarrollador de proyectos web antes de entrar a formar parte del Departamento de Sistemas del Centro de Supercomputación de Catalunya, donde ahora coordina el área de repositorios y portales universitarios.

### Introducción

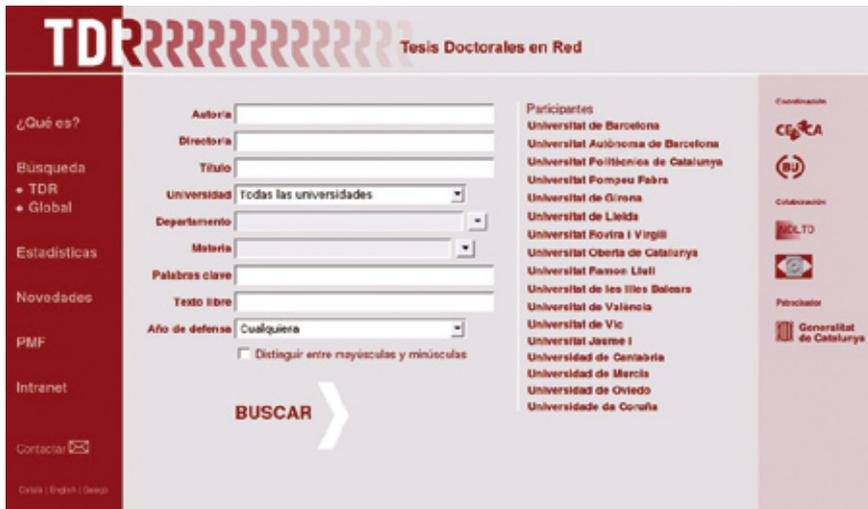
**LAS BIBLIOTECAS SON ORGANIZACIONES QUE ORDENAN y almacenan colecciones de documentos, tanto impresos como electrónicos, para su preservación y consulta. Concretamente, las bibliotecas universitarias no sólo gestionan aquellas fuentes de información electrónicas generadas por editoriales comerciales, sino que también, y cada vez más, documentos fruto de actividades docentes e investigadoras de sus propias instituciones (Anglada et al, 2005). Todo aquello que hace unos años se generaba sobre pa-**

**pel, hoy en día se realiza de manera digital. Es función de las bibliotecas difundir estos documentos en forma electrónica, pasando previamente por las fases de almacenamiento y organización, así como garantizar su preservación. El procedimiento para hacerlo es mediante los repositorios de e-información.**

Estos repositorios son una parte importante del movimiento del acceso abierto (*Open access*), que desde hace unos años apuesta por una nueva manera de difusión del conocimiento. Las contribuciones del acceso abierto han de cumplir,

según la *Declaración de Berlín*, dos condiciones. Por un lado, los autores y los que retienen el derecho sobre las colaboraciones han de garantizar a todos los usuarios el derecho al acceso abierto con licencia para copiar, usar, difundir, transmitir y exponer los trabajos públicamente con cualquier propósito responsable, a cambio de mencionar como es debido la autoría. Por otro lado, una versión completa de esta obra se tendrá que depositar en formato electrónico en un sitio en línea, esto es, en un repositorio.

En este sentido, el *Centre de Supercomputació de Catalunya*



Tesis Doctorales en Red (TDR)

<http://www.recercat.net>

<http://www.raco.cat>

En 2006 la *Biblioteca de Catalunya* inició otro ambicioso repositorio en colaboración con el *Cesca: Patrimoni Digital de Catalunya (Padicat)* (Llucca, 2006). Éste se estableció para garantizar el acceso permanente a la producción de webs catalanas. De esta manera, se implementa el tradicional ciclo de los documentos (compilación, gestión, preservación y difusión) para el material bibliográfico publicado en internet.

<http://www.padi.cat>

(*Cesca*) y el *Consorci de Biblioteques Universitàries de Catalunya (CBUC)* empezaron en 1999 un repositorio cooperativo para almacenar en formato digital el texto completo de tesis doctorales leídas en universidades de Cataluña y de otras comunidades autónomas para difundirlas mundialmente, en acceso abierto, preservando los derechos de propiedad intelectual de sus autores. Este repositorio, llamado *Tesis Doctorales en Red (TDR)* comenzó a funcionar hace 7 años y hoy en día está plenamente consolidado.

<http://www.tesisenred.net>

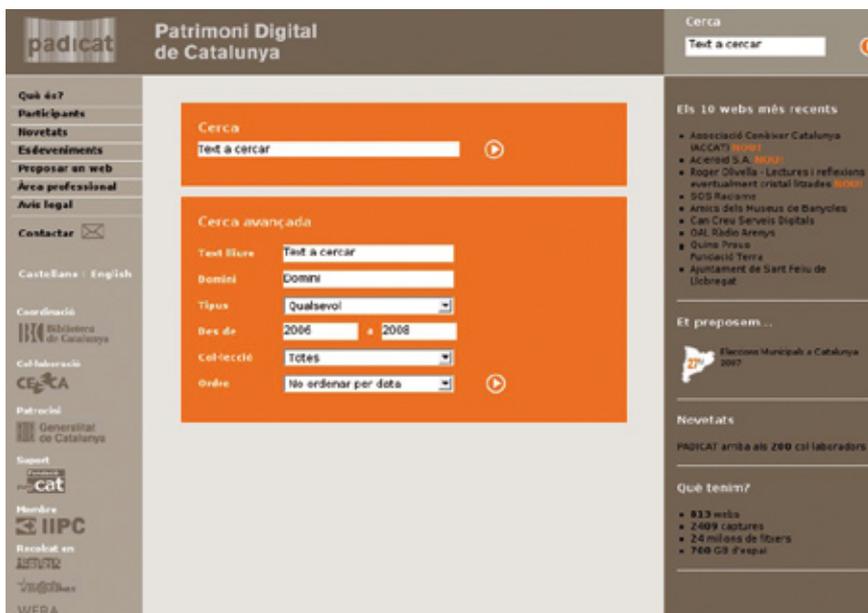
Los resultados positivos de esta iniciativa hicieron posible que la *Generalitat de Catalunya* patrocinase dos nuevos repositorios en acceso abierto: *Dipòsit de la Recerca de Catalunya (Recercat)* y *Revistes Catalanes amb Accés Obert (RACO)*. El primero es un depósito cooperativo de literatura de investigación de las universidades y centros de investigación de Cataluña: artículos aún no publicados, actas, informes de investigación, etc. El segundo permite consultar en acceso abierto los artículos completos de revistas científicas, culturales y eruditas catalanas.

Los cuatro repositorios tienen algunas características comunes: son de acceso abierto; están desarrollados con software libre; cumplen con el protocolo de interoperabilidad de la *Open Archives Initiative (OAI)* para compartir metadatos; y están construidos colectivamente, pudiendo adoptar procedimientos comunes, compartir costes de desarrollo y gestión, así como permitir una mayor visibilidad de los contenidos.

En este trabajo se pretende dar una rápida visión de algunos programas de software libre existentes para el desarrollo de repositorios, centrándose en los que se han usado para la elaboración de los mencionados depósitos cooperativos (Huguet et al, 2007). También se repasan algunas características deseables de estos sistemas, como son el uso de estándares abiertos de comunicación entre ellos, estrategias de preservación de los contenidos digitales o las comunidades de usuarios.

### Software libre para repositorios

La *Universidad de Nottingham* mantiene un listado (dotado con un potente buscador) de repositorios de contenidos en acceso abierto, dentro del proyecto *OpenDOAR*. Aquí puede observarse que la tendencia es desarrollarlos con aplicaciones



Patrimoni Digital de Catalunya (Padicat)

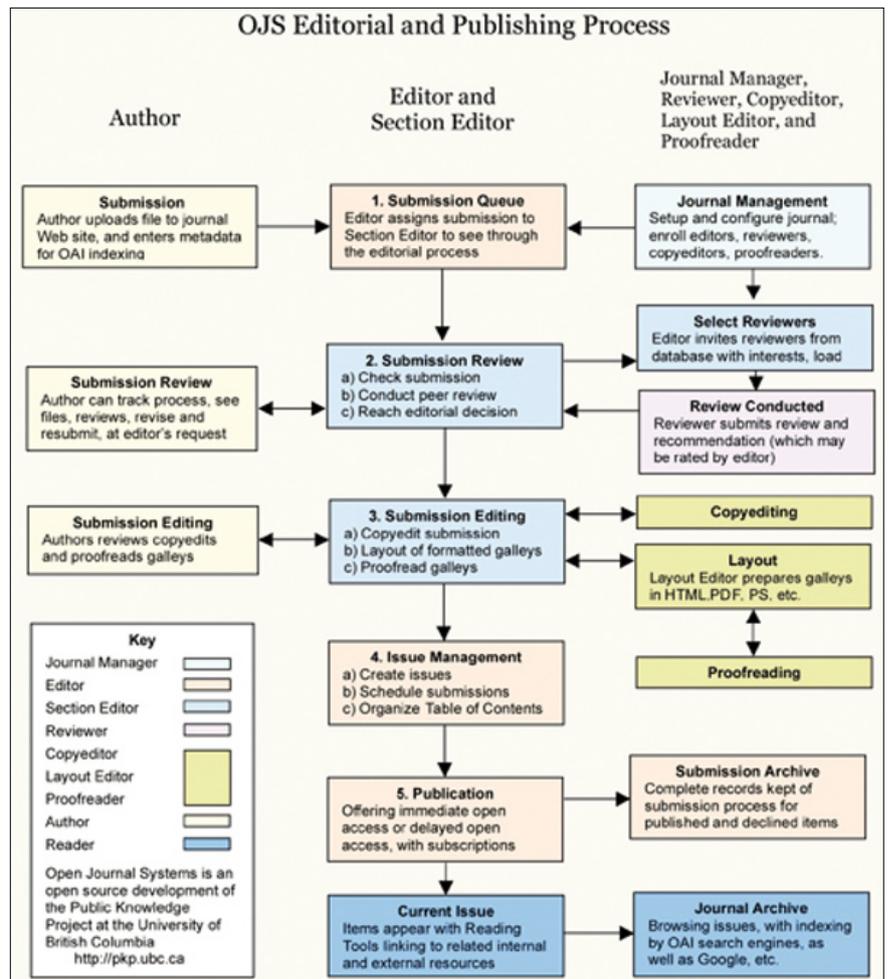
“La principal ventaja que el software libre ha aportado a los repositorios es la adaptabilidad”

libres. De los 689 depósitos que están en estos momentos inventariados y especifican el software con el que están realizados, más del 80% son de software libre (30% *DSpace*, 28% *EPrints*, 7% *OPUS*, 2% *ETD-db*, etc.).

<http://www.openoar.org>

Las ventajas que este tipo de software ha aportado a los repositorios son muchas, pero sin lugar a dudas, la principal ha sido la adaptabilidad. Puede que haya principios básicos genéricos, pero cada proyecto es singular. En los cuatro repositorios cooperativos citados se han realizado mejoras que han acercado el software a los requisitos que se han marcado para cada uno de ellos. Aparte de la adaptabilidad y de la gratuidad, los depósitos también se han beneficiado de otras ventajas (Mas, 2006), como el escrutinio público, donde el código está sometido a un proceso de revisión que aporta dinamismo a la corrección de errores, o la independencia del proveedor, hecho que garantiza su continuidad teórica.

A principios de 2004 se formó un grupo de trabajo para acordar la selección conjunta de un software libre para gestionar repositorios electrónicos de objetos digitales (Borràs et al, 2006). El estudio se centró en los programas *ARNO*, *Cdsware*, *DSpace*, *EPrints*, *Fedora*, *i-TOR* y *MyCore*, todos ellos propuestos en un informe del *Open Society Institute* (2004). Se trataba de un grupo mixto entre bibliotecarios e informáticos del *CBUC* y algunas universidades catalanas. Se quiso una selección conjunta para crear repositorios institucionales con el mismo programa, y de esta manera poder crear un grupo de usuarios



Open Journal Systems (OJS)

que colaborase en la solución de problemáticas comunes. Finalmente se escogió *DSpace*, que describiremos más adelante, por su extenso uso a nivel internacional, por ser un sistema acabado que podía usarse inmediatamente, por su alto nivel de prestaciones y por su facilidad de uso, tanto para el usuario final como para su administrador.

En la literatura pueden hallarse otros trabajos comparativos sobre aplicaciones para el desarrollo de repositorios. Jesús Tramullas y Piedad Garrido hacen un extenso inventario (Tramullas-Garrido, 2006: 171-181) donde después proponen un modelo para evaluar las prestaciones. Para el caso concreto de revistas electrónicas existen programas que además de poder depositar los artículos, sirven también para realizar todo el proceso de revisión por pares, edición y

publicación. En el proceso de selección de *RACO* fueron estudiados *DSpace+CWIS*, *uPortal*, *CONTENTdm* y *Lodel* y finalmente se escogió *Open Journal Systems (OJS)*, al tratarse de la opción más completa y acabada para revistas.

De todos los ejemplos mencionados, se describen más detalladamente los que se han usado en la elaboración de repositorios cooperativos catalanes. Para *TDR* se adaptó el programa *Electronic Theses and Dissertations (ETD-db)* de la *Universidad de Virginia Tech*; para *Recercat*, se seleccionó *DSpace*; para *RACO* se usa *OJS*; y para *Padicat* conviven el recolector *Heritrix*, el indexador y buscador *NutchWAX* y la interfaz de consulta *WERA* y *Wayback*.

De hecho, *TDR* inició su camino como base de datos, pues en 1999

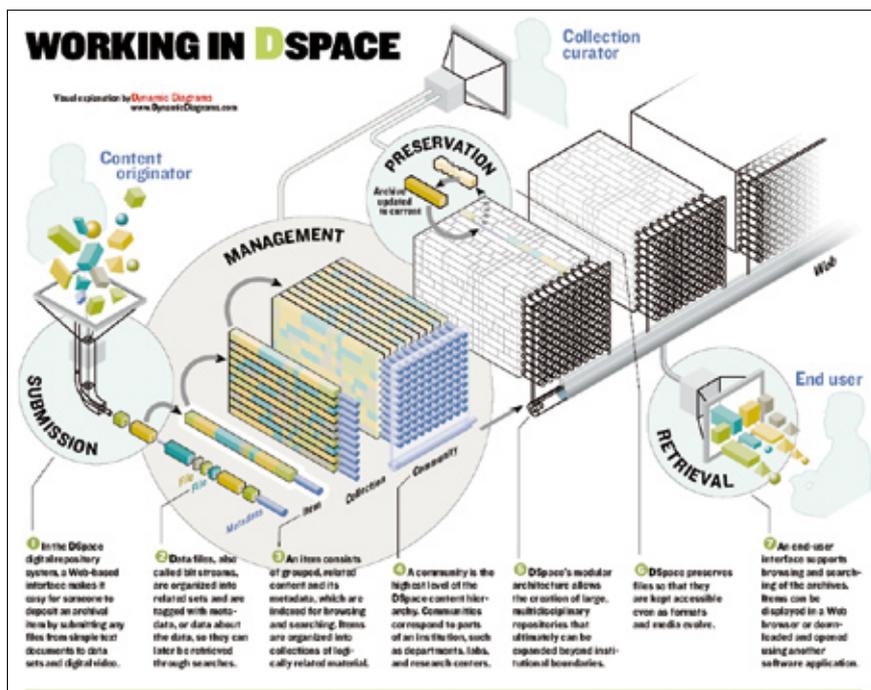
todavía ni la *Scholarly Publishing and Academic Resources Coalition (Sparc)* había definido los repositorios institucionales. El programa escogido para su implementación, *ETD-db*, fue una buena elección en su momento pues cumplía con los requisitos del repositorio y era el único. Hoy en día todavía cumple satisfactoriamente con su misión; sin embargo, la evolución del software en comparación con otros, por ejemplo *DSpace*, ha sido notablemente más lenta, hecho por el cual se está estudiando su sustitución a medio plazo. A continuación nos centramos por lo tanto en *DSpace* como aplicación para repositorios de propósito general; en *OJS* para la edición y publicación de revistas; y en los diferentes programas agrupados bajo el paraguas de *Heritrix* para repositorios de sitios web.

<http://etd.vt.edu>

### 1. DSpace

Es un programa de código abierto creado por el *Massachusetts Institute of Technology (MIT)* y *Hewlett Packard (HP)*. Se trata de un repositorio digital innovador que captura, almacena, indiza, preserva y redistribuye los resultados de la investigación de una organización. Existen 275 instalaciones en 45 países y permite búsquedas por autor, título, universidad, centro de investigación, etc. También es posible suscribirse al servicio de alertas de una colección y otros elementos propios de la biblioteca 2.0.

El hecho de ser de código abierto ha permitido adaptarlo a los requisitos de *Recercat*: personalizar la presentación; normalizar la búsqueda con diacríticos; adaptar la clasificación por materias de los documentos según la CDU; establecer un control de autoridades para los nombres de los autores y las instituciones; implementar las búsquedas por tipo de documento, y realizar un módulo de estadísticas personalizado, pues los diferentes *plugins* que se ofrecen con *DSpace*



DSpace

ce no cumplían con los objetivos deseados para el proyecto. También se resolvieron errores en el módulo de importación de metadatos en los casos en los que se producían modificaciones. Y con todo esto se ha contribuido a la comunidad *DSpace*, tomando un papel activo en el movimiento del *software* libre.

<http://www.dspace.org>

### 2. OJS

Es un programa de código abierto desarrollado por el *Public Knowledge Project (PKP)* con el objetivo de promover el acceso a la investigación, facilitando la administración y publicación de revistas científicas. Permite realizar todo el proceso de revisión por pares y publicación de revistas. En su vertiente de repositorio permite búsquedas por autor, título de artículo, título de publicación o tema. Para cada artículo es posible consultar la bibliografía, las estadísticas, el metadato que lo describe, recomendar su lectura a colegas, etc.

Para *RACO*, aparte de normalizar las búsquedas con diacríticos y traducirlo al catalán, se han llevado a cabo dos nuevos módulos. El primero, para implementar la clasificac-

ción temática de las revistas (según la CDU), puesto que la aplicación estaba más orientada a artículos que a revistas, y características como éstas no se encontraban disponibles. El segundo módulo simplifica el proceso de publicación para aquellas revistas que no tengan la suficiente estructura o volumen como para atacar el detallado proceso de publicación de *OJS*.

<http://pkp.sfu.ca/ojs>

### 3. Heritrix, NutchWAX, Wera, Wayback, etc.

*Padicat* es un proyecto pionero en España, pero las aplicaciones usadas por proyectos similares internacionales son bastante comunes. Un conjunto de programas informáticos permite recoger, almacenar, organizar, preservar y ofrecer acceso permanente a las webs publicadas en la Red. *Heritrix* es el recolector de *Internet Archive* y su misión es arrastrar y almacenar en archivos comprimidos las páginas de las webs seleccionadas. *NutchWAX* (basado en *Google*) genera los índices que más tarde serán usados para las búsquedas. Y finalmente, *Web Archive Access (WERA)* es una solución libre para buscar y navegar por las

colecciones de documentos de los distintos sitios web archivados.

Se ha evaluado, y pronto estará también en producción el buscador *Wayback*, que a diferencia de *WERA* sólo permite las búsquedas por url en lugar de a texto completo. Ambos se complementan, pues *Wayback* será usado para hacer búsquedas a partir de un listado catalogado de contenidos. En estos programas se están realizando también adaptaciones: para *Heritrix* se ha implementado un módulo de estadísticas y se está produciendo una herramienta de catalogación.

<http://crawler.archive.org>

<http://archive-access.sourceforge.net/projects/nutch>

<http://archive-access.sourceforge.net/projects/wera>

<http://archive-access.sourceforge.net/projects/wayback>

## Características deseables de los repositorios

### 1. Interoperabilidad

La manera más efectiva de aumentar la visibilidad de los contenidos de los depósitos es comparando metadatos con otros; de esta manera los materiales se pueden dejar en local y ser buscados desde otros repositorios internacionales. Para hacer esto posible se usa un protocolo de interoperabilidad creado por la *Open Archives Initiative (OAI)*, organismo que desarrolla y promueve estándares que se proponen facilitar la diseminación eficiente del contenido. *OAI* tiene sus raíces en el acceso abierto y los movimientos de repositorios institucionales. Con el tiempo, sin embargo, su trabajo se ha ampliado para promover el acceso amplio a recursos digitales para la escolarización, la educación y la ciencia.

<http://www.openarchives.org>

Uno de estos estándares es el protocolo de *OAI* para la recolec-

ción de metadatos (*OAI-PMH*), un mecanismo de bajo nivel para la interoperabilidad entre repositorios. Hay dos actores posibles en el protocolo: por un lado, los proveedores de datos (*data providers*) son repositorios que exponen metadatos estructurados vía protocolo *OAI-PMH*. Por otro, los proveedores de servicio (*service providers*) efectúan peticiones *OAI-PMH* a los proveedores de datos para recolectar esos metadatos.

Existen ya diversas iniciativas de recolectores de metadatos. En España, desde mayo de 2006 se pueden buscar en *TDR* las tesis de todos los repositorios españoles que publican a través de *OAI-PMH*, convirtiéndose así en el primer recolector de tesis españolas. En la Comunidad de Madrid, *e-Ciencia* permite buscar entre la producción científica de sus universidades. A nivel europeo, el proyecto *Driver*, financiado por la Comunidad Europea, quiere recoger de la red de repositorios de los diferentes países los del ámbito de la investigación y la educación superior, e incluso ha publicado unas directrices para ello (*Driver*).

### 2. Preservación a largo plazo de los datos

Es un aspecto complejo, y precisamente se encargó un estudio sobre el tema para los repositorios cooperativos catalanes (**Gallart; Keefer, 2007**). Existen diferentes aproximaciones al problema, algunas de las cuales son resueltas por los propios programas; en otras ocasiones se necesitan aplicaciones adicionales, aunque hay que tener en cuenta además otros aspectos de carácter organizativo que nada tienen que ver con unos y ceros.

Desde el punto de vista tecnológico, algunas de las estrategias que pueden ser adoptadas son la migraciones de formatos, los controles de la persistencia de los datos (*checksums*) y la diversidad, ya sea en repositorios oscuros (no

accesibles) o mediante dispersión geográfica, como el modelo *Lots Of Copies Keep Stuff Safe (Lockss)*. El *Open Archival Information System (OAIS)*, elaborado por el *Consultative Committee for Space Data Systems* de la *NASA*, ha sido adoptado por diferentes instituciones que tienen la misión explícita de preservar información, y se ha convertido en el estándar *ISO 14721:2003*.

La *Universidad de Cambridge* colabora activamente en el proyecto *DSpace*, y está especialmente interesada en los aspectos de preservación de recursos digitales al ser una de las bibliotecas receptoras del depósito legal de su país. En este marco está al frente del proyecto *Digital Preservation Tools and Strategies*, enfocado a mejorar la funcionalidad de *DSpace* como programa capaz de dar soporte a tareas de preservación.

### 3. Comunidad de usuarios

La robustez es garantía de continuidad del software. Al contrario de lo que sucede con los programas comerciales, en software libre no se está atado a una única empresa. Sin embargo, hay muchos programas que no llegan a buen puerto por diferencias en la comunidad, falta de recursos, cambio de inquietudes de los desarrolladores principales, etc. Afortunadamente, la mayoría de proyectos de repositorios institucionales elaborados con aplicaciones libres están siendo llevados a cabo en un entorno corporativo, principalmente gracias a fondos de investigación financiados por administraciones públicas (**Chawner, 2005**). Esto implica unos equipos de desarrollo estables que repercuten positivamente en la calidad de las aplicaciones resultantes. Las principales organizaciones que apoyan estos proyectos son universidades, centros de investigación y bibliotecas universitarias.

Algunas, como por ejemplo la comunidad de *DSpace*, han creado

fundaciones para liderar y dar soporte al crecimiento del software. Además, se organizan eventos periódicos como el *DSpace user group meeting*, ocasión que sirve para poner en común todas las iniciativas que las diferentes instituciones están realizando en el programa y donde se dibuja el *roadmap* del producto. El último se ha realizado en octubre de 2007 en Roma.

Además de que la comunidad esté respaldada por grandes instituciones también es muy importante contar con el apoyo de grupos locales, pues muchas veces determinados problemas preocupan más a unos que a otros. Por ejemplo, *DSpace* es un programa de origen anglosajón y la problemática de las búsquedas correctas con diacríticos salió de un miembro del *Grupo de usuarios DSpace de España (GUDE)* quien, como es lógico, tenía una escala diferente de urgencia por resolver el problema.

<http://sod.upc.es/gude/index.php/Portada>

## Conclusiones

Los repositorios de e-información son una parte importante del movimiento de acceso abierto. Además, ayudan a las bibliotecas a cumplir con su misión de difundir los recursos electrónicos generados en las propias instituciones, pasando antes por las fases de almacenamiento y organización, así como de garantizar su preservación. Para poder adaptar el programa informático a los requisitos concretos del re-

**“Si unimos los conceptos de acceso abierto y software libre obtenemos un binomio ganador, pero si además le añadimos el concepto de estándares abiertos, obtenemos una apuesta robusta y segura”**



Dipòsit de la Recerca de Catalunya (Recercat)

positorio es importante la elección de software libre, que es el tipo de programas que más se está usando para estos sistemas. Si unimos los conceptos de acceso abierto y software libre obtenemos un binomio ganador (Vega, 2006), pero si además le añadimos el concepto de estándares abiertos (Corrado, 2005), necesario para una buena interoperatividad entre depósitos, obtenemos una apuesta robusta y segura para la creación de repositorios institucionales.

Existen varias aplicaciones libres para implementar estos productos. Aunque están en evolución continua –condición intrínseca del software libre–, pueden ser usadas en entornos de producción con resultados plenamente satisfactorios. *TDR* está operativo desde 2001 y cuenta con 4.800 tesis de 17 universidades de diferentes comunidades autónomas españolas.

*Recercat* funciona desde 2005. En los primeros meses se llegó rápidamente a la cifra de 1.000 documentos de investigación y hoy ya cuenta con más de 3.800 de 15 instituciones. Un año más tarde llegaron *RACO* y *Padicat*. En el primero participan hasta 31 instituciones con 143 revistas y más de 38.000 artículos; el segundo ofrece 2.400 capturas de 810 sitios web, lo que

significa 24 millones de ficheros (páginas web, imágenes, etc.).

Finalmente un apunte: este artículo se centra en tecnología, pero es importante no perder la perspectiva de que la creación de repositorios desde el punto de vista tecnológico no es un problema (*CD-LOR*). Como se ha descrito, existen muchas opciones más o menos desarrolladas que lo permiten. Y existen ejemplos en producción que lo demuestran, como son *TDX*, *Recercat*, *RACO* y *Padicat*. Sin embargo, hay otros aspectos que son más complejos, como dotarlos de contenidos, ponerse de acuerdo en los metadatos, establecer circuitos de trabajo, etc. Elaborar de forma cooperativa los repositorios ayuda en gran medida a alguno de estos aspectos, pues se pueden adoptar procedimientos comunes, desarrollar políticas de preservación a largo plazo y compartir los costes del desarrollo y su gestión. Además, el hecho de estar en un solo repositorio documentos de diferentes instituciones aumenta su visibilidad, consiguiendo de esta manera una mayor difusión de la investigación realizada y ayudando al desarrollo de la e-ciencia y de la sociedad de la información del país.

## Agradecimientos

La creación de los repositorios que han servido de ejemplo a lo lar-

go del artículo no hubiera sido posible sin el esfuerzo de mucha gente, tanto de dentro de los consorcios (*Cesca* y *CBUC*) como de fuera. De los consorcios, hay que citar a **Marc Boix, Joan Cambras, Núria Comellas, Jesús Martín, Maite Novo, Jordi Prats, Sandra Reoyo, Ramon Ros, Marc Terrones, Xavier Torelló y Natalia Torres**, además de los directores **Miquel Huguet y Lluís Anglada**, artífices de los proyectos. Fuera de los consorcios, la lista crece demasiado exponencialmente como para individualizar el agradecimiento, pues en cada institución participante en alguno de los repositorios hemos hallado la colaboración necesaria sin la cual no hubiera sido posible darles vida.

## Referencias

**Anglada, Lluís.; Reoyo-Tudó, Sandra; Ros-Gomé, Ramon.** “Los repositorios como componentes esenciales de las bibliotecas digitales: la experiencia de las bibliotecas universitarias de Cataluña”. En: *3º Jornada sobre la biblioteca digital universitaria*, 2005.

**Borràs, Antoni; Carreño, Juan; Jorba, Ferran; Prats, Jordi; Ros, Ramon.** “Programari de codi lliure per gestionar dipòsits digitals: el procés de tria dut a terme al CBUC”. En: *BiD: textos universitaris de biblioteconímia i documentació*, 2006, junio, n. 16.

*CD-LOR. Community dimensions of learning objects repositories*. Consultado en: 04-10-07. <http://www.academy.gcal.ac.uk/cd-lor>

**Chawner, Brenda.** “F/OSS in the library world: an exploration”. En: *Open source application spaces: fifth workshop on open source software engineering (5-Wosse)*. ACM: St. Louis, 2005.

**Corrado, Edward M.** “The importance of open access, open source, and open standards for libraries”. En: *Issues in science and technology librarianship*, 2005, n. 42. Consultado en: 04-10-07. <http://www.istl.org/05-spring/article2.html>

*Declaración de Berlín sobre el acceso abierto al conocimiento en el ámbito de las ciencias y las humanidades*. Consultado en: 04-10-07. <http://www.zim.mpg.de/openaccess-berlin/berlin-declaration.html>

*Driver. The Driver guidelines*. Consultado en: 04-10-07. <http://www.driver-support.eu/en/guidelines.html>

**Gallart-Marsillas, Núria; Keefer, Alice.** *Preservació de dipòsits digitals*, 20 marzo 2007. Consultado en: 04-10-07. <http://hdl.handle.net/2072/4078>

**Huguet, Miquel; Anglada, Lluís; De-la-Vega-Sivera, Ricard.** “Catalan policies and experiences on cooperative repositories”. En: *Ist Iberian*

*grid infrastructure conference (Ibergrid)*, 2007, pp. 63-75.

**Lluca, Ciro.** “Archivando la web, el proyecto Padicat (Patrimonio Digital de Cataluña)”. En: *El profesional de la información*, 2006, noviembre-diciembre, v. 15, n. 6, pp. 473-478.

**Mas-Hernández, Jordi.** “Programari lliure: tècnicament viable, econòmicament sostenible i socialment just”. Barcelona: Generalitat de Catalunya, Escola d'Administració Pública de Catalunya, 2006, 140, p. ISBN 84-393-6995-6.

*Open Society Institute. A guide to institutional repository software. V. 3.0. Enero 2004*. Consultado en: 04-10-07. <http://www.soros.org/openaccess/software>

*Sparc. The case for institutional repositories: a Sparc position paper*. Consultado en: 04-10-07. <http://www.arl.org/sparc/publications/papers.html>

**Tramullas, Jesús; Garrido-Picazo, Piedad.** “Software libre para repositorios institucionales: propuestas para un modelo de evaluación de prestaciones”. En: *El profesional de la información*, 2006, mayo-junio, v. 15, n. 3, pp. 171-181.

**De-la-Vega-Sivera, Ricard.** et al. “Dipòsits col·lectius d'e-informació”. En: *V Jornades de programari lliure*, 2006. Consultado en: 04-10-07. <http://hdl.handle.net/2072/2195>

**Ricard de-la-Vega-Sivera, Centre de Supercomputació de Catalunya, Barcelona.**

[rdelavega@cesca.es](mailto:rdelavega@cesca.es)

## Próximos temas centrales

Marzo 2008

Mayo 2008

Julio 2008

Septiembre 2008

Noviembre 2008

**Innovación en bibliotecas públicas**

**Presente y futuro de la profesión**

**Libros electrónicos**

**Información en la empresa**

**Redes sociales**

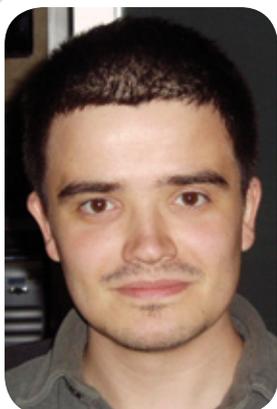
Los interesados pueden remitir notas, artículos, propuestas, publicidad, comentarios, etc., sobre estos temas a:

[epi@elprofesionaldelainformacion.com](mailto:epi@elprofesionaldelainformacion.com)

# Software libre en las bibliotecas de la UPC

Por Javier Clavero, Eulàlia Formentí y Toni Prieto

**Resumen:** Se exponen los proyectos y experiencias de implantación y uso de software libre en las bibliotecas de la Universitat Politècnica de Catalunya (UPC). El software libre es un movimiento que tiene cada vez mayor importancia, sobre todo en ámbitos universitarios. El uso de estándares potencia la interoperabilidad entre sistemas, lo que redundará en una mayor difusión de los datos. Las bibliotecas de la UPC están creando depósitos institucionales para dar acceso a la producción docente y de investigación a través de plataformas de software libre como DSpace. Las bibliotecas de la UPC seguirán trabajando en proyectos de software libre y promoviendo el uso del mismo entre sus usuarios y entre la comunidad universitaria.



**Javier Clavero**, licenciado en ingeniería informática por la Facultad de Informática de Barcelona. Es técnico IC de la Biblioteca Rector Gabriel Ferraté del campus nord de la Universitat Politècnica de Catalunya.



**Eulàlia Formentí** es licenciada en informática por la Universitat Politècnica de Catalunya en 1985 y responsable de la Càtedra de programari lliure de la UPC.



**Toni Prieto**, licenciado en ingeniería informática por la Facultad de Informática de Barcelona, es técnico IC del Servicio de Bibliotecas y Documentación de la Universitat Politècnica de Catalunya.

**Palabras clave:** Software libre, DSpace, Biblioteca, Universidad, Depósito, Linux, Interoperabilidad, Estándares.

**Title:** Free software at the UPC libraries

**Abstract:** We reviewed projects and outcomes in the implementation and use of free software in the Technical University of Catalonia (UPC) libraries. Free software is a growing movement with steadily increasing significance, especially in university environments. The use of standards improves the possibility of integration with other systems and enhances the efficient management of library data. At the moment UPC libraries are developing their institutional repositories, using open source software like DSpace to disseminate the institution's teaching and research activities. The UPC libraries will continue to implement this cost effective software and promote its use among library users and the university community.

**Keywords:** Free software, DSpace, Library, University, Repository, Linux, Interoperability, standards.

**Clavero, Javier; Formentí, Eulalia; Prieto, Toni.** "Software libre en las bibliotecas de la UPC". En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, pp. 56-63.

DOI: 10.3145/epi.2008.ene.06

## 1. El software libre

**FORMA PARTE DE UNA IDEOLOGÍA más universal conocida como "conocimiento libre". La idea fundamental de esta filosofía es que el conocimiento no puede ser un bien comerciable y, por tanto, la figura de los derechos de autor tiene un impacto negativo sobre la generación y difusión del conocimiento y del progreso.**

La *Wikipedia* define así el conocimiento libre<sup>1</sup>: "Hasta hace pocas décadas el conocimiento era

patrimonio de la humanidad, y existía libertad de uso y acceso al mismo. Con la aparición de límites artificiales como las patentes y los derechos de autor, entre otros (mal denominados en conjunto propiedad intelectual), el conocimiento perdió esta característica. Mientras que el avance de la tecnología por lograr una sociedad digitalizada y conectada brinda al conocimiento humano nuevas posibilidades de crecimiento, estos límites artificiales están restringiendo la posibilidad de aprovechar este potencial".

Internet pone de manifiesto este problema: la tecnología nos permite difundir el conocimiento con una facilidad y velocidad jamás imaginadas, pero la *Ley de propiedad intelectual*<sup>2</sup> (*LPI*) pone cortapisas a esta difusión. Tenemos que revisar el marco legal de la propiedad intelectual para adaptarlo a las necesidades actuales y permitir que el conocimiento sea patrimonio de la humanidad.

El contexto legal que ampara el software es también la *LPI*: los

programas legalmente son obras intelectuales, al igual que un libro, un cuadro o una fotografía. Las licencias de software hacen las veces de contrato entre el desarrollador o el distribuidor del mismo y el usuario; en ellas se establecen las condiciones de uso y los derechos de autor que se transfieren o restringen al usuario.

**Richard Stallman**, creador del proyecto *GNU*<sup>3</sup> y máximo impulsor del concepto del software libre, lo definió como aquel que otorga al usuario los siguientes derechos para:

- Ejecutarlo para cualquier propósito.
- Estudiar cómo funciona y adaptarlo a sus necesidades. Es indispensable el acceso al código fuente.
- Redistribuirlo.
- Modificar el programa y distribuirlo con las modificaciones realizadas. En este caso también es necesario el acceso al código fuente.

Las licencias de software libre garantizan estos derechos a los usuarios. La *Free Software Foundation*<sup>4</sup> redactó la primera licencia de software libre, y la más utilizada, la *General Public License (GPL)*<sup>5</sup>.

El movimiento del software libre se organiza en comunidades de desarrollo organizadas por proyectos, cada una generalmente con un líder, los desarrolladores y los colaboradores. El líder tiene que ser una persona muy activa, encargada de las relaciones personales con el resto de colaboradores, de los temas relacionados con nuevas versiones y de la promoción (documentación, sesiones informativas, *software parties*, *sprint sessions*, etc.). La organización de las comunidades es poco jerárquica y el reconocimiento entre sus miembros es por meritocracia. Las personas que forman una comunidad están altamente motivadas, técnicamente prepara-

das y fuertemente comprometidas con el objetivo a conseguir.

## 2. Las universidades y el software libre

Las universidades nacionales e internacionales participan activamente en el debate del conocimiento libre y en particular, del software libre. En un principio, todo se originó persona a persona: investigadores, personal de administración y servicios, y estudiantes. La mayoría del software utilizado en las universidades para los programas de investigación y docencia es libre y muchos de estos proyectos nacen en grupos de investigación, o con investigadores de universidades que participan, a título individual o del grupo. La incorporación de las instituciones a este movimiento es más reciente.

Los programas que se realizan para la investigación forman parte de la misma y su código es utilizado por distintos grupos sin licencias que regulen su utilización. Sólo son conocidos por los grupos afines.

Las universidades nacionales han empezado a crear oficinas para impulsar y dinamizar el software libre. En el proyecto *IRIS-libre*<sup>6</sup> de *RedIris*<sup>7</sup> se recogen las distintas organizaciones de cada una de las universidades con sus proyectos.

## 3. La UPC y el software libre

Como universidad tecnológica, la *UPC* es una entidad consumidora, productora y formadora de software: el movimiento del software libre está presente en cada uno de estos ámbitos. Como en el resto de las universidades, el uso y desarrollo de software libre es común en cada una de las actividades a realizar, pero también es un movimiento ligado a las personas y a los grupos de investigación.

En 2004 la *UPC* creó la *Càtedra de programari lliure (CPL)*

para impulsar y dinamizar esta tendencia. La *CPL* está implementando el *Pla estratègic del programari lliure*, donde se recoge la necesidad de introducir la filosofía del conocimiento libre en cada uno de las áreas de la *UPC* y las líneas estratégicas que deben impulsarse.

Las bibliotecas de la *UPC* han adquirido el compromiso de trabajar para utilizarlo cada vez más, tanto para aplicaciones y plataformas web como en los ordenadores que se ponen a disposición de los usuarios en las bibliotecas.

---

**“En 2004, la UPC creó la Càtedra de Programari Lliure (CPL) para impulsar y dinamizar el uso del software libre”**

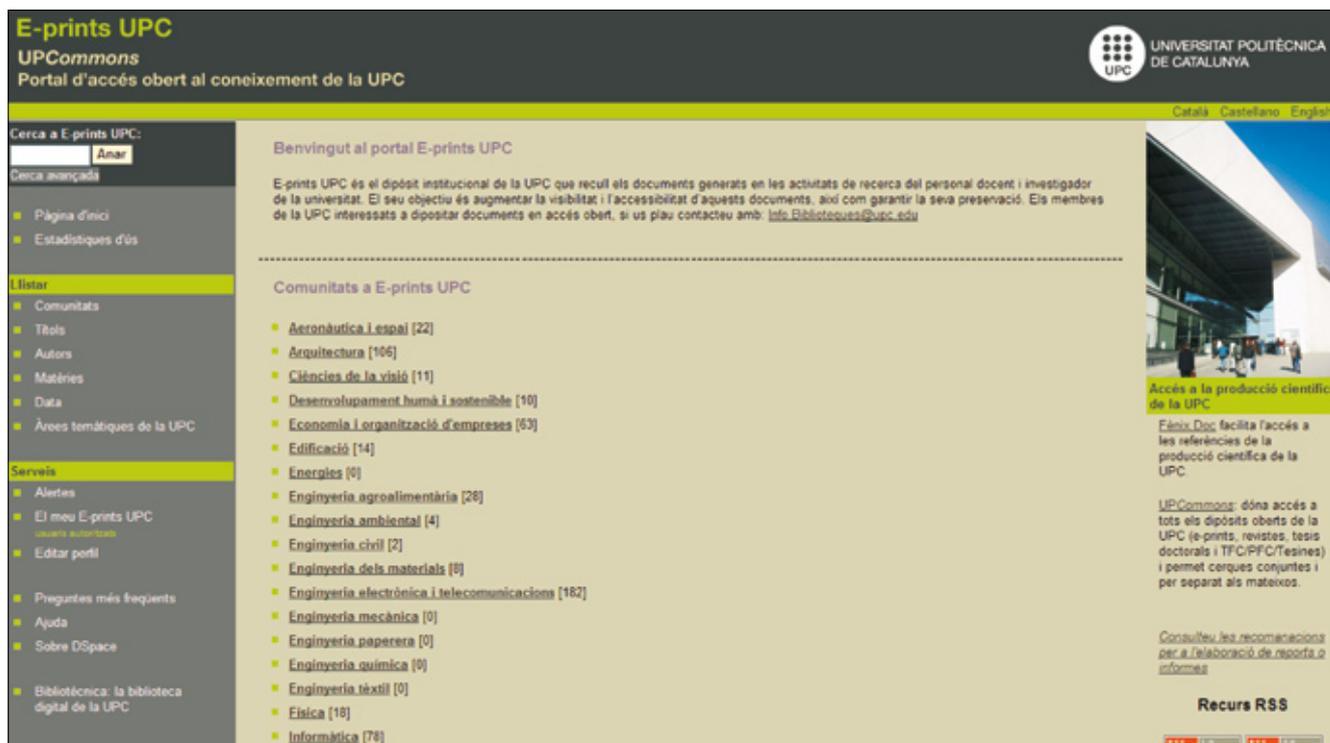
---

## 4. El software libre en las aplicaciones web de las BUPC

En este punto se incluye tanto la producción de nuevas aplicaciones como la progresiva migración de otras ya en explotación, que fueron creadas mediante software propietario, en su mayoría de *Microsoft*. Por otro lado, la proliferación de diversas aplicaciones web en software libre muy completas y fáciles de implantar es otra motivación más en la realización de nuevos proyectos.

### 4.1. Implementación con software libre existente

Puede hallarse un gran número de soluciones muy completas, realmente útiles en el trabajo diario de una biblioteca universitaria. Un ejemplo son los wikis, muy prácticos para el trabajo en grupo del personal de las bibliotecas, o los blogs, excelentes como herramienta de comunicación e interacción. En el caso de las *BUPC*, se eligió *Me-*



Portal E-prints UPC

diaWiki<sup>8</sup> para la creación de wikis y Wordpress<sup>9</sup> para los blogs, que se instalan rápidamente sin apenas hacer modificaciones. Prácticamente sólo se tuvieron que configurar extensiones ya existentes, por ejemplo para integrar la validación de usuarios con el servidor LDAP<sup>10</sup> de la UPC, y personalizar algunos aspectos de la aplicación. Pero es DSpace<sup>11</sup> el programa con el que se ha tenido una mayor implicación.

#### 4.2. DSpace

Es una plataforma en código abierto para la implementación de depósitos documentales, desarrollada por las bibliotecas del Massachusetts Institute of Technology (MIT) y la empresa Hewlett-Packard, y para la cual han creado recientemente una compañía sin ánimo de lucro destinada a liderar su evolución y garantizar su futuro.

A principios de 2004 se creó un grupo de trabajo desde el Consorcio de Bibliotecas Universitarias Catalanas (CBUC) para acordar la selección de un programa en código abierto para los depósitos documentales de las universidades cata-

lanas, que culminó con la elección de DSpace. A partir de entonces, tanto el CBUC como varias universidades catalanas han iniciado sus depósitos mediante esta plataforma. En las BUPC se dispone de tres instalaciones para almacenar, diseminar y preservar los materiales producidos en la universidad. Es el caso de E-prints UPC<sup>12</sup>, que almacena los trabajos producidos por los investigadores de la institución, y el depósito Revistes i UPC<sup>13</sup>, que incorpora las revistas y documentación de congresos realizados en la universidad. Ambos se crearon desde cero, ya que anteriormente no existía ninguna aplicación que realizara sus funciones. En cambio, el tercer depósito, Treballs acadèmics UPC<sup>14</sup>, es el resultado de la migración del contenido de un sistema en ASP ya en explotación, donde se alojaban los trabajos académicos de fin de carrera de los estudiantes de la UPC.

Se está trabajando en la migración hacia DSpace de otro depósito, la Videoteca de la UPC, que inicialmente también se hizo con Microsoft y que alberga vídeos de confe-

**“Las bibliotecas de la UPC disponen de tres instalaciones de DSpace para almacenar, diseminar y preservar los materiales producidos en la universidad.**

rencias, clases, reuniones, etc., de la comunidad universitaria. Se pretende incorporar los documentos en formato Flash e integrar en la pantalla de datos de cada vídeo un visor para su reproducción. A pesar de no ser un formato libre, se ha escogido Flash dado que puede considerarse un estándar de vídeo en internet, ya que es uno de los más usados, sobre todo últimamente, gracias al impulso dado por sitios tan populares como YouTube. Por último, el portal UPCommons<sup>15</sup>, que es el encargado de dar acceso a todos estos depósitos y al resto de los gestionados por las BUPC y responsable de ofrecer un acceso unificado a sus contenidos, es también DSpace.

Dipòsit de Materials Docents UPC

Durante la realización de todos estos depósitos se han personalizado y modificado partes del software para adaptarlo a las necesidades de las bibliotecas. Estas innovaciones han sido tanto pequeños programas propios como instalaciones de parches disponibles en la web para solucionar fallos o incluir nuevas funciones. Se ha procurado que los cambios realizados sean mínimos, con el fin de que la futura actualización a las nuevas versiones, que a menudo ya incluyen las mejoras requeridas por los usuarios, sea lo más fácil posible.

El hecho de pertenecer a una gran comunidad de usuarios tiene la ventaja de que muchas de las necesidades y los problemas que surgen durante la puesta en marcha ya se hayan planteado y solucionado en otras instituciones. En este sentido, han sido muy útiles las herramientas que ofrece la comunidad de *DSpace* para la comunicación entre sus miembros, como las listas de distribución o el wiki. Además, en estas listas de distribución, especialmente en la destinada a dudas en el desarrollo técnico, los máxi-

mos responsables de los cambios que se establecen en la plataforma están informando y discutiendo continuamente las decisiones tomadas, haciendo que su evolución sea muy abierta a la participación del resto de la comunidad.

### 4.3. GUDE

Tras observar que un gran número de universidades españolas tenían proyectos de implementación de depósitos institucionales mediante *DSpace*, se decidió impulsar la creación del *Grupo de Usuarios de DSpace de España (GUDE)*<sup>16</sup>. Éste nació con el objetivo de compartir, colaborar en las mejoras, participar en las comunidades mun-

**“Dado el gran número de universidades españolas con proyectos de depósitos mediante *DSpace*, se decidió impulsar la creación del Grupo de Usuarios de *DSpace* de España”**

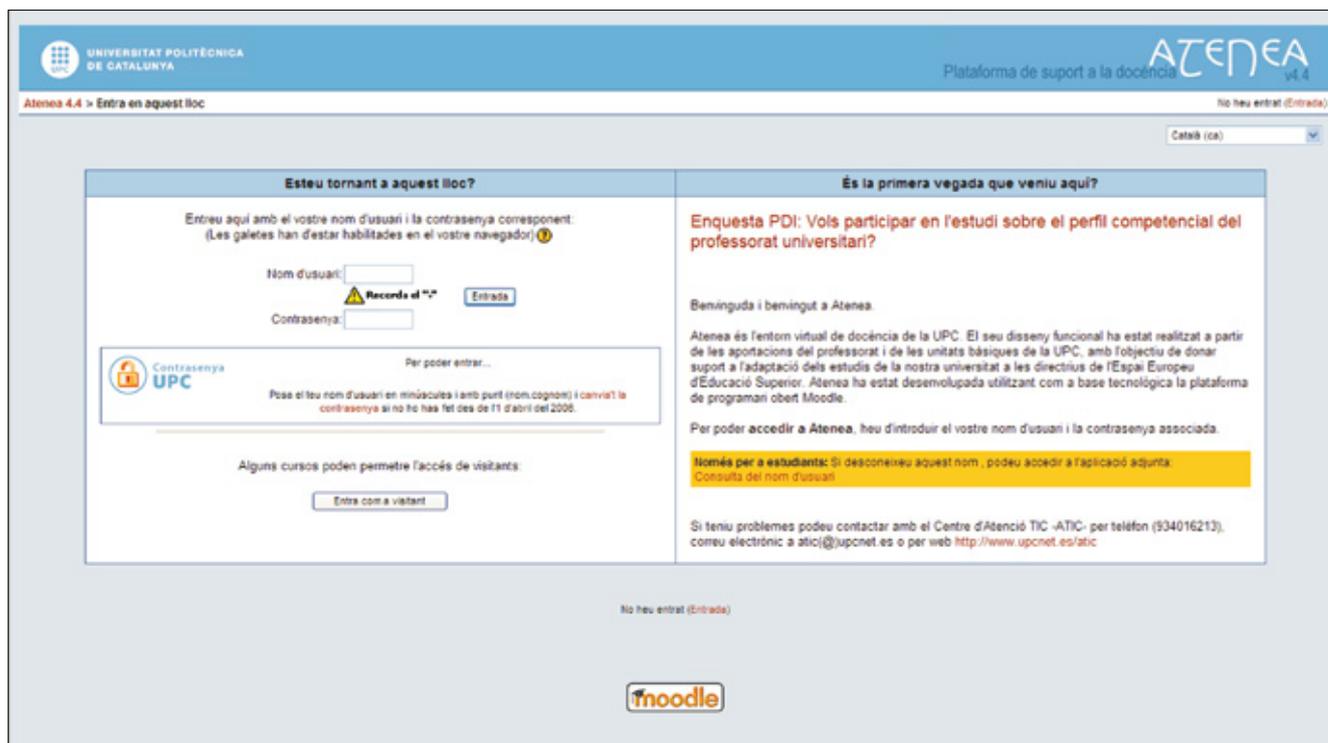
diales *DSpace* y organizar *workshops* de usuarios a escala española. Como canales de comunicación se dispone de una lista de distribución y un wiki donde se localizan las instalaciones y proyectos españoles, directorios, recolectores y otros.

Fruto del trabajo de varios miembros del grupo, se han traducido al castellano las nuevas versiones de las dos interfaces de usuario ofrecidas, así como algunos documentos de ayuda. En la lista de distribución, activa desde octubre del 2006, se comparten tanto dudas técnicas como otras de carácter más funcional, además de noticias y novedades.

### 4.4. Desarrollos propios

En algunas ocasiones se ha optado por hacer desarrollos propios, al no disponer de software que se ajustara a las necesidades requeridas. En estos casos se ha utilizado *PHP*<sup>17</sup> como lenguaje de programación y el gestor de base de datos *MySQL*<sup>18</sup>.

El ejemplo más significativo es el *Dipòsit de Materials Docents*<sup>19</sup>, que recopila y da acceso a los mate-



Atenea: el campus virtual de la UPC

riales de las asignaturas producidos por profesores y alumnos de la universidad en su actividad docente. Realiza y amplía las funciones que llevaba a cabo otro producto basado en software propietario, al que ha sustituido, y fue inicialmente creado como parte de un proyecto de fin de carrera de una estudiante de informática. Está formado internamente por tres partes separadas: el portal, donde se puede acceder a los materiales disponibles de las asignaturas mediante búsquedas; el depósito, donde se da entrada a los documentos y sus metadatos, y el gestor, en el que se listan y gestionan los materiales que se muestran en cada una de las asignaturas cursadas en la UPC.

Aunque inicialmente se consideró la opción de utilizar también *DSpace*, sobre todo para realizar las funciones de la parte de depósito, finalmente se desechó por dos razones: por un lado, no permitía la gestión de objetos compuestos hasta el nivel que se necesitaba y por otro, el sistema de circuitos de validación de metadatos no se ajustaba a las necesidades requeridas.

#### 4.5. Interoperabilidad y estándares abiertos

Uno de los aspectos más destacables del *Dipòsit de Materials Docents* es su integración a distintos niveles con tres aplicaciones de la universidad:

- *Atenea*: el campus virtual de la UPC<sup>20</sup>. A través de una pasarela de autenticación se permite introducir contenidos en su depósito y referenciarlos de manera transparente para el usuario.

- *Prisma*: el sistema de gestión académica. Fuente de datos para la obtención de las asignaturas con las que trabaja el gestor de contenidos.

- *Gauss*: sistema para la gestión e identificación de usuarios de la UPC desde donde se obtienen archivos para la actualización de privilegios de los profesores y se controla su autenticación a través del protocolo LDAP.

La utilización de software libre por otras entidades de la universidad ha facilitado, sin duda, esta integración. Un ejemplo claro es *Atenea*, que se migró hace unos años a la plataforma *Moodle*<sup>21</sup>.

En este aspecto juega también un papel fundamental el uso de estándares abiertos, muy ligados al concepto de software libre ya que, entre otras cosas, exigen la existencia de alguna implementación en abierto. Representan la herramienta más adecuada para la interoperabilidad de sistemas, ya sean aplicaciones de la propia institución o externas. El *Dipòsit de Materials Docents* sirve de ejemplo, ya que utiliza el protocolo estándar *SOAP*<sup>22</sup> para la comunicación entre las diferentes partes que lo forman (el depósito, el gestor y el portal) con el fin de que sean independientes y pueda reemplazarse alguna si es necesario sin tener que cambiar el resto.

Otro caso es la comunicación entre el depósito *E-prints UPC* y la aplicación *Fénix*, que gestiona la actividad académica del personal investigador de la universidad. Aquí, *Fénix* recupera e incorpora los metadatos de los materiales incluidos en el repositorio, tales como artículos, comunicaciones de congresos, etc., a través del protocolo estándar *OAI-PMH*<sup>23</sup>, permi-

tiendo así disponer, además de los metadatos, de un enlace al depósito para tener acceso al texto completo y así evitar duplicar el trabajo en el registro de datos. Este protocolo estándar también es el responsable de la incorporación de los metadatos en el portal *UPCommons* desde el resto de repositorios para tener un acceso unificado a todos los materiales disponibles en acceso abierto, así como de hacerlos públicos para que los recolectores y directorios tengan acceso a ellos.

## 5. El software libre en los ordenadores de las BUPC

Tal y como se ha comentado antes, también se ha experimentado en aplicaciones locales o de escritorio. La más interesante fue la creación e implantación de imágenes (ver definición)<sup>24</sup> basadas en *Linux*<sup>25</sup> para los ordenadores denominados *PUB*, que se encuentran en las 13 bibliotecas de la *UPC* y son utilizados por los usuarios para consultar el catálogo e internet. Inicialmente funcionaban con imágenes basadas en sistemas operativos *Windows*, lo que generaba dos grandes problemas para el personal técnico de las bibliotecas.

### 5.1. Problemática inicial

El primer y gran inconveniente estaba relacionado con mantener actualizada la imagen ya que, como se sabe, un sistema *Windows* consume muchos recursos, y más a cada nueva versión, lo que requiere ordenadores más y más potentes. Esto planteaba un dilema a las bibliotecas: o bien invertían parte del presupuesto en renovar cada cierto tiempo estos ordenadores, pudiendo tener así un sistema y unas aplicaciones actualizadas y funcionando con un rendimiento aceptable para el usuario, o bien no invertían y continuaban con los mismos ordenadores, pero con un sistema y unas aplicaciones desactualizadas e incluso obsoletas. Por desgracia, este último caso era lo más común,

ya que a menudo la inversión requerida era demasiado grande para la biblioteca.

Esto, como es fácil deducir, generaba problemas tanto de seguridad como de incompatibilidad. Se daban muchos casos de usuarios que intentaban abrir documentos realizados con una versión más actual que la aplicación disponible en la imagen, con el resultado de que no se visualizaba o lo hacía incorrectamente. Esto generaba insatisfacción y obligaba a la biblioteca a buscar rápidamente una alternativa, como la de dejar abrir al usuario su documento en un ordenador del propio personal. Por otro lado, también daba lugar a agujeros de seguridad, lo que nos lleva al segundo gran problema.

---

### “El software libre se basa en la existencia de estándares abiertos”

---

Dejando de lado el debate de si los sistemas *Windows* son seguros o no, la cuestión de la seguridad en estos ordenadores *PUB* se dividía en dos frentes. Por un lado, al tener muchas veces un sistema en una versión anterior hacía que los ordenadores fueran más vulnerables a los nuevos virus, poniendo en riesgo al propio ordenador y a los datos de los usuarios. Esto implicaba que el personal técnico tenía que asegurarse que tanto el sistema como el antivirus estuvieran actualizados y también obligaba a revisarlos periódicamente en busca de virus o fallos de seguridad. Por otro lado, a veces y por la misma razón, era propiamente un usuario malintencionado el que, con un poco de esfuerzo, conseguía vulnerar la seguridad y provocar fallos en el funcionamiento del sistema. Todo esto unido a que un sistema *Windows* de por

sí requiere cada cierto tiempo una reinstalación para un correcto y eficaz funcionamiento, provocaba que el personal tuviera que reinstalar la imagen muy a menudo, dejando al ordenador fuera de servicio durante un tiempo, a veces horas. Esto reducía su disponibilidad y provocaba ofrecer un peor servicio al usuario.

### 5.2. Migración a *Linux*

Por todos estos motivos y dado que en el entorno de la *UPC* cada vez hay una tendencia mayor a usar software libre y se conocían otras experiencias de migración a *Linux* de ordenadores de consulta como en el caso de las bibliotecas de la *Universidad Autónoma de Barcelona*, se decidió probar la alternativa de crear una imagen en *Linux*, concretamente una primera versión basada en la distribución *Debian*<sup>26</sup> y una segunda en *Ubuntu*<sup>27</sup>, que es la que actualmente se usa.

Esta nueva imagen es mucho más ligera, permitiendo su uso en ordenadores más antiguos con un rendimiento más que aceptable, lo que reduce considerablemente la inversión que se realiza en la adquisición de máquinas nuevas. También se puede decir que es más robusta y segura ya que, en el año y medio que lleva funcionando, no se ha detectado ningún problema de seguridad. Esto hace que en muy pocos casos se haya tenido que reinstalar, con lo que aumenta la disponibilidad de los ordenadores y mejora el servicio. Asimismo, incorpora herramientas de administración remota que permiten cambiar fácilmente el sistema sin tener que ir físicamente a cada una de las máquinas y de forma transparente al usuario. En conjunto, todo ello ha reducido el tiempo que el personal de las bibliotecas debe emplear en estos ordenadores, permitiendo invertirlo en otras tareas y en la realización de nuevos proyectos.

Es cierto que se ha mejorado en muchos aspectos, pero debe reconocerse que no todo han sido venta-

jas. A pesar de que *Linux* soporta la mayoría de dispositivos y cada día que pasa mejora, todavía existe incompatibilidad con ciertas tarjetas o periféricos, principalmente debido a que las empresas fabricantes todavía no apoyan firmemente al software libre. Concretamente se encontraron problemas con ciertas memorias *USB*, sobre todo en la primera imagen *Debian*, con varias impresoras y con algunas placas base y ciertas tarjetas gráficas, que todavía no eran soportadas en las versiones *Linux* que se usaron. En el caso de las memorias se tuvo que optar por dejar al usuario usar un ordenador del personal con sistema *Windows* para que pudiera acceder a ellas y se enviara los documentos vía correo electrónico. Gracias a la actualización de la imagen en su versión con *Ubuntu*, este problema parece haberse solucionado y raramente se encuentran dispositivos de este tipo incompatibles con el sistema.

Para el caso de las impresoras la solución no fue tan fácil, ya que a pesar de que la empresa fabricante sí proporcionaba controladores de sus máquinas para *Linux*, éstos no tenían las mismas funcionalidades que los equivalentes en *Windows*. Ello obligó a tener que montar un sistema complementario al controlador del fabricante que las añadiría. Obviamente, a pesar del trabajo que comportó, esto fue posible gracias a que se pudo modificar parte del código del gestor de impresión, cosa que en otros sistemas no habría sido posible. Estos periféricos, a pesar de ciertos fallos puntuales, funcionan a la perfección con este sistema modificado. Por último, se han encontrado algunos casos en que ciertas incompatibilidades con la placa base o con la tarjeta gráfica han provocado problemas, como la no detección del disco duro, un funcionamiento defectuoso del sistema gráfico de las *X-Windows*<sup>28</sup> o la imposibilidad de hacer funcionar la disquetera o el cd-rom. Normal-

mente estos inconvenientes se han solucionado modificando ficheros de configuración del propio *Linux*. En algún caso se han tenido que actualizar ciertos controladores y aún hay algún caso que está pendiente de ser solucionado.

Otro inconveniente, quizás el más importante que se ha ido encontrando, ha sido el desconocimiento por parte del propio personal y de los usuarios del funcionamiento de un sistema *Linux*, ya que la mayoría habían utilizado siempre máquinas *Windows*. Por ello, en las primeras fases de implantación de la nueva imagen se tuvo que invertir mucho tiempo y esfuerzos en la formación del personal y de los usuarios. Carteles, atención personalizada y páginas web fueron algunos de los métodos para ir acercando poco a poco el nuevo sistema a los usuarios. A pesar de que al principio éstos parecían resistirse o, en algún caso, hasta tenerle miedo, cada vez son más los que lo usan con total libertad y agilidad, como cualquier otro sistema. Es cierto que todavía hay usuarios que por comodidad, desconfianza o desconocimiento, siguen prefiriendo usar otros ordenadores de las bibliotecas que todavía tienen sistemas *Windows*.

## 6. Conclusiones

El software libre está cada vez más presente en las aplicaciones web y de escritorio que se utilizan en las *BUPC*, ya sea por el desarrollo de nuevos proyectos o por la progresiva migración de productos inicialmente llevados a cabo con software propietario.

En general, de la experiencia en las *BUPC*, se puede concluir que el uso y la migración de proyectos a software libre han sido positivos, pero debe indicarse que no es un proceso tan sencillo como a veces se quiere hacer creer. En el caso del uso de aplicaciones web ya creadas, es necesario en muchas ocasiones invertir tiempo en comprender

la aplicación antes de poder adaptarla, personalizarla o modificarla para las necesidades requeridas. En el caso de migraciones a software libre, al trabajo que ya supone una migración a veces hay que añadirle el de estudiar qué formato o estándar es el que más conviene y sobre todo, cuál es el que tiene más futuro, para evitar tener que migrar otra vez toda una aplicación si éste desaparece. En el caso de aplicaciones locales o de escritorio, hay que tener en cuenta que, sobre todo al principio, va a ser necesario invertir tiempo en formación y aclimatación de los usuarios, antes de conseguir un sistema estable, robusto y amigable para ellos y para el propio personal de las bibliotecas.

---

**“La migración a software libre ha de ser progresiva para que los usuarios puedan ir asumiendo poco a poco el cambio y descubriendo que el nuevo sistema puede ser tan útil como el anterior”**

---

También hay que tener en cuenta que la migración a software libre debe ser progresiva, sobre todo en este último ámbito de las aplicaciones de escritorio, ya que en el contexto de la web su uso es casi transparente al usuario. De este modo, todos los usuarios podrán ir asumiendo poco a poco el cambio y descubriendo que el nuevo sistema puede ser tan valioso y útil como el anterior. Es por eso que en las bibliotecas de la *UPC* aún se conservan ordenadores con sistemas *Windows* para aquellos usuarios que les cuesta más el cambio y también porque se prefiere ofrecer todas las alternativas para que se pueda escoger aquella que más gusta o conviene.

A pesar de estos pequeños problemas, las *BUPC* están plenamente satisfechas con el camino que se ha escogido hacia el software libre, ya que abre un mundo de posibilidades, sobre todo de interoperabilidad, hasta ahora muy difíciles de conseguir con las herramientas propietarias. Esto es muy importante en el ámbito universitario, ya que permitirá interactuar y enlazar con sistemas de otras unidades, y de otras universidades, unificando esfuerzos para crear plataformas más completas para compartir y difundir conocimientos y materiales de una forma mucho más potente de lo que se podría haber pensado hace unos años.

---

### “La apuesta de las *BUPC* por el software libre está presente en el nuevo plan estratégico *Aprèn* (2007-2010)”

---

La apuesta por el software libre está presente en el nuevo plan estratégico *Aprèn*<sup>29</sup>, donde se contempla la elaboración de una nueva versión de *Biblioteca, la biblioteca digital de la UPC*<sup>30</sup> en software libre, que ya está en progreso. A nivel técnico se pretende migrar progresivamente la plataforma a *PHP* utilizando un gestor de base de datos libre como *MySQL* o *PostgreSQL*<sup>31</sup> y adaptando si es posible una herramienta de gestión de contenido (*CMS*) en software libre para facilitar su mantenimiento.

Por último, las *BUPC* animan a, como mínimo, probar esta alternativa y las posibilidades que ofrece, así como a valorar los beneficios que puede reportar una migración. Las *BUPC* continuarán con su la-

bor, ya que se considera que ofrece muchas ventajas y su filosofía se complementa perfectamente con la de las bibliotecas y las universidades, que es difundir el conocimiento sin barreras ni límites.

## 7. Notas

1. Wikipedia.  
[http://es.wikipedia.org/wiki/Conocimiento\\_libre](http://es.wikipedia.org/wiki/Conocimiento_libre)
2. Ley de propiedad intelectual.  
<http://civil.udg.es/normacivil/estatal/reals/Lpi.html>
3. GNU.  
<http://www.es.gnu.org>
4. Free Software Foundation.  
<http://www.fsf.org>
5. GPL.  
<http://www.opensource.org/licenses/gpl-license.php>
6. IRIS-libre.  
<https://forja.rediris.es/plugins/wiki/index.php?OpcionesDeSoftwareLibre&id=205&type=g>
7. RedIris.  
<http://www.rediris.es/rediris>
8. MediaWiki.  
<http://www.mediawiki.org>
9. Wordpress.  
<http://wordpress.org>
10. LDAP.  
<http://es.wikipedia.org/wiki/LDAP>
11. DSpace.  
<http://www.dspace.org>
12. E-prints UPC.  
<http://upcommons.upc.edu/e-prints>
13. Revistes i Congressos UPC.  
<http://upcommons.upc.edu/revistes>
14. Treballs Acadèmics UPC.  
<http://upcommons.upc.edu/pfc>
15. UPCommons.  
<http://upcommons.upc.edu>
16. GUDE.  
<http://sod.upc.es/gude>
17. PHP.  
<http://www.php.net>
18. MySQL.  
<http://www.mysql.com>
19. Dipòsit de Materials Docents UPC.  
<http://e-md.upc.edu>
20. Atenea: el campus virtual de la UPC.  
<http://atenea.upc.edu>
21. Moodle.  
<http://moodle.org>
22. SOAP.  
<http://www.w3.org/TR/soap>

23. OAI-PMH.

<http://www.openarchives.org>

24. Imagen: archivo en el que figura toda la información del ordenador. Se usa como “molde” para configurar nuevos ordenadores sin tener que realizar todos los procesos de instalación de un sistema nuevo. A la vez sirve como copia de seguridad de los datos de ese ordenador antes de realizar alguna operación sobre él.

25. Linux.

<http://www.linux.org>

26. Debian.

<http://www.debian.org>

27. Ubuntu.

<http://www.ubuntu.com>

28. X-Windows.

<http://www.x.org>

29. Plan estratégico *Aprèn* de las *BUPC*.

[http://biblioteca.upc.edu/Pla\\_estrategic/apren2010.pdf](http://biblioteca.upc.edu/Pla_estrategic/apren2010.pdf)

30. *Biblioteca, la biblioteca digital de la UPC*.

<http://biblioteca.upc.edu>

31. PostgreSQL.

<http://www.postgresql.org>

## 8. Bibliografía

**Borràs, Antoni; Carreño, Juan; Jorba, Ferran; Prats, Jordi; Ros, Ramon.** “Programari de codi lliure per gestionar dipòsits digitals: el procés de tria dut a terme al CBUC”. En: *BiD: textos universitaris de biblioteconomia i documentació*, 2006, n. 16. Consultado en: 04-11-07.  
[http://www2.ub.edu/bid/consulta\\_articulos.php?fichero=16borras.htm](http://www2.ub.edu/bid/consulta_articulos.php?fichero=16borras.htm)

**De la Vega, Ricard,** et al. “Dipòsits col·lectius d’e-informació”. En: *V Jornades de programari lliure*, 2006. Consultado en: 04-11-07.  
<http://portal.jornadespl.org/ponencies/htm>

**Prats, Jordi; Prieto, Toni; Rico, Oriol.** “UP-Commons: portal d’accés obert al coneixement de la UPC”. En: *V Jornades de programari lliure*, 2006. Consultado en: 04-11-07.  
<http://portal.jornadespl.org/ponencies/upc-commons-portal-d-acces-obert-al-coneixementde-la-upc/>

**Jorba, Ferran.** “Els kiosks Linux a les biblioteques UAB”. En: *II Jornades de programari lliure*, 2003. Consultado en 04-11-07.  
<http://www.jornadespl.org/biblioteca/ii-jornades/ponencies/fjorba-2003.pdf>

**Javier Clavero, Eulàlia Formentí, Toni Prieto,** Universitat Politècnica de Catalunya.

[javier.clavero@upc.edu](mailto:javier.clavero@upc.edu)

[eulalia.formenti@upc.edu](mailto:eulalia.formenti@upc.edu)

[antonio.juan.prieto@upc.edu](mailto:antonio.juan.prieto@upc.edu)

# Biblioteca virtual cooperativa descentralizada con software libre Greenstone

Por Dominique Babini, Florencia Vergara-Rossi, Flavia Medici y Jessica González

**Resumen:** Trabajar con software libre fue una decisión sencilla para crear la Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe hace 10 años, pues el Consejo Latinoamericano de Ciencias Sociales (Clacso) es una red académica de 195 centros de investigación en 21 países de la región que está comprometida con el uso de plataformas libres y con el acceso abierto a los textos digitales de sus investigaciones. En este artículo se describe cómo se desarrolló y cuáles son los procesos de trabajo en esta biblioteca virtual cooperativa que funciona con software libre Greenstone para la sala de lectura, con buscador avanzado, y con software libre Plone para la interfaz gráfica. Ambos permiten la gestión descentralizada de contenidos: 9.000 textos digitales (libros, artículos de revistas, documentos de trabajo, ponencias) y bases de datos.

**Palabras clave:** Bibliotecas virtuales, Bibliotecas digitales, Software libre, Greenstone, Plone, ISIS, Edición electrónica, Clacso, Ciencias sociales, América Latina.

**Title: Use of the open source software Greenstone to develop a decentralized cooperative digital library**

**Abstract:** When the Latin America and the Caribbean Social Sciences Network of Virtual Libraries was created 10 years ago, it was an easy decision to work with open software because the Latin America and the Caribbean Social Science Council (Clacso), an academic network with 195 member institutes located in 21 countries, promotes the use of open software and open access to communicate and exchange research results in digital formats. The development and daily operations of this cooperative digital library are described, including Greenstone for the reading room with advanced search capacity and Plone for the graphical interface; both allow decentralized content management with 9.000 full text publications (books, journal articles, working documents, papers) and databases.

**Keywords:** Digital libraries, Virtual libraries, Open software, Greenstone, Plone, ISIS, Electronic publishing, Clacso, Social sciences, Latin America.

**Babini, Dominique; Vergara-Rossi, Florencia; Medici, Flavia; González, Jessica.** "Biblioteca virtual cooperativa descentralizada con software libre Greenstone". En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, pp. 64-70.

DOI: 10.3145/epi.2008.ene.07



**Dominique Babini, Florencia Vergara Rossi, Flavia Medici y Jessica González** son miembros del equipo de trabajo de la Oficina Coordinadora de la Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe.

## Introducción

EL CONSEJO LATINOAMERICANO DE CIENCIAS SOCIALES, *Clacso*, es una red de centros de investigación y docencia en ciencias sociales ubicados en 21 países de América Latina y el Caribe.

Resultó siempre muy difícil hacer circular publicaciones en for-

mato impreso para una red de casi 200 centros tan dispersos geográficamente, con una creciente producción académica. Esto se debe a las limitaciones propias de este formato: bajas tiradas de libros y revistas académicas, elevados costos de correo, complicaciones en préstamos interbibliotecarios entre distintas ciudades y países, y la falta de difusión impresa de documentos de

trabajo. Asimismo, con el formato impreso es difícil llegar al público interesado en conocer los aportes académicos para comprender mejor los problemas y necesidades sociales, políticas y culturales de América Latina. Debido a esta dificultad, se aprovechó el surgimiento de la Web para ir organizando la distribución del formato digital de los libros, revistas y documentos de

trabajo producidos por los centros de investigación y docencia de esta red académica.

Hace 10 años surgió la *Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe*<sup>1</sup> de *Clacso* que funciona con software libre *Greenstone*<sup>2</sup>. Este programa fue creado por la *Universidad de Waikato*, en Nueva Zelanda, para formar colecciones de textos digitales con un servicio de buscador avanzado, que realiza la búsqueda en texto completo y metadatos. Para la interfaz gráfica de la biblioteca virtual se ha optado por el software libre *Plone*<sup>3</sup> que permite realizar cambios de formato y contenidos vía web de forma totalmente descentralizada.

---

**“La decisión de trabajar con software libre fue la más apropiada para una red académica regional que contempla distintas realidades económicas y culturales en cada país”**

---

Hoy sabemos que el desarrollo de una biblioteca virtual cooperativa es el camino más adecuado para acelerar y ampliar la visibilidad e impacto de los resultados de las investigaciones. Por un lado, fue una forma de apoyo al desarrollo de nuevos conocimientos y a las actividades de estudio o docencia y, por el otro, una herramienta para la sociedad que utiliza cada vez más la web en busca de respuestas a las preguntas que plantea su quehacer profesional. A su vez, la decisión de trabajar con software libre fue definitivamente la más apropiada para una red académica regional ya que contempla las distintas realidades económicas y culturales de cada país de América Latina y el Caribe.

## **Trabajando con Greenstone**

A partir de la decisión de difundir virtualmente la producción académica, se comenzó a pensar en la elección de la plataforma más adecuada. Teniendo en cuenta las políticas institucionales, los costos, y la necesidad de descentralizar los procesos a medio plazo, se buscó el software libre que mejor se adaptara a estas condiciones y que organizara de forma eficiente las publicaciones digitales de la red.

Por esta razón se eligió *Greenstone* para desarrollar íntegramente la sala de lectura de la biblioteca virtual. Esta sala ofrece 9.000 textos completos de acceso libre: artículos de revistas académicas, libros, documentos de trabajo, ponencias en congresos, tesis y el año próximo incluirá también los contenidos de tres bases de datos, originalmente desarrolladas en software libre *ISIS*: referencias bibliográficas, proyectos de investigación e investigadores de *Clacso*. *Greenstone* es un conjunto de programas diseñado para crear y difundir colecciones digitales, de código abierto, por lo que permite modificar tanto su apariencia como sus funciones, adaptándolo a las necesidades concretas.

---

**“Greenstone es un conjunto de programas en código abierto para crear y difundir colecciones digitales”**

---

Este conjunto de programas funciona tanto en forma local como en servidores web y se pueden realizar CDs con el motor de búsqueda incorporado; se presenta en diferentes versiones: *Windows*, *Linux*, *Solaris* y *Mac*, y ocupa poco espacio. La información se organiza en colecciones y tanto éstas como el resultado de las búsquedas se definen con una estructura propia. Además indexa

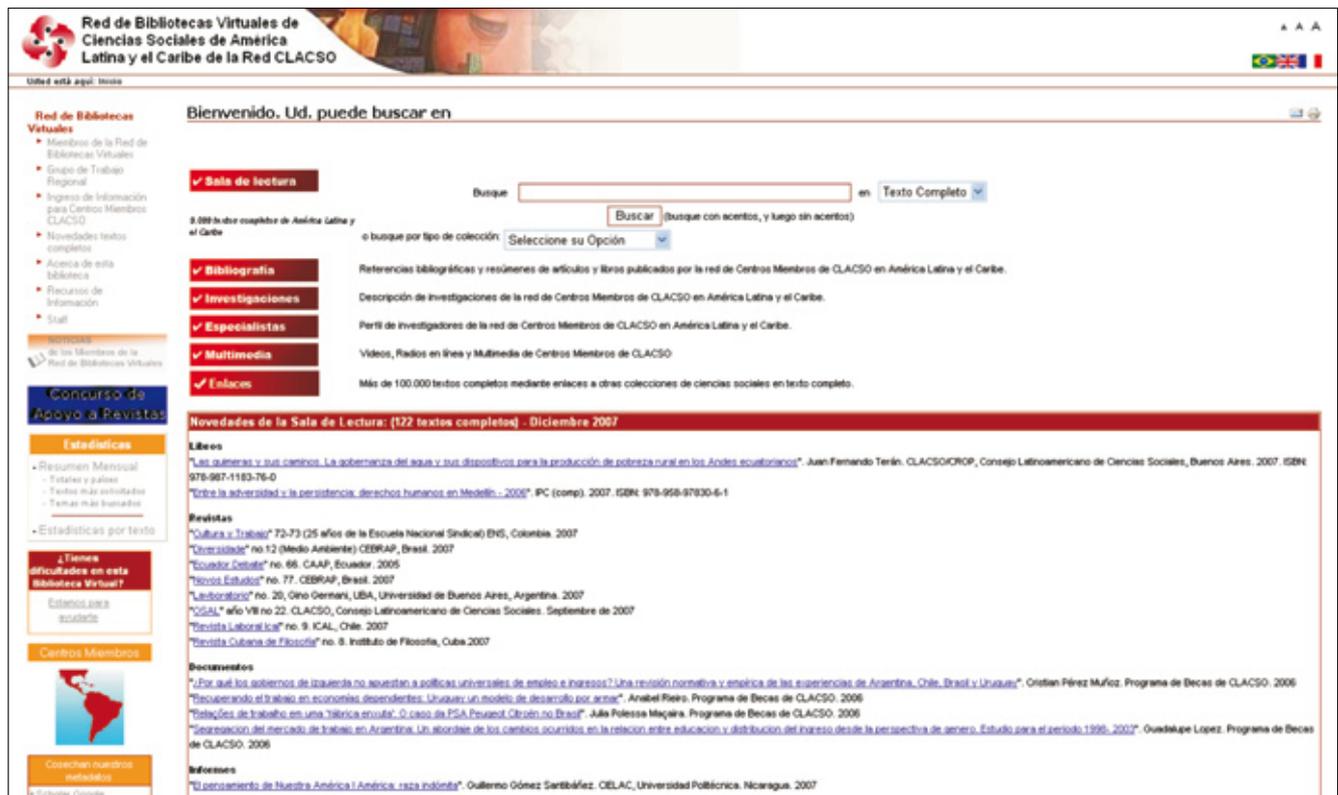
archivos de diferentes formatos: texto, vídeos, bases de datos, música, fotos, páginas web y e-mails. A su vez, trabaja con metadatos, que pueden ser incorporados manual o automáticamente y son éstos los que le permiten elaborar listados para navegar por la información en cada colección, con la ventaja de que procesa toda la información en xml<sup>4</sup>. *Greenstone* tiene incorporado el sistema *OAI (Open Archives Initiative)* para transferencia de datos desde y hacia colecciones remotas para formar grandes redes de información de textos digitales.

Contiene una interfaz del bibliotecario, diseñada para definir y procesar colecciones en forma local en un ordenador. En nuestro caso utilizamos directamente líneas de comando por consola en un servidor *Linux*, por lo que se llevan a cabo dos procesos, generalmente realizados por distintas personas, ya que primero se trata el texto y luego se agrega a una colección ya en línea.

## **Estructura de las colecciones**

La *Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe* está pensada con un doble objetivo: por un lado es un servicio a los usuarios de la web que buscan textos de ciencias sociales y a la vez es un servicio a los centros miembros de *Clacso* que buscan una mayor visibilidad y citaciones de sus publicaciones digitales. En *Greenstone* se crean colecciones para cada institución y para cada tipo de documento.

Desde la página principal de la biblioteca virtual el usuario tiene tres posibilidades para realizar su búsqueda: en toda la sala de lectura, por tipo de documento o por la colección de un centro de investigación en particular. Cada colección posee su propia página principal donde se encuentra la información general, el logo institucional y los datos parti-



Web de la Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe de Clacso

culares para contactar. Además, se ofrece una explicación de las formas de buscar y las condiciones de uso establecidas por la licencia *Creative Commons*<sup>5</sup> que protege los contenidos de la biblioteca virtual.

La búsqueda se puede hacer de dos formas:

- por metadatos (autor, título y tema);
- por un concepto en el texto completo de todas las publicaciones y dentro de cada colección, en listados por autor y por título.

Desde ambas se puede acceder al texto completo.

En esta biblioteca la información se procesa en html, lo que permite cortar el texto en capítulos y realizar así una catalogación analítica de los libros y revistas. A cada parte se le consignan sus propios metadatos, que se utilizan a su vez para crear automáticamente la cita bibliográfica.

La razón principal por la cual procesamos cada artículo de revista

y cada capítulo de libro es que al realizar una consulta se recuperan los distintos capítulos por separado y se evita así que los usuarios se vean obligados a bajar el extenso archivo de un libro o una revista completa. De esta forma se simplifica el acceso a la información ya que la mayoría de los usuarios de América Latina y el Caribe no cuentan con banda ancha y por lo general están interesados sólo en uno de los artículos o capítulos del texto. Además con html es sencillo insertar la portada, las imágenes y cuadros que el texto posea. La finalidad de trabajar así e incluir en el encabezado de cada capítulo el enlace al texto en su formato original (pdf, rtf, doc, etc.), es poder ofrecer a los usuarios la versión en pantalla y la posibilidad de que puedan guardarse una copia para imprimir.

Las colecciones funcionan de forma independiente pero pueden enlazarse unas con otras utilizando la opción que en *Greenstone* se llama *super collection*. De esta forma se puede hacer una búsqueda en

todas las colecciones a la vez o en aquellas que se programen. Por tipo de documento realizamos diferentes súper búsquedas:

- En todas las colecciones.
- Artículos de revistas académicas.
- Libros y documentos de trabajo.
- Ponencias de congresos.
- Tesis.

Se facilita así la consulta en tiempo y forma, porque el usuario puede realizar una sola búsqueda sobre las 190 colecciones con las que contamos en este momento, en lugar de recorrerlas una por una para encontrar la información que necesita.

### Transferencia de tecnología

La gran ventaja de trabajar con colecciones es que la información puede estar alojada en un servidor y ser ofrecida en forma de enlace desde cualquier página. Esto es im-

portante para la red de bibliotecas virtuales de *Clacso* porque no todas las instituciones que la conforman cuentan con servidores propios para instalar esta plataforma, ni con el personal capacitado para desarrollarla. Entonces, las colecciones de los centros cooperantes, que estén alojadas en los servidores de la *Oficina coordinadora de la biblioteca virtual cooperativa*, pueden ser ofrecidas por los centros como enlaces desde sus webs institucionales. De esta forma el centro puede dar acceso a toda su producción académica ordenada dentro de este poderoso buscador como un servicio más.

Sin embargo, algunos centros ya han comenzado a instalar su propia biblioteca digital utilizando *Greenstone*. Para poner en práctica la estructura de cada colección se les proporciona el archivo de configuración, que se puede copiar exacto o se le pueden hacer los cambios que se consideren necesarios. Para modificarlo se trabaja sobre el archivo de configuración de cada colección con cualquier editor de texto plano. En este sentido también se realizan capacitaciones con manuales en castellano, se brinda apoyo vía correo electrónico y chat con los centros miembros.

### **OAI**

Los centros miembros que tienen otras bases de datos en texto completo pueden incorporarse a la red de bibliotecas virtuales por medio de *OAI*, un protocolo de intercambio que permite compartir la información independientemente de la plataforma que cada centro tenga. Se realizaron pruebas con la *University of Texas - Latin American Network Center (UT-Lanic)* y en este momento estamos experimentando con la *Universidad Autónoma del Estado de México*, que tiene un software propio. Este proceso se realiza únicamente a nivel de metadatos, e incluimos dentro

de ellos la dirección web del documento original. De esta manera los portales se enriquecen, ofreciendo a sus usuarios una mayor variedad y cantidad de información proveniente de toda la red.

Cuando se utilizan programas libres se deben considerar los costos de adaptación a las necesidades locales y de capacitar a quienes deben utilizarlo para actualizar las colecciones. La ventaja de que todos los miembros de la red utilicemos la misma plataforma es que de esta forma se pueden realizar transferencias e intercambios de los conocimientos adquiridos por cada uno.

### **PI05**

Desde la secretaría ejecutiva de *Clacso* se está impulsando un proceso de descentralización del trabajo hacia los centros de investigación de ciencias sociales que conforman la red. Con este objetivo, en la *Oficina coordinadora de la biblioteca virtual cooperativa* comenzamos a desarrollar distintas herramientas para facilitar la incorporación de la producción de cada uno de los centros miembros a la sala de lectura de la biblioteca.

En este marco se desarrolló una interfaz gráfica de usuario llamada *PI05*. Fue específicamente pensada como un sistema para facilitar la preparación de los documentos y el trabajo necesario para generar los tres archivos requeridos por *Greenstone* cuando se procesan los textos en formato html: html, xml y jpg. Éstos son necesarios para la inclusión de documentos de trabajo, publicaciones periódicas o libros de cada uno de los centros miembros de *Clacso* en sus respectivas colecciones, para ser luego presentados en la sala de lectura.

Aunque la interfaz se encuentra aún en etapa beta funcional (sujeta a mejoras futuras), se logró agilizar la técnica de trabajo, ya que hizo más simple tanto la metodología

para ingresar metadatos en cada parte del documento, como la tarea de darle forma con etiquetas html. Esto significó un importante avance en la descentralización del trabajo de incorporación de libros, documentos y artículos.

*PI05* se puede bajar libremente, junto con su manual de uso, de la sección “Recursos de información”<sup>7</sup> de la biblioteca. Esta interfaz ofrece diferentes esquemas de metadatos predefinidos para cada tipo de publicación que pueden ser adaptados a distintas necesidades, tanto para publicaciones específicas como para utilizarlos en otros servicios. Así se simplifica enormemente la inclusión de metadatos.

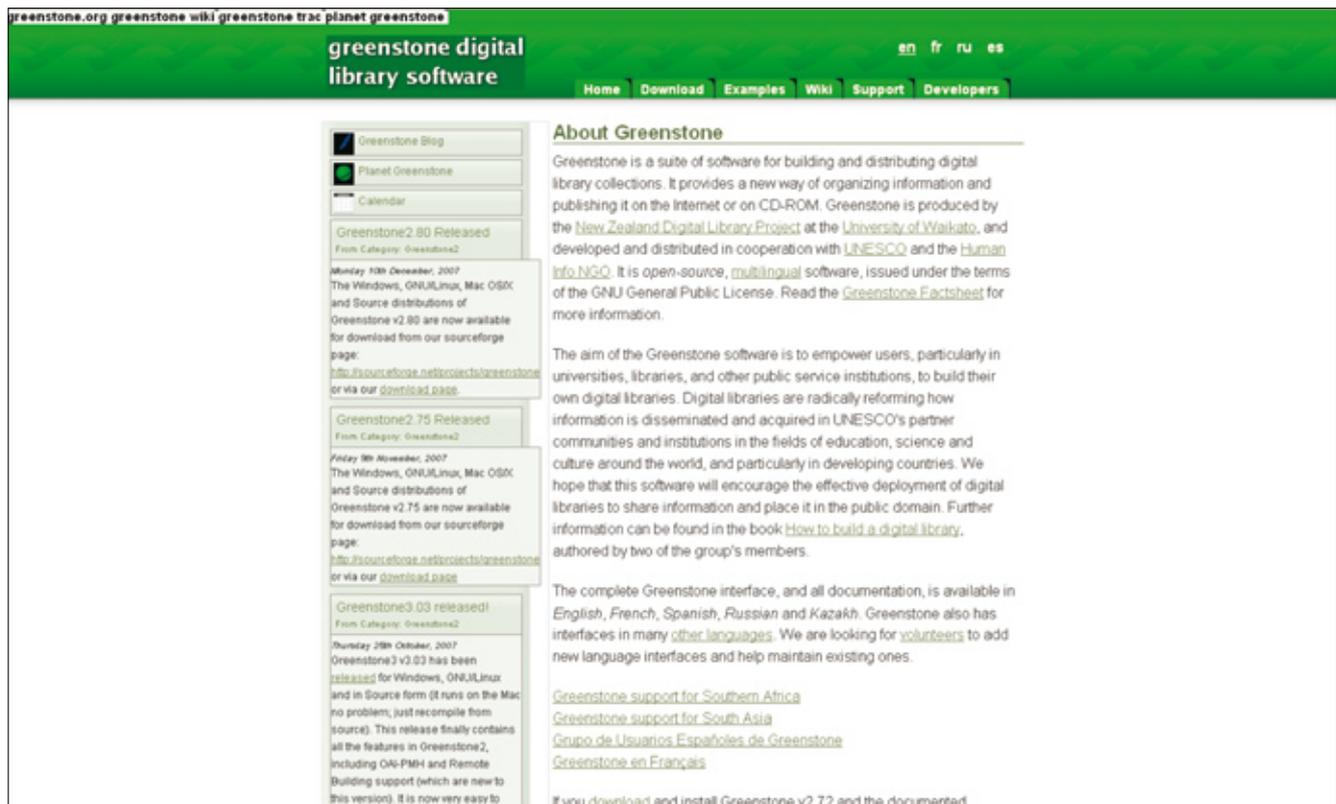
---

**“PI05 permitió que cada vez sean más los centros cooperantes que trabajan a distancia preparando sus propios documentos para *Greenstone*”**

---

Esta interfaz gráfica presenta dos grandes partes. En la primera se completan los datos generales de la publicación, tanto sus metadatos como la presentación general del documento; y en la segunda se divide el texto en secciones y se trabaja de forma analítica cada capítulo o contribución del documento general. Esta segunda parte está pensada para los libros colaborativos, que tienen capítulos de distintos autores, así como para las publicaciones periódicas académicas, que generalmente tienen división de artículos por autores o secciones especiales que engloban distintos documentos.

La finalidad de este proceso de división en secciones es que se genere automáticamente en *Greenstone* un índice interactivo del texto con sus distintos capítulos. De esta



Web de Greenstone

manera, al realizar una búsqueda el resultado serán capítulos específicos sobre un tema, un autor o un título, y no todo el volumen completo. Se trabaja en dos planos, por un lado se incluyen los metadatos específicos de la sección y por el otro, se gestiona el estilo del documento que ha sido perdido al guardarlo en texto plano.

Para tratar los documentos con *PI05* no hace falta conocer en profundidad html. En lugar de trabajar con el propio código, *PI05* cuenta con una barra de herramientas, similar a la de un procesador de texto, que facilita la inclusión de enlaces, la activación de direcciones electrónicas, la alineación y marcas del texto, etc. Así, se logra que la corrección o modificación del estilo en html sea una tarea cómoda y familiar para cualquier usuario. Teniendo en cuenta que fue diseñado sobre la base de una larga experiencia, esta herramienta va creando, a medida que se delimitan las secciones, un “árbol” a la izquierda del área de trabajo que sirve de guía

para saber qué parte del texto se ha trabajado.

Como resultado, *PI05* transforma los documentos y cada una de sus partes en un archivo único de extensión xml donde reúne los elementos necesarios para su procesamiento en *Greenstone*:

- Un archivo html, que contiene el documento completo con los metadatos de cada sección.
- Un archivo xml con los metadatos del documento general.
- Un archivo de imagen (la portada del documento).

Es decir, en el marco del proceso de descentralización del trabajo para la incorporación de libros, documentos y artículos en la red de bibliotecas virtuales, *PI05* permitió que todo el material pueda ser enviado en un solo archivo xml.

Finalmente, esta herramienta nos permite ampliar el servicio de textos completos de la red, brindando acceso a mayor cantidad de documentos cada mes porque cada

vez son más los centros cooperantes que trabajan a distancia preparando sus propios documentos para *Greenstone*.

### Trabajo con metadatos

Al detenernos a reflexionar sobre la aplicación práctica que *Greenstone* significa para esta biblioteca virtual cooperativa, resulta imprescindible pensar en el trabajo vinculado con los metadatos temáticos. La alternativa que brinda esta biblioteca cooperativa para el procesamiento y publicación de textos permite acceder a los últimos desarrollos en temas vinculados a las ciencias sociales que son generados por los centros miembros de *Clacso*.

Por tratarse de una biblioteca especializada en ciencias sociales, se ofrece el valor agregado de la clasificación humana, que permite al usuario realizar una búsqueda temática bien precisa. Por otra parte, para quienes analizan los textos, esta clasificación implica un trabajo creativo e intelectual de permanen-

te especialización y perfeccionamiento.

Existen diferentes formas de localizar la información dentro de una biblioteca. La búsqueda en texto completo es el trabajo que hace automáticamente el programa; mientras que las consultas por título, autor, tema, es decir mediante los metadatos, requieren indización manual.

---

**“La aplicación práctica de Greenstone hace que pensar en el trabajo con metadatos temáticos sea imprescindible”**

---

Uno de los medios de acceso a la publicación es a través de una búsqueda por tema o temas, ya que a partir de las palabras asignadas a cada texto se agruparán los documentos por temáticas.

Las palabras clave o descriptores asignados para cada texto provienen de un proceso de reflexión interdisciplinario cuya premisa es la selección de hasta diez descriptores como máximo para cada documento. La tarea, que se realiza desde la *Oficina coordinadora de la biblioteca cooperativa*, junto con los centros cooperantes, intenta establecer conceptos comunes o generales que sean comprensibles para toda la región. Esta actividad se encuentra ligada a la utilización de un lenguaje específico y controlado, recurriendo en este caso al *Macrothesaurus* de la *OCDE (Macro)*<sup>8</sup>. Éste es nuestro sustento de trabajo ya que es el único material de clasificación específico de ciencias sociales que incluye descriptores de diversas disciplinas del área.

Frente al constante desarrollo de las diferentes disciplinas de ciencias sociales, y debido a que la última actualización del *Macrothe-*

*sauro* fue realizada en 1991, surgió la necesidad de enriquecer la terminología mediante la creación de un listado de descriptores libres<sup>9</sup>. Es utilizado como complemento al proceso de asignación de descriptores, es decir, se consulta cuando el *Macro* no posee descriptores puntuales o términos modernos para el tema tratado.

Cuando el texto que se analiza hace referencia a un lugar geográfico en particular se asigna la región, el país, la provincia y la ciudad (ejemplo: “América Latina, Bolivia, La Paz”) porque, como se trata de una biblioteca virtual, el usuario puede acceder desde cualquier parte del mundo y no conocer las ubicaciones específicas de América Latina y Caribe.

Como muestra veremos los descriptores generales asignados para el texto *Los tormentos de la materia. Aportes para una ecología política latinoamericana* de **Héctor Alimonda** (compilador): “Ecología política, Gestión ambiental, Política ambiental, Recursos hídricos, Canal de Panamá, Medio ambiente, Desarrollo sustentable, América Latina, Caribe”.

Asimismo podemos apreciar la diversidad temática en el interior de dos de sus capítulos y los descriptores temáticos analíticos propuestos para cada caso:

– ”III. Teoría crítica y crítica política en la cuestión ambiental: problemas y perspectivas de **Enoch Adames-Mayorga**: Ecología política, Medio ambiente, Ciencia y tecnología, Naturaleza, Intelectuales, Teoría política, América Latina, Caribe.

– X. Con la soja al cuello: crónica de un país hambriento productor de divisas de **Domínguez, Diego; Sabatino, Pablo**: Ecología política, Medio ambiente, Política agraria, Economía, Recursos naturales, Nueva tecnología, Soja, Transgénicos, América Latina, Argentina”.

Los usuarios de esta biblioteca virtual cooperativa utilizan el servicio de búsqueda de información a través de internet. Por tal motivo, la respuesta que obtengan en su búsqueda estará ligada en gran parte a la calidad de la clasificación que se logre en cada texto. Es por esto que la asignación de metadatos<sup>10</sup> se realiza analíticamente, es decir, cada capítulo o parte del texto tiene sus propios metadatos.

La mirada interdisciplinaria para la asignación de descriptores enriquece la tarea y tiende a enfatizar las distintas aristas desde las que un fenómeno social puede ser analizado.

### **Conclusión**

Trabajar con software libre fue una decisión sencilla en la *Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe* porque *Clacso* es una red académica comprometida con el uso de plataformas libres y el acceso a los textos digitales de las investigaciones que se realizan en los centros miembros de la red es abierto.

---

**“Clacso es una red académica comprometida con el uso de plataformas libres y el acceso abierto a los textos digitales”**

---

Esta filosofía es compartida por las agencias que apoyan financieramente los programas de la red *Clacso*: principalmente *ASDI* (gobierno sueco), *NORAD* (gobierno noruego), *IDRC* (gobierno de Canadá) e *Inasp (International Network for the Availability of Scientific Publications)*.

Cuando esta biblioteca virtual cooperativa comenzó, los centros miembros entendían muy poco de

estas nuevas formas de producir y difundir textos en formato digital. Por eso *Clacso* organizó actividades de promoción y capacitación del personal de las áreas de publicaciones, biblioteca e informática para que colaboraran en los nuevos sistemas de trabajo. Diez años después, el eje se ha invertido y este servicio se encuentra desbordado para atender la demanda de los centros que quieren difundir rápidamente la versión digital de las publicaciones, siendo cada vez más habitual que ésta sea la única versión difundida. Esto requiere una detallada revisión de procedimientos y programas utilizados para acelerar y simplificar los procesos. En esta tarea estamos trabajando actualmente.

### Notas

1. <http://www.clacso.org.ar/biblioteca>
2. <http://www.greenstone.org>
3. <http://www.plone.org>

4. Xml: eXtensible Markup Language.
5. <http://creativecommons.org>
6. <http://www.vim.org>
7. <http://sala.clacso.org.ar/biblioteca/pi05/pi05>
8. <http://bibliotecavirtual.clacso.org.ar/ar/oeecd-macroth>
9. <http://sala.clacso.org.ar/biblioteca/Members/claves/desclib/desliba>
10. Guía de asignación de descriptores temáticos. Clasificación Temática o Indización: Humana: [http://bibliotecavirtual.clacso.org.ar/ar/biblioindra/documentacion/analisis\\_documental.pdf](http://bibliotecavirtual.clacso.org.ar/ar/biblioindra/documentacion/analisis_documental.pdf)

### Bibliografía

**Babini, Dominique; Fraga, Jorge** (comp.). "Edición electrónica, bibliotecas virtuales y portales para las ciencias sociales en América Latina y El Caribe". Buenos Aires: Consejo Latinoamericano de Ciencias Sociales, 2006, pp. 271. ISBN 987-1183-53-4. <http://bibliotecavirtual.clacso.org.ar/ar/libros/secret/babini/babini.html>

**Babini, Dominique.** "Producción, difusión y uso de investigación colaborativa en red: el caso de la red académica del Consejo Latinoamericano de Ciencias Sociales". En: **Albornoz, Mario; Alfaraz, Claudio** (eds.). *Redes de conocimiento: construcción, dinámica y gestión*. Buenos Aires: Red Iberoamericana de Indicadores de Ciencia y Tecnología (Ricyt), 2006, pp. 249-258.

<http://www2.ricyt.org/docs/libroRC2006/libroRC2006.pdf>

**Garrido-Picazo, Piedad; Tramullas, Jesús.** "Un experimento de creación de biblioteca digital con Greenstone". En: *El profesional de la información*, 2004, marzo-abril v. 13, n. 2, pp. 84-92. <http://eprints.rclis.org/archive/00008299/>

**Jose, Sanjo.** "Adoption of open source digital library software packages: a survey". En: **Kumar, Manoj K.**, eds. *Proceedings Caliber 2007: 5th International convention on automation of libraries in education and research institutions*. Febrero 2007, pp. 98-102. <http://eprints.rclis.org/archive/00008750/>

**Tramullas, Jesús.** "Software libre para gestión de recursos de información digital". En: *Encuentro internacional sobre conocimiento libre, II Conferencia internacional de software libre*, 2006, pp. 346-356. <http://eprints.rclis.org/archive/00006272/>

**Dominique Babini, Florencia Vergara-Rossi, Flavia Medici y Jessica González,** *Oficina Coordinadora de la Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe, Buenos Aires.* [biblioteca@clacso.edu.ar](mailto:biblioteca@clacso.edu.ar)

**nature.com**  
es ciencia

**nature.com**  
es innovación



Todo está en nuestro **nature.com**

Conéctate al acceso a nature.com con una suscripción de Acceso y elige entre las revistas de Investigación Científica y Médica de más alta calidad. Descarga y compartición de documentos. Busca y evalúa el contenido online por palabras clave. nature.com ofrece una nueva aplicación imprescindible en Web 2.0: el sitio que encontrarás.

- Sigue como en el centro Nature Now! Blog
- RSS feeds
- Podcast
- Videos
- Organiza y comparte tus datos con SciData
- Busca y comparte documentos con CrossRef
- Redes sociales en Nature Network
- Askus Research: un foro que le permite compartir sus trabajos de investigación antes de su publicación

T: +44 (0)20 7845 1700 | E: [institutions@nature.com](mailto:institutions@nature.com) | W: [www.nature.com/institutes](http://www.nature.com/institutes)

# Laboratorio virtual de software libre para bibliotecas

Por Patricia Russo-Gallo, Josep-Manel Rodríguez-Gairín y Andreu Sulé-Duesa

**Resumen:** La elección de un programa de gestión de bibliotecas se ve afectada muchas veces por una serie de condiciones sociales, económicas y políticas que hacen que la elección no sea la más adecuada para las necesidades, características y funciones de la biblioteca. El software libre está siendo una de las soluciones más optadas, por sus libertades de copia, modificación y distribución, además de la libertad de licencias y las posibilidades de interoperación con otras aplicaciones. Esta nueva tendencia hacia el software libre en bibliotecas se refleja también en los estudios de biblioteconomía y documentación, en los que desde diferentes asignaturas se aportan conocimientos sobre programas de automatización, de gestión de repositorios, incluso del sistema operativo Linux/GNU, entre otros. Esta combinación entre las necesida-



**Andreu Sulé-Duesa** es profesor de la Facultat de Biblioteconomia i Documentació de la Universitat de Barcelona en donde imparte asignaturas relacionadas con los sistemas de automatización. Profesor del curso virtual del Cobdc, Introducció al format Marc 21.

**Patricia Russo-Gallo** es coordinadora del Grupo de Trabajo de Software Libre para Profesionales de la Información (Cobdc). Trabaja como arquitecta de la información en una empresa de experiencia de marcas y diseño web (Abilbo) en Barcelona.

**Josep-Manel Rodríguez-Gairín** es profesor de la Facultat de Biblioteconomia i Documentació de la Universitat de Barcelona donde coordina también el Màster en gestió de continguts digitals, y es profesor del Master online de documentación digital del IDEC, Univ. Pompeu Fabra.

des de los centros y la tendencia al uso de software libre, es la que un grupo de profesores de la Facultat de Biblioteconomia i Documentació (Universitat de Barcelona) y estudiantes, miembros del Grup de Treball sobre Programari Lliure per als Professionals de la Informació (Cobdc), han querido aportar a la comunidad profesional, creando un laboratorio virtual para el uso de software libre de aplicación en bibliotecas.

**Palabras clave:** Software libre, Automatización de bibliotecas, Gestión documental, Tesoros, Depósitos digitales, Repositorios, Gestión de bibliografías, Gestión de recursos electrónicos, Educación virtual, Laboratorios virtuales.

**Title: Virtual laboratory of free software for libraries**

**Abstract:** The selection of a library management system is often affected by social, economic and political conditions that may result in an inappropriate choice for a library's needs, characteristics and functions. Free software, or freeware, is one of the more often selected solutions, given the freedom to copy, modify and distribute it in addition to free licensing and the possibilities for integration with other applications. This trend is reflected in librarianship curricula, in which automation software, repository management software, and even Linux/GNU, among others, are explained in a variety of courses. This combination of organizational needs and freeware trends led a group of professors from the Faculty of Library and Information Science (UB) and members of the Work Group on Free Software for Information Professionals (Cobdc) to create a virtual laboratory for the use of free software for library applications, as a contribution to the professional community.

**Keywords:** Free software, Freeware, Library automation, Document management, Thesaurus, Digital deposit, Repositories, Bibliography management, Electronic resources management, E-learning, Virtual lab.

**Russo-Gallo, Patricia; Rodríguez-Gairín, Josep-Manel; Sulé-Duesa, Andreu.** "Laboratorio virtual de software libre para bibliotecas". En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, pp. 71-77.

DOI: 10.3145/epi.2008.ene.08

## Introducción

CON MOTIVO DEL FES-INTERNET 2007, miembros del Grup de treball sobre programari lliure per als professionals de la informació del Col·legi Oficial de

Bibliotecaris-Documentalistes de Catalunya y de la Facultat de Biblioteconomia i Documentació de la Universitat de Barcelona decidimos crear un laboratorio virtual de software libre orientado a la au-

tomatización de unidades de información (principalmente bibliotecas y centros de documentación).

<http://www.fesinternet.net/>  
<http://www.soft-libre.net/>  
<http://www.ub.edu/biblio>

Web del laboratorio virtual

El objetivo era dar a conocer, tanto entre los alumnos de la *Facultat* como entre los profesionales del sector, diferentes soluciones gratuitas de automatización. Y como la mejor manera de conocer un programa es utilizándolo, además de las presentaciones de los diversos productos, centramos nuestros esfuerzos en la creación de un entorno virtual que permitiera a los interesados su testado directo, sin intermediarios. Durante los días que el laboratorio estuvo activo (del 15 de mayo al 8 de junio de 2007), los alumnos y los profesionales que accedieron a él pudieron gestionar los programas con todos los privilegios, administrando el sistema, creando cuentas de usuario, editando registros, creando estadísticas, importando registros, etc.

Para conseguir este objetivo se instalaron en un servidor (al que se le denominó *iGNUcius*, en honor del personaje que, a manera de broma, interpreta **Richard Stallman**) los programas siguientes:

- *PMB* y *Openbiblio* (automatización de bibliotecas).
- *Potnia* (gestión de recursos electrónicos).

- *Alfresco* (gestión de documentos electrónicos).
- *Dspace* y *E-prints* (gestión de repositorios).
- *Tematres* (gestión de tesauros).
- *Rebase* (gestión de bibliografías).

[http://es.wikipedia.org/wiki/San\\_Ignacio](http://es.wikipedia.org/wiki/San_Ignacio)

**“Centramos nuestros esfuerzos en la creación de un entorno virtual que permitiera a los interesados experimentar en directo”**

Además del acceso a todas sus funciones, en la misma interfaz web se ofrecía una breve descripción de las principales características de cada producto, acceso a su web oficial y a los manuales de instalación y uso.

## Programas expuestos

A continuación se describen las principales funciones y características de todos ellos.

### 1. *PMB*

Es un sistema de automatización de bibliotecas creado en Francia en 2002 bajo licencia *CeCILL*, una licencia de software libre del gobierno francés que recoge las mismas condiciones y libertades de la que *GPL*.

<http://www.sigb.net>

Está destinado a bibliotecas medianas y grandes, permitiendo también la gestión de una red de bibliotecas integradas en un catálogo colectivo. El formato de catalogación que utiliza es Unimarc pero incorpora un sistema de conversión e importación de formatos Usmarc y xml entre otros.

Lo podemos encontrar traducido a varios idiomas, entre ellos francés, inglés, español y catalán. Está disponible la versión 3 desde septiembre de 2006.

a. Requisitos técnicos mínimos de la máquina y programas necesarios para la instalación:

- Procesador: *Intel Pentium 1 GHz*.
- Memoria: 256 MB.
- Sistema operativo: *GNU-Linux, Windows 2000/XP, MacOS*, etc.
- Lenguaje de programación: *PHP v. 4.2*.
- Base de datos: *MySQL v. 4.0*.
- Navegador: *Mozilla Firefox, IE 6.0, Netscape 7*, etc.

#### b. Características.

- Gestión de autoridades (por autores, editores, colecciones).
- Gestión de un tesoro de materias.
- Uso del protocolo Z39.50.
- Automatización del servicio de DSI.
- Control de publicaciones en serie.
- Gestión de vaciado de revistas.
- Sistema de copias de seguridad.

#### c. Módulos.

- Administración del sistema, desde donde se puede parametrizar toda la configuración general, gestionar a los usuarios, realizar las copias de seguridad, etc.
- Catalogación, para monografías, publicaciones en serie, captura de registros desde el catálogo Z39.50, creación de listas de registros (de novedades, búsquedas guardadas, etc.).
- Préstamo.
- Autoridades para autores, editores, colecciones y clasificación. Incluye la creación de un tesoro para materias.
- DSI.
- Opac.

## 2. Openbiblio

Es un sistema integrado de gestión de bibliotecas. Creado en 2002 en inglés y bajo licencia *GNU-GPL*.

Está disponible en diversas lenguas, entre ellas español y catalán.

<http://obiblio.sourceforge.net>

Está destinado a pequeñas bibliotecas, sobre todo escolares. El formato de catalogación es *Unimarc* pero permite la importación de registros en otros compatibles. Está disponible la versión 0.6.0 desde febrero de 2007.

#### a. Requisitos.

- Sistema operativo: *GNU-Linux, Windows 2000/XP*.
- Lenguaje de programación: *PHP v. 4.2*.
- Base de datos: *MySQL v. 4.0*.
- Navegador: *Mozilla Firefox, IE 6.0, Netscape 7*, etc.

#### b. Características.

- Avisos a los usuarios a través de mensajes prediseñados en el sistema.
- Control de multas por retraso en la devolución de documentos.
- Diversos modelos de estadísticas de uso de la biblioteca y el material.

#### c. Módulos.

- Préstamo. Administración de datos de los usuarios, sus préstamos, devoluciones, reservas y multas.
- Catalogación. Control de registros bibliográficos y las respectivas copias.
- Administración. Configuración y gestión del sistema, incluidas las autorizaciones del personal administrador del sistema.
- Informes. Gestión de estadísticas, informes y mensajes a los usuarios.
- Opac.

## 3. Potnia

Es un programa para la creación de directorios temáticos pensado para la gestión de recursos científicos, revistas, publicaciones, etc. Creado en 2003 bajo un proyecto

de la *Dirección General de Universidades del Ministerio de Educación, Cultura y Deportes* (España). Se rige por la *Mozilla public license 1.1 (MPL 1.1)*. Originalmente en español, no se conocen otras traducciones.

<http://potnia.sourceforge.net>

La estructura y la descripción de los registros siguen un estándar definido por el conjunto de metadatos Dublin Core. Se encuentra disponible la versión 1.0 desde octubre de 2003.

#### a. Requisitos.

- Servidor: *Apache*.
- Sistema operativo: *GNU-Linux, Windows 2000/XP*.
- Lenguaje de programación: *PHP v. 4.2*.
- Base de datos: *MySQL v. 4.0*.
- Navegador: *Mozilla Firefox, IE 6.0*, etc.

#### b. Funciones.

- Búsqueda de registros. Además de la consulta básica también se incluye una avanzada que permite concretar los campos donde se quiere buscar (título, palabra clave, descripción o tema) y la combinación de éstos a través de los operadores booleanos *and*, *or* y *not*.
- Administración del sistema. Permite añadir, modificar o borrar registros de la base de datos y también gestionar la lista de temas.
- Página de inicio. Ofrece la lista de temas de los registros de la base de datos; una vez dentro de uno de los temas se presenta la lista de los registros categorizados bajo el mismo, desde donde se podrá acceder al detalle.

## 4. Alfresco

Se define como un gestor de contenidos empresariales (*enterprise content management, ECM*) y entre sus aplicaciones está la gestión de documentos. Creado en

2005 por un equipo de miembros de *Documentum* e *Interwoven*. Originalmente en inglés, no se conoce una versión oficial en español.

<http://www.alfresco.com>

Es un software de código abierto. El acceso al código fuente permite un alto nivel de calidad en el desarrollo y evolución del sistema. El beneficio comercial se halla en el empaquetado junto con los manuales, formación y mantenimiento del programa.

Dispone de un espacio dedicado a las transformaciones que la comunidad de usuarios realiza sobre el programa, las cuales están bajo una licencia *Alfresco public license*. Disponible la versión 2.1 desde junio de 2007.

<http://forge.alfresco.com/>

#### a. Requisitos.

- Servidor: *Apache Tomcat*.
- Sistema operativo: *GNU-Linux*, *Windows 2000/XP*.
- Lenguaje de programación: *PHP* v. 4.2.
- Base de datos: *MySQL* v. 4.0.
- Navegador: *Mozilla Firefox*, *IE 6.0*, etc.

#### b. Características.

- Depósito de documentos electrónicos en su formato original (doc, xls, tif, html, jpg, pdf, vídeo, audio, cad, etc.).
- Visualizadores de documentos.
- Agrupaciones por carpetas y cuadro de clasificación.
- Sistema de recuperación de la información, búsqueda básica y avanzada, busca ficheros, carpetas y contenidos.
- Gestión y control de flujos de trabajo.
- Gestión de usuarios y niveles de acceso.
- Histórico de versiones.

#### c. Funciones.

- Espacio común desde donde clasificar la documentación por carpetas.
- Permite añadir documentos de la red local o crearlos nuevos desde el mismo sistema (en formato html, txt o xml).
- Gestiona reglas de contenidos con las que transformar formatos de forma automática, crear flujos de trabajo, etc.
- Espacios de foros para las carpetas o documentos concretos.
- Edición y modificación de los metadatos de cada documento, desde donde también ver el histórico de las versiones creadas.
- Creación de copias de trabajo.
- Gestión de los perfiles de usuarios del sistema y sus privilegios y autorizaciones.

### 5. DSpace

Es un sistema de gestión de depósitos institucionales (repositorios) creado por la *Biblioteca del MIT (Massachusetts Institute of Technology)* y *Hewlett-Packard* en 2002. Se distribuye bajo *BSD Open source licenses (Berkeley Software Distribution)*, que se encuentra dentro de la familia de las consideradas como software libre, pero con algunas restricciones en comparación con la licencia *GNU-GPL*. Permite el uso y modificación del código fuente y el uso comercial del producto.

<http://www.dspace.org>

Es posible la gestión de diferentes tipos de materiales (*preprints*, informes técnicos, imágenes, vídeos, etc.) y contribuye a la preservación digital de la producción científica de una organización. Lo podemos encontrar traducido a diversos idiomas, entre ellos el español, catalán y gallego.

Como sistema de preservación de material digital, a todos los re-

gistros se les asigna un identificador persistente. *Dspace* trabaja, como *Handles*, con el *Protocol for Metadata Harvesting (PMH)* de la *Open Archive Initiative*. Está disponible la versión 1.4.2 desde mayo de 2007.

#### a. Requisitos.

- Servidor: *Apache Tomcat*.
- Sistema operativo: *Unix*, *GNU-Linux*, *Solaris*.
- Lenguaje de programación: *Java*.
- Base de datos: *postgreSQL*.
- Navegador: *Mozilla Firefox*, *IE 6.0*, etc.

#### b. Características.

- Soporta cualquier tipo de material: documentos, tesis, imágenes, audio, vídeo, etc.
- Los registros quedan guardados en formato Dublin Core.
- Puede interoperar con otros sistemas de la organización.
- Todos los registros tienen un identificador persistente.
- Permite la migración de formatos hacia *IFF*, *SGML*, *xml*, *AIFF*, *pdf*, y, aunque no se asegura, hacia otros propietarios como *Microsoft Word*, *PowerPoint*, *Lotus 1-2-3* o *WordPerfect*.

### 6. E-Prints

Es un sistema de creación de repositorios institucionales de la producción científica, tesis, informes, etc. Creado por la *University of Southampton* (Reino Unido) en 2000 y distribuido con licencia *GNU-GPL*. Originalmente en inglés, podemos encontrar alguna versión traducida al español.

<http://www.eprints.org/>

Utiliza formato de metadatos *xml* y *Dublin Core* basado en la *OAI-PMH*, lo que hace que pueda soportar diferentes tipos de documentos y formatos de datos para su preservación y acceso abierto a los contenidos y su conocimiento. Está

disponible la versión 3 desde enero de 2007.

a. Requisitos.

- Procesador: 1 GHz.
- Memoria: 512 MB.
- Servidor: *Apache*.
- Sistema operativo: *Unix, GNU-Linux*.
- Lenguaje de programación: *PERL*.
- Base de datos: *MySQL*.
- Navegador: *Mozilla Firefox, IE 6.0*, etc.

b. Características.

- Formato de metadatos Dublin Core.
- Repositorio de todo tipo de material: texto, imagen, etc.
- Sindicación de contenidos.
- Incluye un gestor de bibliografías.
- Búsqueda básica y avanzada.

### 7. Tematres

Es una aplicación web para la gestión de tesauros creada en Argentina en 2004 bajo licencia *GPL* y originalmente en español. También está pensada para la creación de estructuras de navegación web, directorios de bibliotecas digitales o control de lenguajes documentales. Disponible la versión 0.9.5 desde mayo de 2007.

<http://www.r020.com.ar/tematres>

a. Requisitos.

- Servidor: *Apache*.
- Sistema operativo: *GNU-Linux, Windows 2000/XP*.
- Lenguaje de programación: *PHP* v. 4.2.
- Base de datos: *MySQL* v. 4.0.
- Navegador: *Mozilla Firefox, IE 6.0*, etc.

b. Características.

- Relación entre los términos por: equivalencia, jerarquía y/o asociación.

- Cantidad ilimitada de términos relacionados, de niveles jerárquicos y de términos no preferentes.

- Múltiples notas de alcance, históricas y bibliográficas por término.

- Control de términos repetidos.

- Navegación sistemática y alfabética.

- Formatos de exportación en xml, Dublin Core, *RDF SKOS-Core, Zthes*, etc.

- Buscador general.

### 8. RefBase

Es un programa de gestión de bibliografías y citas bibliográfica que funciona bajo interfaz web multiusuario. Creado en Alemania en 2002 bajo licencia *GNU-GPL* y originalmente en inglés, no se conoce ninguna versión en otros idiomas.

<http://refbase.sourceforge.net/>

Está pensado para crear colecciones de referencias bibliográficas importadas de diferentes bases de datos académicas y extraerlas en formato de cita para la bibliografía de textos. Está disponible la versión 0.9.0 desde octubre de 2006.

a. Requisitos.

- Servidor: *Apache*.
- Sistema operativo: *GNU-Linux, Windows 2000/XP*.
- Lenguaje de programación: *PHP* v. 4.2.
- Base de datos: *MySQL* v. 4.0.

- Navegador: *Mozilla Firefox, IE 6.0*, etc.

b. Características.

- Búsqueda básica o avanzada.
- Sindicación de contenidos.
- Formato de las citas: html, RTF, pdf, LaTeX.
- Formatos de importación: *BibTeX, Endnote, RIS, RefWorks,*

*PubMed, ISI Web of Science, CSA Illumina, Copac, MODS XML*.

- Formato de exportación: *BibTeX, Endnote, RIS, MODS XML*, Xml de la *Library of Congress* y *OpenDocument*.

### Aspectos técnicos de la puesta en marcha del laboratorio

Fue necesario planificar previamente los requisitos de hardware y software base necesarios para soportar las aplicaciones escogidas. La *Facultat de Biblioteconomia i Documentació* facilitó la adquisición de un servidor exclusivo *DELL PowerEdge* de doble procesador a 3.0 GHz, 2 MB de memoria RAM y dos discos duros de 80 GB.

El sistema operativo escogido fue *Linux* en su distribución *Debian GNU/Linux 4.0 etch* sobre el que se instalaron las aplicaciones básicas requeridas por las distintas aplicaciones: *Perl* y *PHP* como lenguajes de programación, el entorno *Java* de desarrollo, *Apache 2* como servidor web, *TomCat 5.5* como servidor de aplicaciones, y *MySQL 5.0.2* y *PostgreSQL 8.0* como sistemas de gestión de bases de datos. Se utilizó como antivirus *ClamAV* y se implantó un sistema de copias de seguridad.

Se escogió también un sistema de gestión de contenidos para la navegación por el laboratorio que permitiera registro de usuarios y control estadístico sobre el que se elaboraron las fichas descriptivas de cada uno de los programas. El CMS escogido fue *Drupal*.

A nivel técnico podemos diferenciar dos tipos básicos de aplicaciones:

- Las que se distribuyen en forma de *scripts* para ser utilizadas bajo un entorno *Apache-PHP-MySQL*. Su instalación normalmente sólo implica descomprimir los archivos en un directorio web y ejecutar un fichero de configuración

desde el propio navegador que crea la base de datos y tablas necesarias ajustándose a los parámetros. Este tipo de aplicaciones suelen ser las más sencillas de instalar y en este grupo tenemos *PMB*, *OpenBiblio*, *TemaTres* y *Potnia*.

– Las basadas en entorno *Java* que requieren un servidor de aplicaciones como *TomCat* o basadas en *PERL*. Se distribuyen en código fuente y deben ser compiladas y enlazadas en el propio servidor. El proceso no resulta sencillo y puede requerir la instalación de módulos de *PERL* adicionales u otros complementos. En este grupo tenemos *Alfresco*, *DSpace* y *Eprints*.

Compaginar en una misma máquina todas las aplicaciones fue un proceso francamente complicado y tanto la experiencia práctica como los conocimientos compartidos con otros usuarios nos permiten hacer una serie de reflexiones.

Los procesos de instalación requieren unos conocimientos previos informáticos y no están al alcance del público en general, en especial comparando con versiones *Windows* de la misma aplicación (*Alfresco*, *Eprints*).

No resulta nada fácil ajustar compatibilidades entre ellas. Por ejemplo las aplicaciones basadas en *PHP-MySQL* pueden requerir distintos ajustes de configuración de *PHP* (*register globals*) o versiones específicas de *PHP*, ya sea porque

empleen funciones nuevas, solamente presentes a partir de *PHP5*, o bien otras obsoletas no compatibles con éste y que requieren por tanto *PHP4*. Este mismo problema se nos planteaba entre *Alfresco* (que se distribuye con una versión de *TomCat* concreta) y *DSpace* que requiere una preinstalación del mismo.

En el caso de *DSpace*, que podía utilizar distintos gestores de bases de datos como *MySQL*, *Oracle* o *PostgreSQL*, aunque parecía lógico emplear *MySQL* (ya que era el que estábamos utilizando en el resto) los comentarios en listas de distribución y la complejidad de ajustes nos decantaron a usar *PostgreSQL* que es el gestor propuesto por defecto por los diseñadores de la aplicación.

La tabla 1 recoge las aplicaciones básicas empleadas por cada uno de los programas seleccionados.

Una vez superada la fase de instalación se requiere otra de pruebas pues hay que hacer ajustes técnicos algo complejos que no siempre están indicados en el manual y a los que se llega a través de las consultas a foros y listas de distribución.

La mayoría de estos programas demandan un seguimiento importante de las actualizaciones ya que pueden presentar fallos de seguridad. Sirva como ejemplo comentar que durante la exposición el servidor fue atacado y se le instaló software de *phishing* a través de un

agujero de seguridad de *PMB*, detectado y corregido en una versión posterior a la de la instalación.

En este mismo aspecto es importante comentar que en muchos casos las personalizaciones y posibilidades de adaptación y traducción que ofrecen estos programas pueden condicionar las actualizaciones futuras y requerir un esfuerzo adicional importante para adecuar los desarrollos y completar las traducciones en las nuevas versiones del programa.

### Conclusiones

Esta actividad, que empezó sólo para probar y promocionar software libre, ha permitido que la Facultad creara un espacio de aprendizaje tanto para profesores como para estudiantes, ofreciendo material de prácticas docentes para diferentes asignaturas. La posibilidad de interactuar directamente con los programas, así como la facilidad de acceso a todas sus funciones (evitando así el a veces farragoso proceso de instalación en una unidad local), ha permitido a los participantes en este proyecto conocer de primera mano diferentes soluciones gratuitas, comprobar su utilidad ante problemas y necesidades concretas, y resolver dudas e inquietudes de manera individual. Más allá de las simples demos o charlas de difusión, con este laboratorio virtual hemos pretendido difundir este tipo de programas de una manera prác-

	OpenBiblio	PMB	RefBase	TemaTres	Alfresco	Potnia	Eprints	Dspace
Apache	X	X	X	X		X	X	
Tomcat					X			X
PHP	X	X	X	X		X	Opt	
Perl							X	
Java					X			X
MySQL	X	X	X	X	Opt	X	X	Opt
PostgreSQL								X

Tabla 1

tica, de manera que fuera el propio interesado el que fuera descubriendo las posibilidades de cada programa.

## Bibliografía

**Arano, Silvia.** "Los tesauros y las ontologías en la biblioteconomía y la documentación". En: *Hipertext.net*, 2005, n. 3. Consultado en: 29-09-07. <http://www.hipertext.net>

**Crow, Raym.** "Institutional repository software options: questions and answers with the experts, with representatives from ARNO, CDSware, DSpace, Fedora, GNU Eprints, i-Tor, MPG eDoc and MyCoRe". En: *CERN workshop on innova-*

*tions in scholarly communication: implementing the benefits of OAI (OAI3)*, 2004.

**Russo-Gallo, Patricia.** "PMB: programari lliure d'automatització de biblioteques". En: *BiD: textos universitaris de biblioteconomia i documentació*, 2006, desembre, n. 17. Consultado en: 24-09-07. [http://www2.ub.edu/bid/consulta\\_articulos.php?fichero=17russo.htm](http://www2.ub.edu/bid/consulta_articulos.php?fichero=17russo.htm)

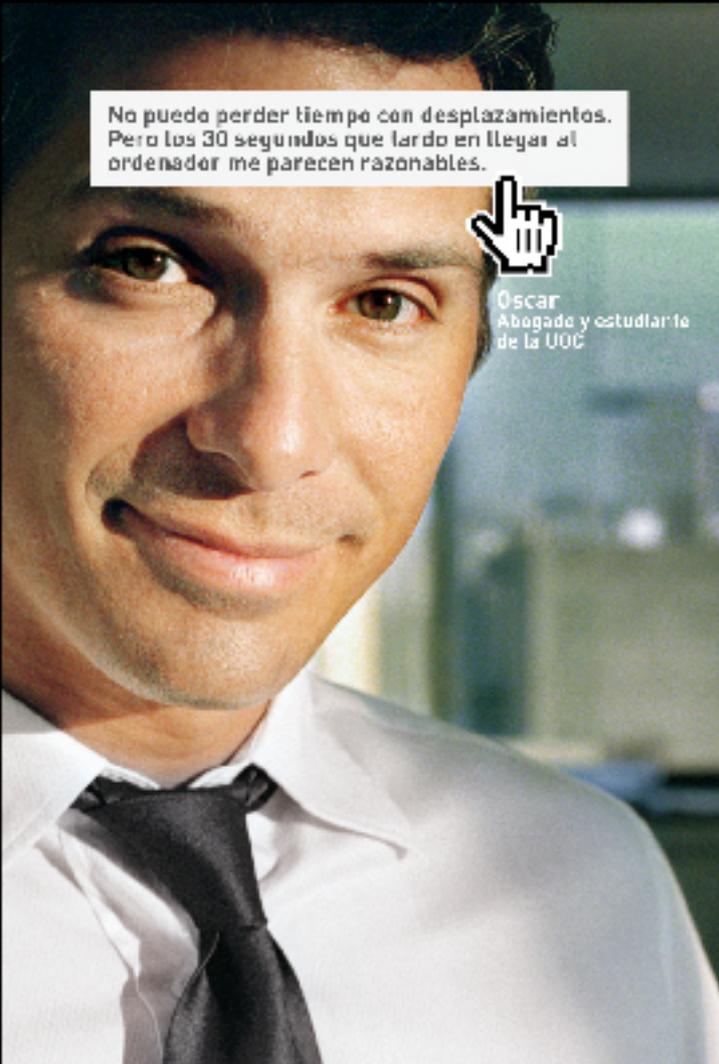
**Serrano-Muñoz, Jordi; Prats-Prat, Jordi.** "Repertorios abiertos: el libre acceso a contenidos". En: *Revista de universidad y sociedad del conocimiento*, 2005, v. 2, n. 2.

**Tramullas, Jesús.** "Herramientas de software libre para la gestión de contenidos". En: *Hipertext.net*, 2005, n. 3. Consultado en: 29-09-07. <http://www.hipertext.net>

**Patricia Russo-Gallo, Abilbo, Barcelona.**  
[prusso@uoc.edu](mailto:prusso@uoc.edu)

**Josep-Manel Rodríguez-Gairín, Facultat de Biblioteconomia i Documentació, Universidad de Barcelona.**  
[rodriguez.gairin@ub.edu](mailto:rodriguez.gairin@ub.edu)

**Andreu Sulé-Duesa, Facultat de Biblioteconomia i Documentació, Universidad de Barcelona.**  
[sule@ub.edu](mailto:sule@ub.edu)



No puedo perder tiempo con desplazamientos. Pero los 30 segundos que tardó en llegar al ordenador me parecen razonables.



Oscar  
Abogado y estudiante de la UOC

## Bienvenidos a la universidad personal

### ESTUDIOS DE CIENCIAS DE LA INFORMACIÓN Y DE LA COMUNICACIÓN

Formamos a los profesionales de la gestión de la información y la comunicación.

#### Licenciaturas de 2º ciclo:

- Licenciatura de Documentación
- Licenciatura de Publicidad y Relaciones Públicas
- Licenciatura de Comunicación audiovisual

#### Programas de Posgrado:

- Máster de Edición
- Máster de Dirección y gestión de la información y el conocimiento en las organizaciones
- Posgrado de Periodismo digital UOC - El Periódico
- Posgrado de Traducción y tecnologías

Inició: febrero 2008



Universitat Oberta de Catalunya

[www.uoc.edu](http://www.uoc.edu)

Llámanos al 902 372 373. Visita nuestra web en [www.uoc.edu](http://www.uoc.edu) o cualquiera de nuestras sedes. Te daremos toda la información que necesites. BARCELONA - Av. Diagonal 6. MADRID - Pta. de las Berlas 4. SEVILLA - Edificio EXFO - Irujo García s/n. VALENCIA - C/ de la Paz 3.

# Hacia una web semántica social

Por Jorge Morato, Sonia Sánchez-Cuadrado, Anabel Fraga y Valentín Moreno-Pelayo



**Jorge Morato** es licenciado en ciencias biológicas por la Universidad de Alcalá de Henares. A partir del año 1991 empezó a trabajar en distintas empresas relacionadas con la documentación. Obtuvo el título de doctor en documentación por la Universidad Carlos III en 1999. Desde ese mismo año ha impartido clases en las licenciaturas de documentación, LADE y en la ingeniería en informática en asignaturas relacionadas con la ingeniería de la información, recuperación de información e ingeniería del software.



**Sonia Sánchez-Cuadrado** es licenciada en filología por la Universidad Complutense de Madrid y doctora en Documentación por la Universidad Carlos III de Madrid. Desde 2002 es profesora ayudante en el departamento de informática de dicha Universidad con docencia en la licenciatura de documentación e informática sobre modelado e ingeniería de la información. Actualmente se halla involucrada en distintos proyectos para la creación automática de ontologías.



**Anabel Fraga** es ingeniero superior en informática y doctorando en ingeniería informática en la Universidad Carlos III de Madrid. Desde 2005 es profesora ayudante en el Departamento de Informática de dicha universidad con docencia en la ingeniería técnica de informática de gestión y en la ingeniería superior en informática, en temas relacionados con ingeniería de información, arquitectura de software, ingeniería del software y reutilización. Actualmente trabaja en proyectos de generación de ontologías y reutilización de conocimiento.



**Valentín Moreno Pelayo** es matemático (especialidad computación), profesor del Departamento de Informática de la Universidad Carlos III de Madrid, investigador en distintos proyectos de inteligencia artificial y recuperación conceptual financiados por Cicyt, y en redes temáticas. Actualmente trabaja en aplicación de la minería de datos al procesamiento del lenguaje natural y a la creación automática de tesauros.

**Resumen:** La evolución de internet hacia un escenario con mayor potencial y con usuarios cada vez más involucrados en su gestión y desarrollo, requiere de la incorporación de documentos con menor grado de ambigüedad semántica. La propuesta de la web semántica, a pesar de haber sido realizada hace casi una década, tiene actualmente una implantación modesta. Por el contrario, la denominada web 2.0, que se refiere a una evolución autónoma de internet hacia entornos colaborativos tiene un gran éxito. La solución propuesta por la web 2.0 ha superado las limitaciones inherentes a la utilización de herramientas de procesamiento del lenguaje natural y a las aproximaciones estadísticas. Por lo tanto, parece lógico analizar las aportaciones que estos enfoques pueden aportar a la web semántica para su implantación. En este artículo se discuten medios para dar el salto de la web social a una web semántica socializada.

**Palabras clave:** Web 2.0, Web semántica, Web semántica social, Entornos colaborativos, Folksonomías, Usabilidad.

**Title:** Toward a social semantic web environment

**Abstract:** The internet's evolution toward a scenario of greater potential, with users increasingly involved in its development and management, demands a lower level of semantic ambiguity in the documents that are provided. The proposed semantic web, a concept that appeared almost a decade ago, has had only modest impact. On the other hand, web 2.0, an autonomous evolution of the web toward a collaborative environment, has met with enormous success. The solution devised by web 2.0 has exceeded the limitations of natural language processing tools and statistical approaches. For that reason, it seems logical to analyze the potential contributions that web 2.0 concepts could make to further the development of the semantic web. Some ways of making the leap from a social web to a social semantic web are discussed.

**Keywords:** Web 2.0, Semantic web, Social semantic web, Collaborative environments, Folksonomies, Usability.

**Morato, Jorge; Sánchez-Cuadrado, Sonia; Fraga, Anabel; Moreno Pelayo, Valentín.** "Hacia una web semántica social". En: *El profesional de la información*, 2007, enero-febrero, v. 17, n. 1, ppp. 78-85.

DOI: 10.3145/epi.2008.ene.09

## Recursos de interés

- NaturalFinder.  
<http://demos.bitext.com/MSN/>
- Vivisimo.  
<http://vivisimo.com/search?tb=vivisimocom&query=&x=15&y=11>
- CleverSearch.  
<http://wdok.cs.uni-magdeburg.de/clever-search/>
- Wikipedia. Web 2.0.  
[http://en.wikipedia.org/wiki/Web\\_2.0#Characteristics\\_of\\_Web\\_2.0](http://en.wikipedia.org/wiki/Web_2.0#Characteristics_of_Web_2.0)
- Swoogle.  
<http://swoogle.umbc.edu/>
- Tabulator: async javascript and semantic web.  
<http://dig.csail.mit.edu/2005/ajar/release/tabulator/0.8/tab.html>
- SKOS Core guide.  
<http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20050510/>
- Published subject indicators for modelling thesaurii.  
<http://www.techquila.com/psi/thesaurus/>
- Zthes-specifications for thesaurus representation, access and navigation.  
<http://zthes.z3950.org/>
- Metadata Authority Description Schema (MADS).  
<http://www.loc.gov/standards/mads/>
- Simile-My Piggy Bank.  
[http://simile.mit.edu/wiki/Piggy\\_Bank](http://simile.mit.edu/wiki/Piggy_Bank)
- W3C: annotea project.  
<http://annotest.w3.org/>
- Comprehensive listing of 250 semantic web tools (updated), AI3 adaptive information.  
<http://www.mkbergman.com/?p=291>
- Searchtools.com. XML query and searching resources.  
<http://www.searchtools.com/info/xml-resources.html>
- OntoSelect.  
<http://olp.dfki.de/OntoSelect/>
- Ontaria.  
<http://www.w3.org/2004/ontaria/>
- Semwebcentral. SemWebCentral: open source tools for the semantic.  
<http://www.semwebcentral.org>
- Webkb. WebKB home page.  
<http://www.webkb.org/>

### Internet y problemas que afronta su evolución

**LAS CARACTERÍSTICAS QUE DIFERENCIAN A INTERNET respecto a los repositorios documentales tradicionales y la trayectoria que ha sufrido la web en los últimos años hacen necesario incorporar planteamientos novedosos para poder ofrecer nuevas funcionalidades. A continuación se pormenorizan algunas de estas diferencias y tendencias:**

a. Internet se caracteriza por la existencia de enormes volúmenes de documentos electrónicos, contabilizados en miles de millones.

Estos poseen una calidad variable, una temática multidisciplinar y una audiencia no predefinida.

b. La naturaleza de los documentos no es siempre textual. Los principales sistemas de indexación automática se basan en la identificación de las palabras que forman parte del texto del documento. Sin embargo, este sistema falla en el caso de documentos no textuales como imágenes, archivos de audio o vídeo. En consecuencia, se requiere de un mecanismo alternativo de indexación, dado que tradicionalmente los motores de búsqueda tienen dificultades para recuperar este tipo de materiales.

c. El extenso volumen de documentos en lenguaje natural expresado mediante html. Su escasa semántica, dada su función principalmente visual, hace que las aplicaciones informáticas no puedan acceder de forma correcta al contenido, ya que éste no se encuentra estructurado. En efecto, la mayor parte de estos documentos carece de estructura salvo por un escaso número de elementos, como el título o las etiquetas *<meta>*.

d. La autoorganización mediante diferentes mecanismos, por ejemplo, los hiperenlaces, la incorporación de páginas a directorios, la estandarización de elementos de di-

seño, o la utilización de metadatos (*tags*, *<meta>*, etc.). Este aspecto de la web en relación a su evolución ya fue descrito por **Flake** (2003).

e. El auge actual de aplicaciones informáticas para tratar la información contenida en los miles de millones de documentos existentes tiene fuertes restricciones debidas a algunas características de la web actual, como:

- La existencia de documentos no textuales.
- Las limitaciones semánticas del html.
- La falta de normalización de la estructura de los materiales presentes en internet.
- Los problemas de polisemia y sinonimia del lenguaje natural.
- En la actualidad, las respuestas a las problemáticas anteriores se producen de manera parcial y local; es decir, se practican soluciones optimizadas para casos concretos y con un grado de reutilización escaso, debido en parte a la poca interoperabilidad entre los formatos de los documentos de distintos repositorios.

f. La web ya ha alcanzado una fase de madurez, como se puede comprobar consultando algunos indicadores como:

- La simplificación de las herramientas para creación de páginas web. Hoy es posible crear un gran sitio web con escaso conocimiento informático, por ejemplo mediante herramientas de gestión de contenido (*CMS*, *Content Management Systems*) como *Joomla*, o aplicaciones colaborativas como *DokuWiki*, o *Moodle*.
- Los usuarios ya no son meros recuperadores-lectores de páginas web, sino que reclaman un papel más activo para crear y compartir recursos.
- Algunos organismos se han convertido en referentes, afectando sus decisiones a la organización de

la información. Un caso claro, es el *W3C*, ya que sus propuestas son frecuentemente aceptadas como estándares de facto.

### Posibles soluciones para la nueva internet

Para conseguir que la Red sea un repositorio con mayores funcionalidades, se han venido realizando diferentes intentos, que se pueden englobar en tres tipos: el procesamiento del lenguaje natural (PLN); la utilización de la estadística y la minería de datos; y la creación de sistemas de organización del conocimiento. Además de su descripción, a continuación se plantean algunos de los problemas que presentan estos enfoques:

a. Dotar a las aplicaciones, sobre todo buscadores, de capacidades de PLN. Aunque con evidentes atractivos, se deben afrontar distintos problemas:

– La ambigüedad terminológica inherente al lenguaje natural, agravada por el hecho de que los documentos de la web pertenecen a múltiples dominios.

– La falta de estructuración de los contenidos dificulta la interpretación por los sistemas automáticos. Además, en algunos formatos como las tablas, la expresión de los conceptos se hace visualmente en vez de utilizar estructuras lingüísticas.

– Las estructuras lingüísticas en un entorno tan heterogéneo no son previsibles. Diferentes sintaxis con los mismos términos pueden expresar el mismo concepto en unas ocasiones, o ideas diferentes en otras. Este problema, unido a la falta de estructuración de los documentos y la ambigüedad semántica, hace que sea muy complicado desarrollar una recuperación tipo pregunta-respuesta: que ante una pregunta el sistema ofrezca una respuesta, y no un conjunto de documentos que probablemente la contendrán.

– La brevedad de las consultas, frecuentemente formadas por uno o dos términos, invalida el análisis por contexto.

– Se trata de enfoques muy dependientes del lenguaje, que obtiene peores resultados en repositorios multilingües. Hay que tener en cuenta que los mejores rendimientos se obtienen cuando a la aplicación de conocimiento lingüístico se le suministran corpus etiquetados, colecciones de prueba y frecuencias de uso, y que todo este conocimiento varía según la zona geográfica, el dominio y la disponibilidad de recursos en el idioma objetivo.

– Es una aproximación con costes computacionales evidentes.

– Carece de mecanismos efectivos para mejorar la interoperabilidad entre bases de datos o para su procesamiento por aplicaciones informáticas; por ejemplo, para realizar inferencias a partir de un conjunto de documentos.

A pesar de todos estos inconvenientes, actualmente las aplicaciones web integran mecanismos de PLN. Un ejemplo es *Google*, que incorpora identificación de palabras vacías y lematización de términos no ambiguos. Así, si se escribe “nació Juan” encuentra documentos con los términos “nacido” o “nace”. Otro ejemplo se puede ver en la demo de *NaturalFinder* para buscar en lenguaje natural con el buscador *MSN*.

b. La utilización de estadísticas y de minería de la web tiene actualmente gran popularidad. Estas técnicas son frecuentemente utilizadas para la identificación de clusters de términos. Este enfoque, con un gran auge actual en los motores de recuperación web, tiene bastantes problemas, algunos comunes con el PLN, entre los que destacan:

– La escasa mejora que facilitan para optimizar la comprensión que tienen las máquinas de los documentos no estructurados.

– Los problemas ya mencionados de ambigüedad del lenguaje.

– Funcionan mejor en dominios bien representados en internet pero no en otros muy específicos o con escasa presencia de documentos.

Como en el caso anterior se puede ver el funcionamiento de los clusters en numerosas aplicaciones web, que varían desde las técnicas de clasificación documental (Steinbach, 2000) a la incorporación de estos recursos a motores de recuperación. Ejemplos en recuperadores se pueden ver en *Google* con la opción “Ud. quiso decir” o en algoritmo de posicionamiento; o en *Vivísimo*, que crea una taxonomía de forma automática a partir de la frecuencia de concurrencia de los términos de las páginas.

c. La utilización de sistemas de organización del conocimiento. Aunque parece ser la solución que más éxito tiene, no ha podido eludir la perenne dialéctica entre la utilización de lenguaje libre o controlado, o si se quiere, ir de lenguajes menos a más estructurados, de acuerdo con la graduación que se muestra en este ya tradicional esquema:

– Palabras-clave independientes, que se usan en indización libre tanto por extracción como por asignación, uno de cuyos tipos son las folksonomías.

– Listas de palabras, como los glosarios, de nombres o los diccionarios.

– Categorizaciones y clasificaciones: con el objetivo de crear conjuntos temáticos, por ejemplo las clasificaciones bibliotecarias (Zeng, 2003).

– Grupos de relaciones, basados en las establecidas entre términos y conceptos, de estructura más compleja, entre los que se sitúan los tesauros, las redes léxicas (por ejemplo, *WordNet*), los *topic maps*, y las ontologías. Como se discutirá en el siguiente punto, hoy por hoy, es la opción que más aceptación tiene entre los usuarios. Ejemplos de su empleo se encuentran en las listas de sinónimos (se puede comprobar con la búsqueda “~auto”: encuentra términos como “car”, “automobile”, etc.), o con la aplicación de la red léxica *WordNet* a la expansión de consultas en *CleverSearch*.

## Iniciativas para la evolución de internet

Hasta el momento se han planteado dos grandes iniciativas para incorporar a internet los sistemas de organización del conocimiento: la web 2.0 y la web semántica (tabla 1):

a. La evolución de la Red hacia la web 2.0, es decir, la creación de una red social que concede a los usuarios mayor control sobre los recursos. **Dale Dougherty** de *O'Reilly Media* creó el término web 2.0 en una conferencia con **Craig Cline** de *MediaLive*. Durante la charla sobre la evolución de la Red se percataron que habían surgido numerosos servicios caracterizados por un grado de colaboración y la voluntad de compartir recursos por parte de los usuarios. El término se afianzó en la *Conferencia web 2.0* en 2004 y está basado en una arquitectura de participación que anima a los usuarios a dotar a las aplicaciones de valor añadido mediante el uso que hacen del recurso, por ejemplo mediante la descripción con *tags*.

	Web 2.0	Web semántica
<b>Origen</b>	Constatación de la evolución natural de la web	Propuesta de <b>Tim Berners</b> para evolucionar la web
<b>Implantación</b>	Muy alta	Escasa ( <b>Palacios</b> , et al., 2006)
<b>Coordinación</b>	No existe	Centralizada, sobre todo por el W3C
<b>Foco</b>	Personas	Aplicaciones informáticas
<b>Primeras menciones</b>	2003, primera conferencia 2004	1999 ( <b>Berners-Lee</b> , 1999)
<b>Expresión</b>	Lenguaje libre, expresado mediante folksonomías, palabras clave denominadas etiquetas ( <i>tags</i> ), con problemas de sinonimia y polisemia	Lenguaje controlado, mediante lenguajes para expresión de ontologías, <i>KOS</i> y vocabularios de metadatos
<b>Algunas características</b>	<ul style="list-style-type: none"> <li>- Descripción de los recursos para mejorar su distribución gratuita, se comparten conocimiento y desarrollos</li> <li>- Arquitectura de colaboración</li> <li>- Usabilidad alta</li> <li>- Un recurso es más útil cuanto más uso tenga</li> </ul>	<ul style="list-style-type: none"> <li>- Utilización de un lenguaje estandarizado con sintaxis uniforme y semántica no ambigua</li> <li>- Interoperabilidad: intercambio de información entre cualquier repositorio</li> <li>- Usabilidad escasa</li> </ul>

Tabla 1. Algunas diferencias entre la web 2.0 y la web semántica

---

## “La web semántica responde a un entorno centralizado, a diferencia de la web 2.0”

---

b. La propuesta de la web semántica o web 3.0. Es decir, ampliar la audiencia de los documentos de la web, en un principio dirigido a su lectura por personas, para que sea interpretado por aplicaciones informáticas, mejorando de este modo la calidad de la recuperación y la interoperabilidad entre distintos repositorios.

### Problemas de las soluciones propuestas

Ambas iniciativas tienen ventajas e inconvenientes. En este apartado analizaremos las soluciones propuestas por cada una y las problemáticas que presentan.

#### 1. Solución aportada por la web 2.0

Internet tradicionalmente ha presentado deficiencias en la falta de semántica en la descripción de algunos recursos. La solución 2.0 a estas carencias son las folksonomías. Se trata de etiquetar recursos mediante términos de descripción procedentes del lenguaje natural. Han alcanzado gran popularidad en los últimos tiempos, probablemente por la sencillez en su aplicación, ya que no utilizan un lenguaje controlado que deba asimilarse. Entre sus ventajas destacan:

– Permitir que un usuario cualquiera, especialmente quien publica un recurso, por ejemplo una foto o un vídeo, pueda añadir términos descriptivos sin un entrenamiento previo sobre un determinado vocabulario controlado. La facilidad para añadir términos provoca que un gran volumen de usuarios se animen a etiquetar recursos a un coste nulo.

– Suprimen los costes derivados de la construcción, consenso terminológico y familiarización de un vocabulario controlado con la dificultad que supondría para el dominio de internet caracterizado por la heterogeneidad temática.

Como solución, las folksonomías se adecuan a las necesidades que internet requiere, ya que permiten:

– Etiquetar de forma económica un medio que cuenta por miles de millones los recursos disponibles.

– Adecuar las etiquetas a un término concreto, a pesar del entorno multidisciplinar y multifuncional. Pueden ajustarse a múltiples perfiles de usuarios, ya que son ellos mismos los que las incorporan según sus intereses.

– Etiquetar documentos no textuales como vídeos y fotos que de otra manera no serían correctamente recuperados por los *crawlers* de los motores web basados en indexación automática de textos.

– Facilidad para comprender el empleo e interpretación de las *tags* de la folksonomía.

– Mayor implicación de los usuarios en la gestión de los recursos.

Pero a pesar de esta simbiosis, presenta determinadas carencias debidas a que no permiten:

– Eliminar la ambigüedad terminológica.

– Una interpretación correcta por parte de las aplicaciones informáticas.

– Expandir o restringir consultas.

– Suprimir los problemas para fusionar dos folksonomías.

– La incorporación de inferencias mediante la representación de información.

– Asegurar una mejor calidad de la descripción del recurso.

Además, muchos de los términos descriptivos ocasionan ruido al no tener sentido únicamente más que para un usuario concreto (por ejemplo: “Juan”, “casa de Pepe”, etc.).

#### 2. Solución aportada por la web semántica

Uno de los pilares sobre los que se asienta es la creación de sistemas de organización del conocimiento como las ontologías. Para que la web semántica sea una realidad se necesita una sintaxis normalizada, conseguida mediante xml-rdf (W3C, 2005) y una semántica no ambigua, obtenida mediante vocabularios de metadatos, es decir, lenguajes controlados.

Sus ventajas más destacadas son:

– Interoperabilidad: algunos autores la consideran como un proyecto para crear un mediador universal para el intercambio de información. Esto es factible mediante la creación de documentos interoperables con una semántica no ambigua para las aplicaciones informáticas de la www. Es decir, se trata de convertir la web en una gran base de datos.

– Recursos semánticos: se necesita que el conocimiento semántico sea expresado en documentos escritos en un lenguaje orientado al modelado de conocimiento como rdf. Las ontologías tienen un rol importante en la web semántica porque soportan el conocimiento semántico para ser utilizado por las aplicaciones informáticas resultando información con menos ruido y mayor pertinencia para el usuario.

Lamentablemente, esta propuesta planteada en el año 1999 (Berners-Lee, 1999) no ha tenido el éxito esperado. Una búsqueda en Google muestra que, en junio de 2007, existen 2.770.000 documentos con extensión rdf, 41.500 con OWL, 2.330 XTM, 4.540.000 RSS y 212.000 ATOM. Es decir 7,5 millo-

nes de documentos en un conjunto de 10.000 millones. Con *Swoogle*, un recuperador especializado en la web semántica, los resultados no son mucho mejores: 158.000 documentos semánticos conteniendo el término rdf y 2.323.857 en el caso de *OWL*.

---

**“La web semántica es ya una propuesta de 1999 que lamentablemente no ha tenido el éxito esperado”**

---

Las causas son diversas, pero las principales a tener en cuenta son:

- Falta de legibilidad de los lenguajes rdf y *OWL*, lo cual supone un cuello de botella para que los expertos validen las ontologías (**Gómez-Pérez**, 2004). En 2005 **Mika** subrayó la importancia que tienen los usuarios (denominada “dimensión social” en el artículo original) para la aceptación de la web semántica, la cual tiene diferentes grados de complejidad en la creación de recursos, inversamente proporcional a la proximidad al usuario. En la figura 1 se muestra gráficamente este hecho: existe una tendencia que provoca que cuando se incrementa la complejidad en la representación semántica se produce una disminución en la dimensión de contacto con el usuario. Esta dimensión social engloba diferentes afectos como usabilidad, legibilidad o necesidad de conocimientos previos para su interpretación.

---

**“La web semántica tiene diferentes grados de complejidad, la cual es inversamente proporcional a la proximidad al usuario”**

---

- Escasez de herramientas que faciliten la creación de documentos semánticos mediante formularios usables, un ejemplo de un entorno más amigable se puede ver con *Protégé* o con *Tabulator*.

- La migración de folksonomías a folkontologías es un tema aún por desarrollar, aunque ya existen estudios (**Damme**, 2007; **Matsuo**, 2006). Básicamente, este proceso se hace mediante herramientas estadísticas (**Bagelman**, 2006) y lingüísticas, o incluso proponiendo una normalización para asignar etiquetas (**Xu**, 2006).

- La necesidad de incorporar técnicas semiautomáticas para la creación de sistemas de organización del conocimiento basadas en PLN y minería de datos (**Sánchez-Cuadrado**, 2007), ya que la carencia de estos recursos, junto con la lentitud en su creación dificulta la implantación de la web semántica.

- Presencia de duplicidades en los vocabularios de metadatos y ontologías, que provoca la desconfianza y confusión del usuario que no sabe cuál es el idóneo o más generalizado. Como ejemplo están los vocabularios de metadatos para

expresar tesauros. Actualmente existen, entre otros: *SKOS-Core* del W3C, *PSI* de los *topic maps*, *Zthes* y *MADS*.

**Discusión. Hacia un nuevo conjunto de soluciones para un nuevo concepto de internet**

Como consecuencia, parece evidente que la única forma de popularizar la web semántica sería mediante una aproximación a la web social, para implicar a los usuarios en la creación y gestión de documentos semánticos. Actualmente, se están realizando distintas propuestas de conseguir una mayor usabilidad de los documentos semánticos, obtener una mayor implicación de los usuarios, y suministrarles recursos que les ayuden a mejorar la localización y el grado de popularidad de los contenidos.

- Una opción prometedora en esta dirección son los enfoques híbridos que combinan ambas tendencias. Un ejemplo es *Piggy Bank* (figura 2), una aplicación que captura localmente las etiquetas de los rdf visitados, para organizarlos en una ontología local, y que puede ser compartida en el *Semantic Bank*

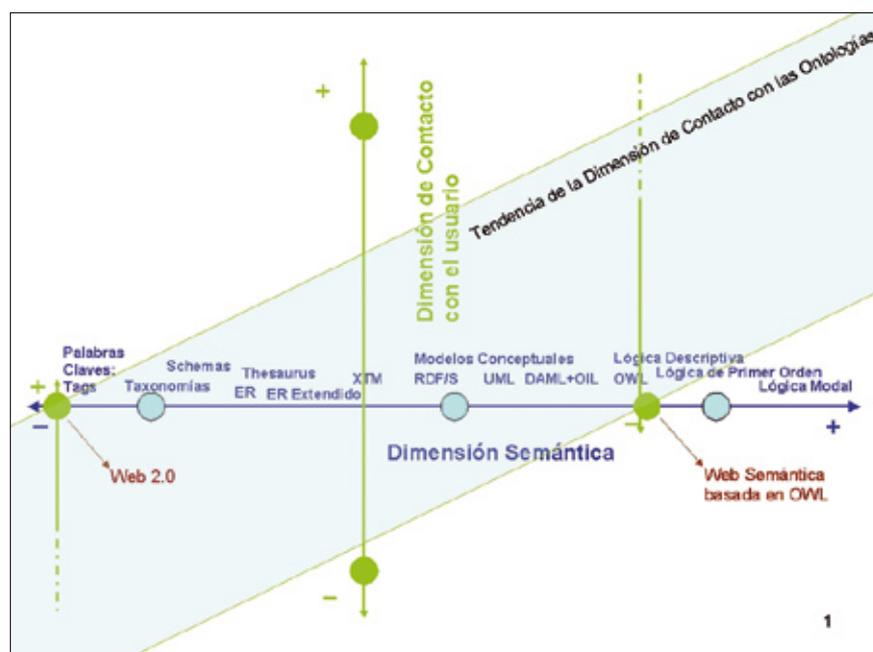


Figura 1. Relación entre la complejidad de la web semántica y la proximidad al usuario

**“La única forma de popularizar la web semántica sería mediante una aproximación a la web social para implicar a los usuarios en la creación y gestión de documentos semánticos”**

(Senso, 2007). Esta herramienta permite poner *tags* a las url seleccionadas para recuperarlas posteriormente. Un proyecto muy similar ha sido propuesto por el W3C bajo la denominación de *Annotea*, instalable mediante un plugin en *Firefox* denominado *Annozilla*. El problema de ambas aproximaciones son las inherentes a las *tags*, es decir, la sinonimia y la polisemia.

– Como ya se mostró en la figura 1, es necesario cambiar la trayectoria natural de la dimensión de contacto con el usuario mediante aplicaciones informáticas usables o la simplificación. La simplificación (Senso, 2007) se puede dar de diferentes formas, por ejemplo, prescindiendo de los lenguajes formatos a favor de expresar el contenido mediante documentos Xhtml (una solución denominada microformatos). Esta opción, dado su carácter de solución local, no consigue uno de los principales objetivos de la web semántica, la interoperabilidad. Por el contrario, la interoperabilidad no se ve mermada en determinados documentos semánticos que han tenido un gran éxito gracias a la simplicidad del vocabulario subyacente; un ejemplo es la sindicación de contenidos mediante *RSS*.

– La creación de nuevos recursos que faciliten la localización y popularización de los documentos semánticos: un ejemplo es el ya mencionado buscador *Swoogle*, o los también conocidos *OntoSelect* u *Ontaria*. Como en el caso de *PigggyBank*, *OntoSelect* es a su vez un

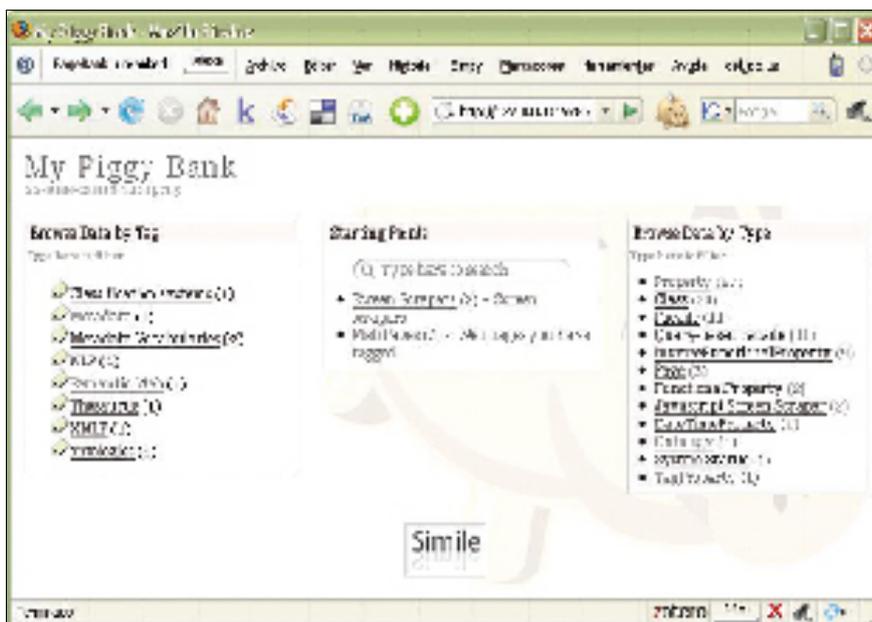


Figura 2. Visualización de la captura de una página con MyPiggyBank

repositorio de documentos semánticos.

Existen propuestas, en fase de experimentación, que incorporan varias de estas tecnologías. Estas soluciones combinan repositorios de documentos semánticos con buscadores sobre popularidad y ayudas para traducir las *tags* de los usuarios a un vocabulario genérico centralizado basado en recursos similares a *WebKB* (Palacios, 2006). Sería capaz de proporcionar estadísticas de popularidad de cada ontología, lo que evitarían problemas de confusión del usuario debido a duplicidades de vocabularios.

### Consideraciones finales

La trayectoria de internet en los últimos años hace necesario incorporar nuevos planteamientos a la resolución de problemas clásicos para tratar el gran volumen de información de todo tipo en diferentes formatos y dirigida a un público heterogéneo. Por otro lado, los usuarios y las aplicaciones informáticas necesitan de un mayor contenido semántico para dar mayor funcionalidad a la web.

Las soluciones adoptadas han seguido varios caminos: el desa-

rollo de herramientas de procesamiento del lenguaje natural, la incorporación de técnicas estadísticas, y la creación de recursos terminológicos para dotar a los recursos de una semántica adicional.

Entre las propuestas para la creación de recursos terminológicos destacan las folksonomías y las ontologías, procedentes respectivamente de cada una de las dos grandes iniciativas planteadas hasta el momento: la web 2.0 y la web semántica. La primera de ellas ha logrado un éxito mayor debido a su proximidad al usuario y a su simplicidad, pero tiene el gran inconveniente de no poder subsanar problemas de interoperabilidad, sinonimia y polisemia. Por el contrario la web semántica no consigue alcanzar el mismo grado de implantación a causa de los mismos factores, poca legibilidad y complejidad, pero carece de los problemas de la web 2.0.

**“Se propone la creación de recursos semánticos más usables, simples y con potencialidades de búsqueda”**

Como soluciones, en este documento se proponen la creación de recursos semánticos más usables, simples y con potencialidades de búsqueda. Una posibilidad sería la traducción de las folksonomías de los usuarios a ontologías, algo que se podría realizar mediante dos procedimientos:

– La incorporación de procesamiento del lenguaje natural y estadístico a la migración (folksonologías).

– La asistencia al usuario para alinear su vocabulario en lenguaje natural con otro en lenguaje controlado.

## Bibliografía

**Begelman, G.; Keller, P.; Smadja, F.** “Automated tag clustering: improving search and exploration in the tag space”. En: *WWW2006*, 2006. <http://www.rawsugar.com/www2006/20.pdf>

**Berners-Lee, T.; Fischetti, M.** *Weaving the web*. 1st edition. San Francisco: Harper, 1999.

**Damme, C.**, et al. *Ontology: an integrated ap-*

*proach for turning folksonomies into ontologies*, 2007. <http://www.heppnetz.de/files/vandammehepp-siorpaes-folksonology-semnet2007-crc.pdf>

**Flake, G. W.; Pennock, D. M.; Fain, D. C.** “The self-organized web: the yin to the semantic web’s yang”. En: *IEEE intelligent systems*, 2003, July/Aug, pp. 72–86.

**Fumero, A.; Roca G.; Encinar, J.** *Web 2.0*. Colección Fundación Orange, 2007. [http://www.fundacionauna.com/areas/25\\_publicaciones/indice\\_web2.asp](http://www.fundacionauna.com/areas/25_publicaciones/indice_web2.asp)

**Gómez-Pérez, A.**, et al. *Ontological Engineering*. London: Springer, 2004.

**Matsuo, Y.**, et al. “Spinning multiple social networks for semantic web”. En: *AAAI-06*, 2006.

**Mika, P.** “Ontologies are us: a unified model of social networks and semantics”. En: *Proceedings of the 4th International semantic web conference (ISWC 2005)*, 2005. <http://www.cs.vu.nl/~pmika/research/papers/ISWC-folksonomy.pdf>

**Palacios, V.; Morato, J.; Sánchez-Cuadrado, S.; Lloréns, J.; Moreiro, J. A.** “An improved methodology for semantic scheme qualification”. En: *The 1st international workshop: semantic information integration on knowledge discovery (SIK 2006)*, 2006.

**Sánchez-Cuadrado, S.** *Definición de una metodología para la construcción automatizada de un sistema de organización del Conocimiento*. Tesis Doctoral Universidad Carlos III, 2007.

**Senso, José A.** “Navegadores semánticos o semantizar el navegador”. En: *ThinkEPI*, 2007. <http://www.thinkepi.net/repositorio/navegadores-semanticos-o-semantizar-el-navegador/>

**Steinbach, M.; Karypis, G.; Kumar, V.** *A comparison of document clustering techniques*. Technical Report 00-034, Department of Computer Science and Engineering University of Minnesota, 2000. <http://glaros.dtc.umn.edu/gkhome/node/157>

*Getting into rdf & semantic web using N3, W3C*. <http://www.w3.org/2000/10/swap/Primer>

**Xu, Z.**, et al. “Towards the semantic web: collaborative tag suggestions”. En: *WWW2006*, 2006. <http://www.rawsugar.com/www2006/13.pdf>

**Zeng, Marcia Lei; Mai Chan, Lois.** “Trends and issues in establishing interoperability among knowledge organization systems”. En: *Journal of the American Society for Information Science and Technology*, v. 55, n. 5, pp. 377 – 395

**Jorge Morato; Sonia Sánchez-Cuadrado; Anabel Fraga; Valentín Moreno-Pelayo**, Departamento de Informática, Universidad Carlos III, Av. Universidad, 30. 28911 Leganés (Madrid).  
[jorge@ie.inf.uc3m.es](mailto:jorge@ie.inf.uc3m.es)  
[ssanchez@ie.inf.uc3m.es](mailto:ssanchez@ie.inf.uc3m.es)  
[afraga@ie.inf.uc3m.es](mailto:afraga@ie.inf.uc3m.es)  
[vmoreno@ie.inf.uc3m.es](mailto:vmoreno@ie.inf.uc3m.es)

## Suscripciones

Renovar (o comenzar) la suscripción a “El profesional de la información” es ágil y sencillo.

Usted puede gestionar online su suscripción conectándose a esta web:

<http://www.elprofesionaldelainformacion.com/suscripciones.html>

Si lo desea puede comunicar con nosotros dirigiéndose a:

[suscripciones@elprofesionaldelainformacion.com](mailto:suscripciones@elprofesionaldelainformacion.com)

o al teléfono: +34-609 352 954

# Ontología para el control y recuperación de información onomástica en televisión

Por Jorge Caldera-Serrano y Rodrigo Sánchez-Jiménez

**Resumen:** Se muestra una propuesta de ontología creada para el control documental de información de personajes visualizados y referenciados en los servicios de documentación de las cadenas de televisión. Para ello se plantea el estado actual del uso de ontologías en el medio audiovisual, para posteriormente ir mostrando y explicando el desarrollo creado. Trabajo práctico donde se ofrece una solución rápida y sencilla para el control documental orientado específicamente a la recuperación de elementos onomásticos, por medio del cual se controlan y desarrollan las relaciones entre personajes. De esta forma se crea una auténtica red de personajes televisivos que puede ser explotada para mejorar los resultados durante el proceso de recuperación.

**Palabras clave:** Ontologías, Información audiovisual, Control onomástico, Recuperación de información, Televisión.

**Title: Development and application of an ontology for onomastic information control and retrieval in television**

**Abstract:** The development of an ontology for information control is proposed to manage visual images of and references to individuals in information management systems of television channels. Research has established precedents for the use of ontologies in the audiovisual field, providing a baseline for measuring and understanding ongoing developments. This practical approach offers a quick, straightforward solution to onomastic control in information retrieval, permitting control and development of relationships between individuals covered in the media. This creates a network of television personalities we can draw upon to improve the outcomes of retrieval processes.

**Keywords:** Ontologies, Audiovisual information, Named entities, Information retrieval, Televisión.

**Caldera-Serrano, Jorge; Sánchez-Jiménez, Rodrigo.** “Desarrollo y aplicación de una ontología para el control y recuperación de información onomástica en televisión”. En: *El profesional de la información*, 2007, enero-febrero, v. 17, n. 1, pp. 86-91.

DOI: 10.3145/epi.2008.ene.10



**Jorge Caldera Serrano** es profesor en la Facultad de Biblioteconomía y Documentación de la Universidad de Extremadura, donde imparte docencia sobre archivos de televisión, filmotecas y fuentes de información audiovisual. Es licenciado y doctor en Documentación por la Universidad de Salamanca y cuenta con la experiencia profesional de dos años en los Servicios Informativos de Televisión Española (Madrid) en el archivo audiovisual.



**Rodrigo Sánchez Jiménez** es profesor del Departamento de Biblioteconomía y Documentación de la Universidad Complutense de Madrid. Imparte docencia en materias de documentación informativa y técnicas de recuperación de información. Ha participado en varios proyectos de investigación relacionados con lenguajes documentales y documentación audiovisual y multimedia. Es autor de artículos sobre metadatos, documentación multimedia y recuperación de información.

## Introducción

**LAS CIENCIAS DE LA DOCUMENTACIÓN SIEMPRE SE HAN BASADO** en el desarrollo de otros campos transversales para abrirse paso y especializarse. El hecho de ser una disciplina de apoyo y auxilio ha obligado a buscar mejores técnicas para la realización de nuestra labor, especialmente en lo relativo a la recuperación de información. La gestión de información en los medios

**de comunicación es un fiel reflejo de la adaptación de herramientas documentales e informáticas al tratamiento documental.**

Son muchos los intentos que se han llevado a cabo en esta línea, como la implementación de los tesauros desarrollados por *Antena 3TV* y *Televisión Española*, o tesauros facetados como el confeccionado por el grupo liderado por **Félix del Valle** y realizado para *Telecinco*, herramientas de clasificación, recu-

peración en texto libre, etc. Pero el gran crecimiento en cantidad y en exigencia de esta información hace que se sigan buscando nuevas vías para su control. Aunque los esfuerzos han sido amplios y ambiciosos, la realidad es que muchas de estas herramientas y técnicas aún no se adecuan a las necesidades de los periodistas ni a los requerimientos de los documentalistas.

El problema fundamental del análisis documental de imágenes en

movimiento es la descripción de acciones, ya que no son temas ni conceptos abstractos sino la representación de iconos que se modifican a lo largo del tiempo. Para resolver este aspecto se ha utilizado en ocasiones una combinación de tesaurus y clasificación facetada, única forma viable de afrontar la descripción e identificación de acciones.

Las ontologías pueden constituir una buena solución a este problema dado su potencial y flexibilidad a la hora de interrelacionar conceptos, ideas y acciones, optimizando así la recuperación. La creación de herramientas ontológicas facilita la recuperación siempre y cuando la ontología creada haya sido planificada y organizada con criterios claros y pragmáticos. No obstante, una ventaja capital es su utilización para el intercambio de información entre empresas audiovisuales que utilicen una misma plataforma.

Las coordenadas de nuestro estudio se derivan de uno de los casos de aplicación más necesarios y evidentes: el control de personas físicas y jurídicas en el ámbito televisivo. No tratamos el ámbito temático por tratarse de un proceso diferente, con estructuras y relaciones ontológicas mucho más confusas y de compleja adaptación al modelo audiovisual.

### **Documentación en televisión**

**López de Quintana** (2000) indica los principios fundamentales que debe regir la documentación en un medio televisivo; por un lado, la preservación del patrimonio documental generado por la cadena y,

por otro, integrar la documentación en sus actividades de producción y gestión para la creación de nuevos productos. La documentación en las empresas televisivas parte para **Hidalgo-Goyanes** (1999) de tres premisas fundamentales para el cumplimiento total de la misión y objetivos de los servicios: rapidez, eficacia y credibilidad.

**López de Quintana** (2000) realiza la división del organigrama atendiendo a diferentes parámetros tales como:

- Separación por tipo de actividad. Consiste en la división orgánica en la que se distingue la videoteca del centro de documentación, ya que la primera se responsabiliza del control físico del material audiovisual mientras que la segunda se encargaría de todo el proceso técnico.

- Atendiendo a la tipología documental. Suele ser habitual en los departamentos de documentación de las televisiones crear áreas de trabajo según la naturaleza del tipo documental, pudiendo encontrar secciones dedicadas a la documentación audiovisual, escrita, fotográfica y sonora.

- Por tipo de programas. Otra división que suele llevarse a cabo en las grandes cadenas televisivas con años de tradición es la separación en el organigrama documental dependiendo del tipo de programa sobre el cual se trabaja. La división se realiza entre programas de entretenimiento e informativos.

- Especialización temática y de técnicas documentales. Aunque sea cada vez menos habitual, en algunos centros el documentalista

está especializado en unas técnicas concretas, como el análisis o la difusión.

Queda claro que la gestión de la información en televisión va mucho más allá de la propia documentación audiovisual como se atestigua en **Caldera y Zapico** (2006), por lo que nos encontraremos en estos servicios con información fotográfica, sonora y textual.

No expondremos aquí de forma más pormenorizada los procesos de descripción en televisión ya que existen documentos más amplios que lo analizan (**Caldera; Nuño**, 2004), (**Fournial**, 1986). No obstante, sí es necesario subrayar la existencia de diferencias en el tratamiento y conservación de registros referenciales para programas informativos y de entretenimiento, consecuencia principalmente de los requerimientos de los usuarios y de los elementos formales intrínsecos de las piezas audiovisuales televisivas de ambas categorías.

### **El control actual de personas físicas y jurídicas en televisión**

En el trabajo diario de las televisiones, la inspección de esta información básicamente se centra en un control de autoridad de los personajes aparecidos y referenciados en la información. Esto puede ser suficiente para muchas unidades informativas, pero no lo es siempre en los sistemas de información de las empresas televisivas. Pongamos un ejemplo representativo de este problema: ante la consulta “recursos de presidentes autonómicos de la década de los 80 en España”, su procesamiento no podría basarse exclusivamente en criterios onomásticos, sino que el documentalista tendría que acotar cronológicamente, consultar textos para identificar parcialmente cada uno de ellos y, por supuesto, guiarse por sus referencias culturales, informativas y formativas.

---

**“Las ontologías cuentan con la ventaja de intercambiar la información entre empresas audiovisuales, además de ser flexibles interrelacionando conceptos, ideas y acciones”**

---

No podemos dejar al criterio personal de los documentalistas este tipo de interrogación, sino que deben confeccionarse y prepararse herramientas de control que nos guíen en el momento de la consulta y recuperación. Aunque existen métodos de investigación documental para poder llegar a concluir con éxito la interrogación anterior, la diferencia fundamental es el tiempo de reacción con el que se cuenta en las televisiones. La existencia como mínimo de tres informativos y de una parrilla estable de programas, mucho de ellos en directo, provoca la necesidad de contestar en minutos a consultas en muchos casos extremadamente complicadas.

**“La premura de tiempo en la resolución de las interrogaciones de los usuarios marca el ritmo y los métodos de trabajo. El ‘futuro’ es el siguiente programa que se emite”**

La información onomástica que podría definirse en la televisión es variada (Caldera; Nuño, 2004), no obstante indicamos las principales: personas físicas/referencias, entidades visionadas/referenciadas. La información que se extrae tanto de la banda de imagen como de la de sonido va en estancos separados, ya que a la hora de la recuperación no es idéntico el requerir una persona que se vea, una persona que hable o de la que trate la información (figura 1).

Estos puntos de acceso deben ser inspeccionados por medio de un control de autoridad, pero además pueden estar integrados en una ontología que aporte una mayor información, potenciando su recuperación por los diferentes elementos asociados a sus personas.

El trabajo con registros de programas informativos y de entretenimiento es bastante diferente debido a la diversa tipología documental. Para el caso que nos ocupa estimamos más factible desarrollar esfuerzos para potenciar una mejor y más coherente recuperación de personajes en los programas de entretenimiento, aunque también es útil y puede implementarse con mayor número de gestores de información para programas informativos.

Las relaciones entre los personajes de los programas de ocio tienen mayor importancia y utilidad. Al mismo tiempo es más factible llevar a cabo un control de autoridad exhaustivo debido a que el número de elementos que se integran semanalmente es más reducido. Piensen en los programas denominados del “corazón”, donde su número es importante y el de personajes es bastante limitado. Igualmente válidos son para los “reality show”, donde interesan los personajes protagonistas y las diferentes relaciones dentro y fuera del concurso. Por el contrario, en programas informativos el número de nuevos personajes es a veces desproporcionado, por lo que crear una ontología para controlarlos es factible, pero se traduce en mucho tiempo de dedicación por parte del documentalista.

**El concepto “ontología” y la información audiovisual**

A continuación se pretende aportar una revisión bibliográfica sobre el concepto de ontología, identificando aquellas definiciones que creemos más acertadas y allegadas al ámbito de lo audiovisual desde el punto de vista de la realización de nuestro trabajo.

Como labor de recopilación sobre este concepto indicamos el trabajo de **Corcho, Fernández y Gómez Pérez (2003)**, que analizan las ontologías desde el concepto, los métodos y las herramientas de construcción. **García Jiménez (2004)** señala el objetivo último de las ontologías en cuanto a su desarrollo, definido como la mejora de la representación y recuperación de la información. **Noy y McGuinness (2000)** en su guía para la creación de ontologías, identifican sus objetivos principales: compartir la comprensión de estructuras entre personas o agentes software, permitir la reutilización del conocimiento perteneciente a un dominio, hacer explícitos los supuestos de un dominio, etc.

Sobre las definiciones aceptadas por la comunidad científica, la de **Gruber (1993)** es la más estandarizada: especificación explícita y formal de una conceptualización compartida. Este mismo autor (1993b) identifica los siguientes componentes:

- Conceptos: ideas que se intentan formalizar.
- Relaciones: representando la interacción entre los conceptos.
- Funciones: ofrecen un tipo concreto de relación, utilizados para la representación de objetivos determinados de un concepto.
- Axiomas: teoremas que identifican las relaciones existentes entre elementos.

Interesante es también la aportación de **Guerrero-Bote y Loza-**

Puntos de acceso onomásticos	
Visionado	Referenciado
Personas físicas	Personas físicas
Personas jurídicas (entidades)	Personas jurídicas (entidades)

Figura 1. Esquema de puntos de acceso onomásticos

no (1999), quienes señalan que el dominio más cercano a la documentación son las ontologías confeccionadas desde el punto de vista terminológico (representación del conocimiento) e informativo (estructuración del almacenamiento de la información en las bases de datos). Será desde la perspectiva informativa desde la que abordemos y elaboremos nuestra propuesta de ontología.

Desconocemos la existencia de experiencias de diseño y creación de ontologías en las televisiones españolas o internacionales. La revisión de literatura científica no aporta resultados, al igual que los contactos directos con departamentos de documentación de diferentes cadenas estatales y autonómicas. Las empresas audiovisuales son compañías en las que el intercambio y difusión de información suele ser escaso, lo que no nos debe extrañar al ser corporaciones con idénticos fines, en lucha por la audiencia y los ingresos publicitarios.

### Creación y desarrollo de la ontología

Se han de tener presentes una serie de elementos a la hora del diseño. Las ontologías deben ser claras, coherentes, adaptables, específicas y precisas. Claras para así comunicar de manera unívoca el significado. Coherentes para inferir de manera consistente entre definiciones. Adaptables para poder ofrecer nuevos usos. Específicas al detallar a nivel de conocimiento y precisas en su concepción y entendimiento.

**“Las ontologías han de ser claras, coherentes, adaptables, específicas y precisas”**

Nuestra labor se va a centrar en implementar una ontología en la que se especifica la estructura del almacenamiento de la información en una base de datos, mejorando así la identificación de las relaciones y optimizando el producto y resultado final ofrecido al usuario (Barchini, 2006)

Por todo ello, y con la necesidad de controlar los diferentes parámetros de personajes aparecidos en pantalla, se desarrolla en una primera fase un esquema que se expone en la figura 2.

Sobre cada uno de los personajes que aparecen en la parrilla televisiva se indicará al sistema tanto fecha y lugar de nacimiento como defunción en su caso, así como otras informaciones de carácter personal que deberán ser perfiladas en cada departamento de documentación atendiendo a las características de las solicitudes de los usuarios. Se definirán las posibles relaciones con entidades determinando las fechas, así como la responsabilidad del personaje en las mismas. Se intentará completar el mayor número de atributos en cada registro, aunque admitiendo que ocasionalmente algunos no se utilicen o puedan ser modificados.

No obstante, para una correcta gestión de esta información es necesaria la puesta en marcha y el se-

guimiento de un estricto control de autoridades, especificando así las diversas formas de denominaciones como variantes ortográficas, de tal manera que las interrogaciones por cada una de ellas siempre ofrezcan idénticos resultados. No queremos entrar en polémicas sobre esta cuestión, pero las normativas actuales de control de autoridades son excesivamente estrictas para la televisión. Además, el control en muchas cadenas es deficitario por lo que se debe comenzar realizando este esfuerzo normativo.

Se presenta en la figura 3 el modelo ontológico para gestionar la información onomástica en la base de datos de información audiovisual en una televisión.

Establecemos 5 clases primarias que representan los conceptos fundamentales para la descripción y recuperación de personajes. Los personajes constituyen la clase fundamental y son descritos en torno a características (*slots*) representados como rectángulos de borde azul para los atributos cuyo valor es tan sólo texto, como el nombre, los datos de control de autoridad (que incluirán variaciones ortográficas y diferentes versiones), y otros datos como las aficiones, su estado civil, etc. Respecto al control de autoridad, hay que decir que cada instancia que representa un personaje

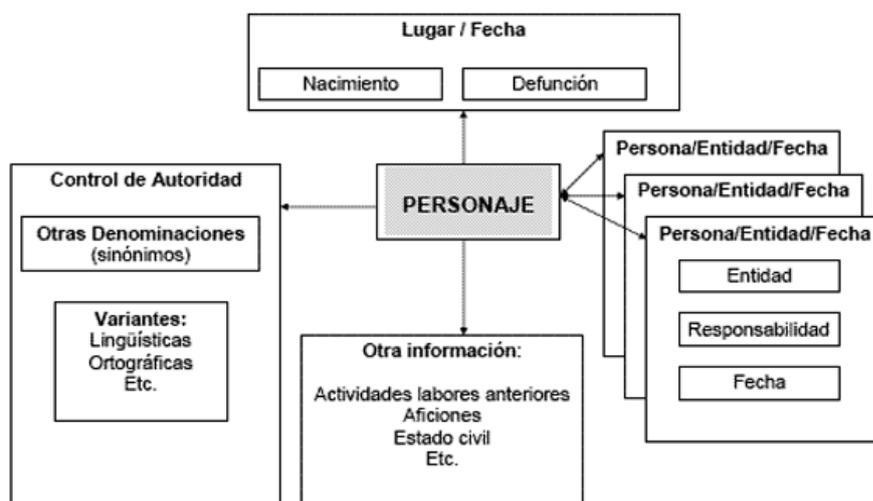


Figura 2. Control onomástico de personaje en TV

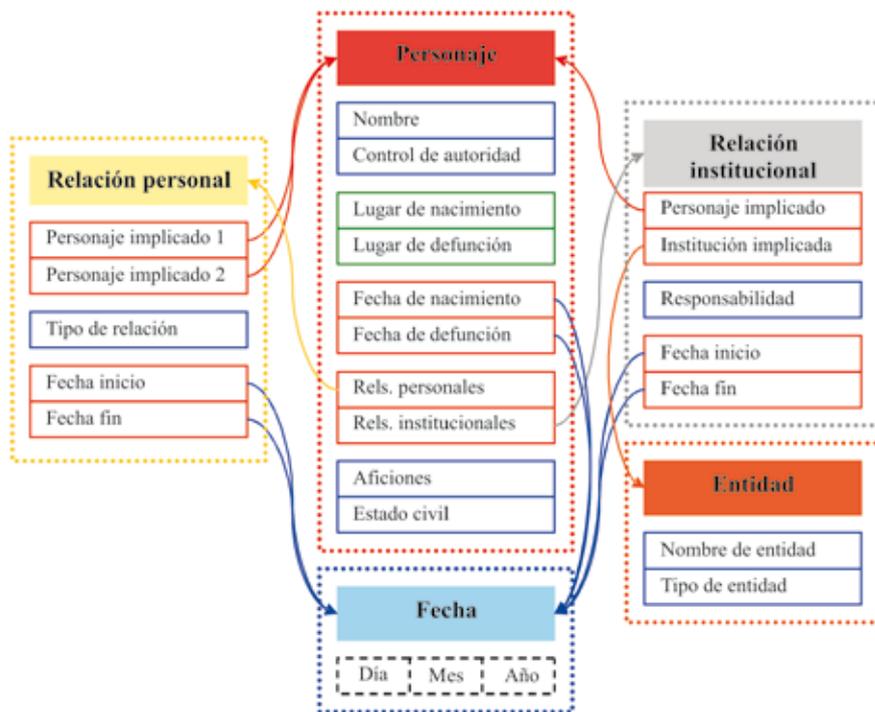


Figura 3. Estructura de ontología de personajes para TV

incluye todas las variantes de su nombre aparecidas hasta la fecha, y que una consulta cualquiera remitirá al mismo independientemente de la opción utilizada.

Las características marcadas como rectángulos de borde rojo definen vínculos con otras clases, otros conceptos de importancia para la recuperación, como relaciones personales e institucionales y las fechas. Por último, los lugares de nacimiento y defunción (rectángulos de borde verde) deberían indicarse utilizando otro recurso que permita el control del vocabulario, para lo que frecuentemente se cuenta con un tesoro geográfico. Esto permitiría mantener la consistencia en la descripción y facilitar la recuperación.

Las relaciones constituyen un concepto importante en sí mismas, por lo que se han representado como clases con sus propias características. Las fechas durante las cuáles tiene lugar la relación, así como el tipo específico de la misma, nos deberían servir para restringir los resultados de las búsquedas. Probablemente sería una buena idea

crear diferentes tipos de relaciones personales e institucionales como subclases de la clase principal, aunque este aspecto variará fuertemente en cada ocasión, por lo que nos limitamos a proponer una solución genérica. En el caso de las relaciones institucionales se vincula a una persona con una entidad, delimitando las fechas y el tipo de responsabilidad. En cuanto a las relaciones personales se relaciona a dos personas diferentes indicando el tipo de relación, que podría ser laboral, amorosa, de familia, etc.

Las entidades pueden en realidad ser descritas mediante muchos más atributos que contribuirán a mejorar los resultados de las consultas, pero de nuevo nos limitamos a ofrecer unos mínimos que nos permiten hacernos una idea de los resultados en un entorno real. El tipo de entidad es de nuevo una característica que podría constituirse en un conjunto de subclases, aunque la profundidad y detalle con que se haga dependerá específicamente de la aplicación que se le quiera dar.

Por último, las fechas se han introducido como clases con un atri-

buto para los días, otro para los meses y otro para los años, cuyos valores deberían ser números enteros. El motivo fundamental es la necesidad de hacer cálculos sobre los valores indicados de manera que podamos resolver en una consulta criterios temporales con exactitud.

La aplicación de una ontología como la propuesta en la práctica pasa fundamentalmente por tres requisitos. En primer lugar es necesario valorar la necesidad de refinar el esquema de clases propuesto para adaptarlo a las necesidades reales de la descripción en un entorno concreto. Como ya anticipábamos podría ser necesario introducir una formalización más enriquecedora de los tipos de relaciones, así como de los tipos de entidades. En segundo lugar será preciso crear las instancias pertinentes para cada una de las clases (es decir, llevar a cabo la descripción de un número más o menos elevado de personajes para poder evaluar el sistema). Finalmente es necesario disponer de una forma de almacenar e interrogar la ontología.

Sobre este último aspecto existen diferentes posibilidades a tener en cuenta. La primera de ellas es establecer una base de datos relacional sobre el esquema propuesto. Es una buena opción si no se dispone de herramientas específicas. Otra posibilidad, bastante más adecuada, es establecer un sistema de interrogación basado en *Sparql*, el lenguaje de consultas desarrollado por el W3C para *rdf* y *OWL*, los lenguajes que constituyen la espina dorsal de la web semántica. Esta solución es bastante más compleja, aunque permitirá explotar mejor la ontología. Por último se puede utilizar un razonador, un programa capaz de llevar a cabo inferencias que se derivan de la estructura de la ontología y de un conjunto de reglas establecidas mediante lógica descriptiva. Es la alternativa más compleja, aunque también la más potente.

## “Las ontologías pueden constituir una opción excelente para la comprensión y utilización de las relaciones y procedencias de los personajes en televisión”

En nuestro caso, para probar la ontología propuesta utilizamos *Protégé*, un software de la *Universidad de Stanford* que permite desarrollarla y llevar a cabo consultas bastante complejas para explotarla. Durante el período de pruebas de la ontología, que se completó con la creación de varias decenas de personajes y relaciones, se observó cómo se podían resolver con precisión consultas como “recursos de presidentes autonómicos de la década de los 80 en España” o “personajes que mantuvieron una relación amorosa con empresarios de la construcción”. Sin embargo es necesario realizar un proceso de experimentación más profundo para llegar a resultados concluyentes.

Evidentemente el grado de éxito de la recuperación se deriva del tiempo y recursos empleados en el refinamiento de la ontología y su adaptación a las necesidades concretas del medio. Sin embargo, en conjunto, las ontologías proporcionan una forma de representar las complejas relaciones sociales y humanas de los personajes televisivos con un gran potencial de explotación.

### Conclusiones

A modo de colofón se puede señalar que las televisiones actuales necesitan un mejor control terminológico de sus herramientas documentales, e igualmente una mejor adaptación a las cambiantes necesidades de los usuarios de los medios audiovisuales. Las herramientas cerradas y poco flexibles hacen que su viabilidad y potencialidad sean escasas. Y es en este escenario don-

de las ontologías muestran un desarrollo muy interesante.

En este trabajo se ha pretendido dar una utilidad parcial a las ontologías, solucionando un problema actual existente en el control y recuperación de los personajes visionados y referenciados en los programas de entretenimiento. Y decimos que parcial ya que la utilización de ontologías debe seguir desarrollándose en otras fases, incorporando elementos temáticos de material visualizado y referenciado para que realmente podamos encontrar por fin una herramienta ágil, ergonómica y adaptable a unas necesidades cambiantes.

Las ontologías pueden constituir una opción excelente para la comprensión y utilización de las relaciones y procedencias de cada uno de los personajes, lo que a veces no es nada sencillo en el contexto televisivo. Sin embargo, la viabilidad final del enfoque sólo puede ser probado y desarrollado sobre escenarios reales, con infraestructuras y datos de los que tan sólo las grandes corporaciones disponen. Por este motivo, nuestra propuesta pretende llamar la atención sobre las excelentes posibilidades que ofrecen las ontologías desde la perspectiva de la representación del dominio de los personajes televisivos, e incitar en lo posible a que se investigue en profundidad en el futuro.

### Referencias

- Barchini, Graciela E.; Álvarez, Margarita M.; Herrera, Susana. “Sistemas de información: nuevos escenarios basados en ontologías”. En: *Revista de gestão da tecnologia e sistemas de informação*, 2006, v. 3, n. 1, pp. 3-18.
- Caldera-Serrano, Jorge; Nuño-Moral, M. V. *Diseño de una base de datos de imágenes para*

*televisión*. Gijón: Trea, 2004, Isbn 84-9704-100-3.

Caldera-Serrano, Jorge; Zapico-Alonso, Felipe. “Seen and heard; duality at the access point to television database”. En: *Aslib proceedings: news information perspective*, 2006, v. 58, n. 4, pp. 30-43.

Corcho, Óscar; Fernández-López, Mariano; Gómez-Pérez, Asunción. “Methodologies, tools, and languages for building ontologies. Where is their meeting point?”. En: *Data & knowledge engineering*, 2003, v. 46, n. 1, pp. 41-65.

Fournial, Catherine. “Análisis documental de imágenes en movimiento”. En: *Panorama de los archivos audiovisuales*. Madrid: Servicios de Publicaciones de RTVE, 1986, pp. 249-258, Isbn 84-85259-19-X.

García-Jiménez, Antonio. “Instrumentos de representación del conocimiento: tesauros versus ontologías”. En: *Anales de la documentación*, 2004, n. 7, pp. 79-95.

Gruber, Thomas R. “Toward principles for the design of ontologies used for knowledge sharing”. En: *Formal ontology in conceptual analysis and knowledge representation*. The Netherlands: Kluwer Academic Publisher, 1993.

Gruber, Thomas R. “A translation approach to portable ontology specifications”. En: *Knowledge acquisition*, 1993, v. 5, n. 2, pp. 199-220.

Guerrero-Bote, Vicente; Lozano-Tello, Adolfo. “Vínculos entre las ontologías y la biblioteconomía y documentación”. En: *La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información (Actas del IV Congreso ISKO-España Eoconsid '99)*. Granada: ISKO-Facultad de Biblioteconomía y Documentación, 1999, pp. 25-31.

Hidalgo-Goyanes, Paloma. “Documentación audiovisual”. En: García-Gutiérrez, Antonio (ed.). *Introducción a la documentación informativa y periodística*. Sevilla: MAD, 1999, pp. 473-485, Isbn 84-8311-460-7.

López-de-Quintana, Eugenio. “Documentación en televisión”. En: *Manual de documentación informativa*. Madrid: Cátedra, 2000, pp. 83-181, Isbn 84-376-1798-7.

Noy, Natalya F.; McGuinness, Deborah L. *Ontology development 101: a guide to creating your first ontology*, 2000.  
<http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinnessabstract.html>

Jorge Caldera-Serrano, Departamento de Información y Comunicación, Universidad de Extremadura, Badajoz.  
[jcalser@alcazaba.unex.es](mailto:jcalser@alcazaba.unex.es)

Rodrigo Sánchez-Jiménez, Departamento de Biblioteconomía y Documentación, Universidad Complutense de Madrid.  
[rsanchezj@ccinf.ucm.es](mailto:rsanchezj@ccinf.ucm.es)

# Ontología del patrimonio de Cantabria

Por Francisca Hernández-Carrascal

**Resumen:** *Experiencia de aplicación de ontologías al dominio patrimonio cultural, el proyecto Ontología del patrimonio de Cantabria está desarrollado y financiado por la Fundación Marcelino Botín. Se presentan los antecedentes del proyecto y una descripción del mismo, destacando el portal semántico y las diferentes aplicaciones que se incluirán en él, como elementos visibles. Se hace una revisión de la normativa internacional utilizada, desde las recomendaciones del W3C para la web semántica a normas específicas para la creación de una ontología sobre patrimonio cultural como la norma ISO 21127 y otros modelos de datos relacionados como FRBR y Dublin Core. Se describe la metodología seguida para la integración de datos procedentes de fuentes diversas y heterogéneas y los procesos de conversión realizados para crear una base de datos de conocimientos sobre el patrimonio de Cantabria.*

**Palabras clave:** *Ontologías, Patrimonio Cultural, Integración de conocimientos, ISO 21127, Modelo Conceptual de Referencia del Cidoc, Cidoc-CRM, Requisitos funcionales para registros bibliográficos orientados a objetos, FRBRoo*

**Title:** **Cantabria's Heritage Ontology**

**Abstract:** *Description of an ongoing project that aims to model the cultural heritage domain in an ontology, the Cantabria's Heritage Ontology, lead by Fundación Marcelino Botín. The ontology will contain data from different and heterogeneous sources (bibliographic records, finding aids, catalogues, etc.). The project will develop a series of applications that get the most out of the ontology, beginning by a semantic portal. The portal will provide easy and intuitive access to a huge volume of information that comes from heterogeneous sources selected by domain experts. People interested in Cantabria will be able to find with as few clicks as possible, not only the information they are interested in, but also some contextual information that complements the one the user is watching, suggesting new navigation directions, and uncovering hidden relations. The ontology has been designed based on the international standards for cultural models, such as CRM, FRBR or Dublin Core, and W3C recommendations for the semantic web.*

**Keywords:** *Ontologies, Cultural heritage, Knowledge integration, ISO 21127, Cidoc conceptual reference model, Cidoc CRM, Functional requirements for bibliographic records object oriented, FRBRoo.*

**Hernández-Carrascal, Francisca.** "Ontología del patrimonio de Cantabria". En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, pp. 92-98.

DOI: 10.3145/epi.2008.ene.11



**Francisca Hernández-Carrascal**, Miembro del Cuerpo Facultativo de Bibliotecas. Ha sido directora del Departamento de Control Bibliográfico de la Biblioteca Nacional (1993-1996), directora de la Hemeroteca Nacional (1996-1999) y directora técnica del Archivo Virtual de la Edad de Plata (1999-2003). Ha participado en diferentes proyectos, nacionales e internacionales, relacionados con ontologías y actualmente es la directora del proyecto Ontología del patrimonio de Cantabria que lidera la Fundación Marcelino Botín.

## Descripción

**EL PROYECTO DE LA ONTOLOGÍA del patrimonio de Cantabria, liderado por la Fundación Marcelino Botín se inició en 2006 y finalizará en 2009. Cuenta con la colaboración del Departamento de Historia Moderna y Contemporánea de la Universidad de Cantabria que ha formado un equipo de investigación dirigido por Manuel Suárez-Cortina y coordinado por Gonzalo Capellán. Este equipo de 'expertos en el dominio' está actualizando la recopilación bibliográfica base Historia**

**de Cantabria: un siglo de historiografía y bibliografía (1900-1994)<sup>1</sup> editada por Manuel Suárez Cortina<sup>2</sup> sobre la que se está construyendo el edificio de la ontología.**

La empresa española *iSOCO* (*Intelligent Software for the Networked Economy*) es la encargada de la construcción e implantación del sistema.

El objetivo del proyecto es aglutinar de una forma estructurada el conocimiento sobre el patrimonio de Cantabria en todos sus múltiples aspectos: desde el arqueológico al industrial, desde el científico-cultu-

ral al etnográfico, pasando por sus actores y períodos históricos. Es un campo muy amplio y heterogéneo y está disponible de forma fragmentada, procede de fuentes como publicaciones oficiales, monografías, artículos, catálogos, inventarios, enciclopedias, bases de datos, sitios web, etc., y se encuentra en diferentes estados de formalización y digitalización, todo lo cual dificulta el acceso a la información y su explotación a través de la Web.

Basándose en las tecnologías de la web semántica, el proyecto está elaborando una base de conoci-

mientos (ontología) procedentes de las diferentes áreas científicas y técnicas relacionadas con el patrimonio (la historia, la geografía, el arte, la gestión del patrimonio, etc.) que han sido, y son, generados por diferentes personas y entidades en el ejercicio de sus actividades (instituciones de la administración pública, universidades, entidades privadas, científicos, escritores, etc.), para ofrecerlos en forma de un portal web del patrimonio de Cantabria.

### **Qué es y para qué sirve una ontología**

Como se sabe, las ontologías definen los conceptos y las relaciones que representan un área de conocimiento. Categorizan las condiciones de uso de los términos y sus posibles relaciones, así como las restricciones en el uso de las mismas. En la práctica las ontologías pueden ser muy complejas<sup>3</sup> (con unos varios miles de términos y condiciones) o muy simples (la descripción de uno o dos conceptos solamente). Así, la Ontología del patrimonio de Cantabria contempla los diferentes tipos de patrimonio tal y como están establecidos en la legislación vigente en Cantabria, los objetos, las piezas, los elementos que componen ese patrimonio; las publicaciones que se han editado a lo largo de la flecha del tiempo para difundir, estudiar, analizar, proteger y gestionar el patrimonio; las instituciones que lo gestionan, sus cometidos, sus actuaciones concretas; las personas que han trabajado, investigado, conservado, restaurado y difundido ese patrimonio. Todo ello exige la definición de las clases apropiadas para expresar entidades temporales como períodos cronológicos, períodos culturales, acciones, etc.; objetos biológicos y objetos realizados por el hombre, etc.

El amplio conjunto de datos relacionados con el patrimonio ordenados semánticamente proporcionarán a cualquier tipo de usuario

una gran potencia de búsqueda y navegación, así como le permitirán proyectar o extrapolar el citado conocimiento a actuaciones de todo tipo, desde la recuperación de un dato concreto a la gestión de un proyecto cultural, o a la promoción de un servicio turístico, y, esto no es lo menos importante, a la realización de un modelo de conocimiento transferible a otros ámbitos territoriales, Comunidades Autónomas o países.

Se quiere también con este sistema de organización de la información contribuir a la potencia y versatilidad de recuperación de la información que pretende la web semántica, en la que los agentes de software harán uso del conocimiento vertido en un entramado de ontologías. Del mismo modo que la web actual ha adquirido su potencia por medio de la creación y transformación de contenidos desde todos los puntos del planeta, la web semántica podrá mostrar su verdadero potencial con un esfuerzo similar de estructuración de información y generación de conocimientos. Los métodos y procedimientos, así como las aplicaciones que se están llevando a cabo en este sentido son muy variados, lo que se refleja, entre otros, en *Semantic Web Education and Outreach Interest Group: Case Studies and Use Cases*<sup>4</sup>, página del W3C destinada a casos de estudio y uso en producción relacionados con web semántica en el ámbito cultural<sup>5</sup>, en la que está incluido el proyecto *Ontología del patrimonio de Cantabria*<sup>6</sup>.

### **Fuentes de información**

Dentro del proceso de construir la ontología está la recopilación del conocimiento e información publicado sobre el patrimonio de Cantabria. Se están siguiendo las técnicas de la bibliografía para la categorización y organización de los diferentes documentos que puedan existir sobre un determinado tema,

desde un nivel superior en el que se encuentran las bibliografías de bibliografías, hasta el nivel más profundo en el que se situarían los documentos primarios, bien entendido que en la sociedad actual esto supone la identificación y registro de documentos que pueden encontrarse en múltiples soportes, incluido el electrónico, o estar organizados en bases de datos. Así, en un primer nivel se dispondría el conocimiento publicado en el que entrarían todas aquellas fuentes (bio-bibliografías, bibliografías especializadas, bibliografías de publicaciones periódicas, biografías, diccionarios y enciclopedias, recopilaciones de fuentes documentales para la historia de Cantabria, directorios, catálogos y bases de datos bibliográficas, etc.) que en su conjunto recopilan todo lo publicado sobre Cantabria. Esto equivale a decir, lo que será muy significativo, todo lo estudiado sobre esta Comunidad. Por ello, y como ya se ha dicho, la elección de la obra *Historia de Cantabria: un siglo de historiografía y bibliografía (1900-1994)* no es un mero punto de inicio sino los cimientos de la ontología.

En un segundo nivel debemos considerar las descripciones de las unidades (materiales bibliográficos, documentos de archivo, bienes de interés cultural muebles e inmuebles, etc.). En este punto los catálogos de bibliotecas y museos; los registros, inventarios y catálogos de bienes de patrimonio; las guías, inventarios e instrumentos de descripción de archivos, etc., serán las fuentes fundamentales de documentación. No deben olvidarse en este punto las estructuras de metadatos normalizadas en las que se transmite esta información (formato Ibermarc, MARC 21, EAD, EAG), puesto que se tendrá que trabajar con ellas. Ni, por supuesto debe dejarse de lado, que muchas de estas fuentes se hallan en bases de datos no normalizadas, en formato texto sin estructura alguna o incluso ni siquiera existen

en forma digital. Esta misma situación es la que podemos encontrar en el tercer nivel, el conocimiento sobre la gestión del patrimonio. En este apartado podemos situar la legislación sobre el patrimonio, los diferentes tipos de administraciones que actúan sobre él y sus instrumentos de difusión y actuación (programas, proyectos, etc.). Por último, en un nivel de conocimiento general, se han identificado las fuentes de información (y sus correspondientes estructuras de información) que permitirán obtener datos concretos sobre lugares geográficos, personas, cronologías, instituciones históricas y actuales, etc.

## Aplicaciones

Construidas sobre la ontología se dispondrá de una serie de aplicaciones que se irán finalizando en el portal a lo largo del proyecto y que permitirán una explotación específica de los datos. Un buscador semántico permitirá búsquedas que habitualmente no pueden realizarse como "libros impresos entre el 1907-1939, escritos por autores nacidos en Santander" y mediante un mapa interactivo se podrán geoposicionar todos los datos del patrimonio de Cantabria, así como establecer rutas entre diferentes puntos. A partir del conocimiento acumulado podrán crearse recursos didácticos digitales, construyendo automáticamente cursos y exámenes sobre los distintos movimientos artísticos, yacimientos, espacios naturales o conjuntos históricos. De los datos obtenidos de proyectos sobre el patrimonio se construirá una herramienta que analice las acciones ya realizadas, y extrapole y realice estimaciones de propuestas concretas de actuación. Igualmente, una aplicación permitirá configurar rutas turísticas según el interés de los usuarios; completar aplicaciones de navegación de vehículos (GPS) con información de interés sobre la zona visitada o aplicaciones automáticas de guías turísticas



Figura 1



Figura 2

para completar la información sobre puntos visitados (transferibles a dispositivos móviles). Por último se generará una Wikipedia semántica con los términos de la ontología con funcionalidades que permitan la actualización de datos, la aportación de comentarios, informaciones, documentos, etc.

## Normativa

No es el objetivo de este artículo entrar en el detalle de los lenguajes de marcado necesarios para expresar ontologías. Por supuesto, el proyecto se vale de las recomendaciones del W3C para la web semántica<sup>7</sup>, y está utilizando RDF y OWL. Inicialmente RDF fue creado como

lenguaje de XML para la expresión de relaciones semánticas y OWL como lenguaje específico para la construcción y transmisión de ontologías en la web. Sin embargo, hay que considerar que el software propio de la web semántica (piezas de software, aplicaciones, librerías, etc.) está haciendo uso de ambos por su capacidad para ser reutilizado en diferentes entornos y aplicaciones. Hay que mencionar que dentro de OWL<sup>8</sup> existen diferentes sabores como OWL Lite, OWL DL, o OWL Full y que algunas aplicaciones específicas como los razonadores funcionan sólo con uno de los sabores mencionados.

Ha sido preciso definir una estructura común para las informaciones y datos mencionados para lo cual, se tomaron los dos modelos más sólidos que existen: los *Functional requirements for bibliographic records* de IFLA y el *Cidoc conceptual reference model (CRM)*<sup>9</sup>. Éste fue elaborado por el *Comité Internacional de Documentación (Cidoc)* del *Consejo Internacional de Museos*, que se transformó en 2006 en la norma *ISO 21127 Information and documentation-A reference ontology for the interchange of cultural heritage information*. La norma ISO 21127 proporciona en una estructura formal las definiciones necesarias para describir los conceptos y relaciones usadas en la documentación sobre patrimonio cultural. Constituye un marco semántico común y extensible al que puede ser convertido cualquier información sobre patrimonio cultural y es una guía para el modelado conceptual de las entidades (*entities*) que forman el patrimonio cultural.

Se ha considerado imprescindible seguir esta norma, primero para asegurar la interoperabilidad de esta ontología con otras análogas, y segundo para poder transferir y ser objeto de transferencia de conocimientos desarrollados por terceros,

incluido el software y los aplicativos elaborados conforme a la referida norma ISO. En otras palabras, su uso permite la transformación de diferentes fuentes en una ontología común e interoperable, lo que es sustancial a la web semántica.

En tanto que norma internacional, permite la interconexión de diferentes ontologías del patrimonio creadas en torno a diferentes ámbitos temáticos, geográficos o administrativos. Es perfectamente razonable pensar una ontología global sobre el patrimonio formada por la suma de ontologías parciales de ámbitos más restringidos. Este es por otra parte el objetivo que anima la web semántica en sí.

En la Figura 3 puede verse la estructura de CRM con las clases y relaciones de más alto nivel del modelo<sup>10</sup>.

La norma *ISO 21127* reconoce la necesidad de desarrollar determinados aspectos de la ontología y prefigura el modo en el que esto puede llevarse a cabo. Indudablemente en el marco de la Ontología del patrimonio de Cantabria ha sido necesario concretar nuevas clases, siempre incluidas dentro de alguna de las definidas en la norma, para

expresar, por ejemplo, relaciones familiares o profesionales entre personas; ocupaciones, profesiones y relaciones de personas con entidades.

Por otro lado el modelo conceptual del CRM ha tenido desde sus inicios una fuerte influencia museística. Como ya se ha comentado están muy detalladas determinadas acciones, pero no así otras, como los aspectos bibliográficos. Obviamente el mejor análisis de la información bibliográfica es el establecido en los *Requisitos funcionales para registros bibliográficos (FRBR)*<sup>11</sup> y en consecuencia, en 2003 se formó un grupo de trabajo, Working Group on *FRBR/CRM Dialogue*<sup>12</sup>, con el objetivo de una diseñar una formulación orientada a objetos de FRBR que actuara como extensión de CRM. En mayo de 2007, aunque ya existía alguna versión anterior, se publicó el borrador de la versión 0.8.1c de *FRBR<sub>OO</sub>* (denominación que la distingue de la versión entidad-relación o *FRBR<sub>ER</sub>*). Hay que mencionar que la incorporación de esta extensión es conforme con la norma *ISO 21127* puesto que las clases específicas definidas en *FRBR* (p. e., F1.Work, F2.Expression, F3.Manifestatio\_Product\_Type o

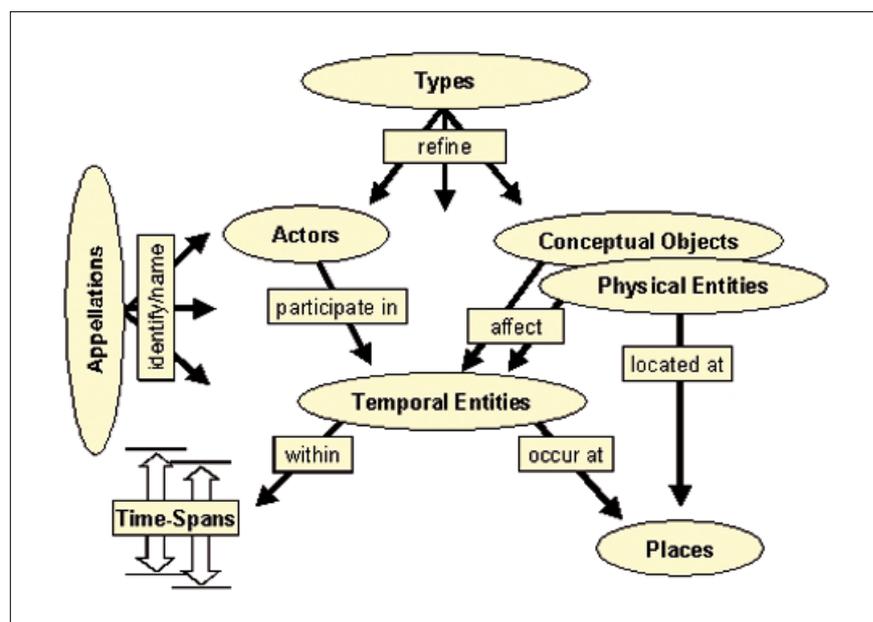


Figura 3

F5.Item) son subclases de clases ya definidas en *CRM* o bien se trata de clases equivalentes. La Ontología del patrimonio de Cantabria ha incorporado esta extensión de *CRM* a la ontología cubriendo así el modelado de la información bibliográfica.

En la concepción del *CRM* y, por tanto, de la norma *ISO 21127*, estaba la posibilidad de convertir datos procedentes de archivos, bibliotecas y museos. Así se han realizado propuestas, más bien teóricas, de transformación desde *Dublin Core*, *Encoded Archival Description*, *MDA Spectrum*, etc. Esto es evidente puesto que el conocimiento de un dominio se obtiene tanto del modelado o estructuración de ese dominio (ontología) como de los datos concretos o instancias que se incorporan a la ontología y esos datos concretos se encuentran en registros con esas estructuras de información. Es más, aunque se trate de procesos completamente diferentes intelectual y técnicamente, son insolubles; una ontología está formada tanto por la estructura de la información, el modelado del dominio, como por los datos propios de ese dominio. El trabajo de armonización entre *CRM* y *FRBR*, no disponible al inicio del proyecto y que todavía se encuentra en estado de borrador, no incluye la definición de los mapeos de información codificada. O mejor dicho, pueden encontrarse propuestas de definición de conversiones (desde *EAD*, *DC*, etc.) previas a la fusión *CRM/FRBR*, que ha sido necesario rehacer en el proyecto, puesto que la inclusión de algunas clases de *FRBR* afecta no sólo a los materiales bibliográficos impresos, sino también a los manuscritos y documentos de archivo.

*Dublin Core (DC)* representa un caso especial dentro de la Ontología del patrimonio de Cantabria porque es una de las estructuras de información en la que están defini-

dos numerosos registros de interés para el patrimonio de Cantabria (pensemos que es la norma de codificación de registros que sustenta los repositorios *OAI*), pero también se utiliza para la gestión de la ontología describiendo cada uno de sus elementos (clases, propiedades e instancias) como recursos informativos. Así, cada instancia tiene asociado a través de *DC* su autor, su título, su fecha de creación, la fuente de procedencia, etc., lo cual permite realizar controles típicos como el seguimiento de un proceso o un programa de conversión o la procedencia de un determinado dato.

## Avances

**Adquisición.** Para la integración masiva de datos en la ontología se ha implementado una serie de herramientas de adquisición de datos, cada una de ellas relacionadas con estructuras típicas de información y compuestas por un conjunto específico de reglas según el tipo de fuente. Así, puede importarse información a partir de sitios web semiestructurados (p. e., tablas), de bases de datos *ODBC* o *JDBC*, de registros en formato *Ibermarc* o *MARC 21* (registros bibliográficos, de autoridades y de fondos y localizaciones), de registros *Dublin Core*, y a partir de ficheros *EAD* o *EAG*<sup>13</sup>. En la actualidad la Ontología del patrimonio de Cantabria está formada por 153 clases y 410 propiedades, así como por más de 110.000 instancias procedentes de registros bibliográficos *MARC 21*, registros *Dublin Core*, y otras fuentes como bases de datos relacionadas, sitios web, etc.

**Registros poco informados.** Hay que mencionar en este punto algunas cuestiones importantes para la gestión de una ontología que no son realmente nuevas porque emergen en cualquier proceso de conversión, como son la calidad y granularidad de los registros, una excesiva flexibilidad en el uso de los

formatos de codificación y una estructuración muy débil en las bases de datos. Por ejemplo, la tendencia a elaborar registros bibliográficos cada vez más reducidos (también en tiempo y coste) hace que estos pierdan información que puede ser sustancial y al menos debería considerarse que uno de los usos de estos registros va más allá de la gestión y control de los procesos de circulación en bibliotecas y que pueden ser reutilizados en otros sistemas de información. Esperemos que el nuevo código de catalogación internacional<sup>14</sup> tenga en cuenta estos usos.

**Duplicados.** Como puede suponerse, al utilizar diversas fuentes de información es frecuente la aparición de instancias duplicadas, para lo que se ha diseñado un programa específico que detecta homonimias y sinonimias, que pueda ser aplicable en los distintos estados de evolución de la ontología.

**Operaciones con fechas.** Algunas de las aplicaciones diseñadas deberán realizar operaciones (sucedió al mismo tiempo, es anterior a, es posterior a, etc.) para lo que se ha desarrollado un programa para la expresión de fechas en un único formato al que pueden trasladarse fechas aproximadas o indeterminadas del tipo 'finales del siglo XX', 'a mediados de la década de los años 60', o bien fechas anteriores o posteriores a Cristo, así como distintas expresiones de indeterminación como 188?. Indudablemente, a lo largo del proyecto se irá completando la definición de diferentes reglas<sup>15</sup>, tanto de análisis de la ontología para la detección de homonimias y sinonimias o para la detección de relaciones no explícitas, como para la visualización, exportación o explotación de los datos

**Visualización.** En cuanto a la visualización de información en la web se ha establecido a través de una ontología de publicación una serie de entidades complejas bási-

cas (periodos cronológicos, periodos culturales, personas, entidades, lugares, temas, y unidades de patrimonio) para las cuales se han definido los elementos básicos de visualización (nombre, fecha de nacimiento y muerte, profesión, actividades, obras, etc.), las preguntas en Sparql<sup>16</sup> que permiten la obtención de datos de clases y propiedades próximas (p. e., los autores cántabros se forman por medio de una ‘query’ que extrae todas las instancias de la clase persona que han sido o son escritores y que han nacido en Cantabria), así como los elementos de navegación en la web. Es preciso decir que si bien todas las clases y propiedades de la ontología son necesarias para estructurar el conocimiento, no todas deben ser directamente visibles en una interfaz web en aras de una mejor usabilidad, comprensión y navegabilidad. De todas formas, el portal web cuenta con una visualización, destinada a usuarios concretos, que permite ver y navegar por cualquiera de los elementos de la ontología. Entre paréntesis, no puede dejarse de mencionar que hace unos años se consideraba absurdo que la información bibliográfica pudiera visualizarse en formato MARC mientras que ahora, todas las grandes bases de datos bibliográficas ofrecen esa posibilidad.

### Edición y gestión

Entre las dificultades más importantes encontradas está el estado de desarrollo de los editores de ontologías, principalmente destinados a la creación y edición de la estructura de información (las clases, propiedades y restricciones) de una ontología, pero no tanto a la gestión y manipulación de gran cantidad de instancias. Los desarrollos ya efectuados en la visualización de instancias en el portal han superado en parte esta dificultad, pero quedan algunas cosas por mejorar en herramientas de base como *Protégé* o *SemanticWorks*).

## Conclusiones

Conviene insistir en el hecho enormemente significativo de que el CRM está ya caracterizado y normalizado en la ISO 21127 lo cual da a todo este trabajo una estabilidad sumamente grande. Es decir, que es posible, dentro de una estrategia de transferencia tecnológica, crear una metaestructura fruto del análisis FRBR/CRM y del mapeo de estructuras normalizadas como MARC 21, y tampoco sería particularmente costoso el esfuerzo que habría que realizar, para trasladar con facilidad este intangible –sumamente laborioso desde el punto de vista de su realización– a otro territorio, a otra Comunidad Autónoma o a otro país, a un estilo arquitectónico, a una escuela literaria o a la recuperación de un tipo concreto de patrimonio. Se ha logrado formalizar e implementar en un sistema automatizado el conocimiento no ya de los expertos que participan en el proyecto, es decir la *Fundación Marcelino Botín*, la *Universidad de Cantabria* y la empresa *iSOCCO*, sino también el de los distintos expertos que han elaborado y formalizado tanto el FRBR como el CRM (por no hablar del formato MARC).

Una de las pruebas para poner de manifiesto la utilidad de todo este trabajo es su tratamiento de información real. En este caso se ha procedido a llevar a cabo pruebas con diferentes catálogos impresos, como por ejemplo el Catálogo de publicaciones periódicas de la *Biblioteca Municipal de Santander* o el *Catálogo de Manuscritos Musicales de Cantabria*.

Naturalmente, la vinculación de este tipo de registros con las grabaciones de un grupo de música gregoriana de los citados manuscritos litúrgicos mostrará la enorme posibilidad informativa de un procedimiento de este tipo y de la metodología ya desarrollada. En este sentido serán muy importantes los mecanismos de colaboración que se

establezcan con diferentes entidades, tanto para la transformación de los datos, estructurados o no, como para la elaboración de programas complementarios conjuntos de digitalización y conversión.

**Ejemplo.** Si bien en la entrada Cabezón de la Sal de la *Gran Enciclopedia de Cantabria* no aparece la relación de escritores nacidos en esta localidad, como tenemos a lo largo de la *Enciclopedia* múltiples entradas que hacen referencia a autores montañeses nacidos en este pueblo, tras crearse las correspondientes instancias podrán suministrarnos de forma inmediata una respuesta a esa cuestión. Pero además también sabremos la música compuesta en torno a esta ciudad, los empresarios, científicos, artistas, políticos... nacidos en ella o que en ella llevaron a cabo su trabajo más relevante, y todo ello establecido a través de la flecha del tiempo. Tendremos en esa misma línea cronológica los procesos comerciales, industriales y más importantes relacionados con Cabezón así como el patrimonio más destacado. Como además pueden establecerse conversiones desde otras fuentes (p. e., desde el *Nomenclator* del INE) y bases de datos geográficas e históricas tendremos, a través del conocimiento estructurado en el sistema, sus representaciones gráficas, cartográficas, sus vías de comunicación tanto actuales como históricas y sus comercios, actividades y vinculaciones con la realidad sociológica que gira en torno a esa entrada. Quien dice un topónimo dice un autor o una obra de arte, o una familia, o un género, o una especie de la fauna y de la flora, o un monumento. Lógicamente será necesario ir ajustando estos procesos a todas las fuentes de información (algunas inéditas como la *Bibliografía de la provincia de Santander*, de Leguina) como las *Guías*, de Coll, los trabajos de erudición local, los nuevos planteamientos técnicos de las universidades, en particular de

la de Cantabria, o las actuaciones históricas y actuales como las de la Comunidad Autónoma.

*“In the Semantic Web, it is not the Semantic which is new, it is the Web which is new”* **Chris Welty, IBM**

### Notas

1. **Suárez-Cortina, Manuel**, ed. *Historia de Cantabria: un siglo de historiografía y bibliografía (1900-1994)*. Santander: Fundación Marcelino Botín, 1995.

2. En *Las bibliografías regionales y locales españolas: (evolución histórica y situación actual)*. Madrid: **Ollero y Ramos**, 2003, **Juan Delgado-Casado** dice: “no es una historia sino una bibliografía sobre la región que incluye casi 8.000 referencias, centrada sobre todo en textos históricos pero sin evitar la descripción de obras de todo tipo y especialmente relacionadas con el arte y los monumentos... Me parece que es el reperto-

rio fundamental sobre esta región y una obra de importancia indiscutible.

3. The Gene Ontology.  
<http://www.geneontology.org>

4. <http://www.w3.org/2001/sw/sweo/public/Use-Cases>

5. El 12 de noviembre de 2007 se convocó en Corea el *First international workshop on cultural heritage on the semantic web*, en el marco de la 6<sup>th</sup> *International semantic web conference*.

6. <http://www.w3.org/2001/sw/sweo/public/Use-Cases/FoundationBotin>

7. <http://www.w3.org/2004/01/sws-pressrelease.html.en>

8. <http://www.w3.org/TR/2004/REC-owl-features-20040210/#s4>

9. <http://cidoc.ics.forth.gr>

10. Tomada de: **Crofts, Nick; Doerr, Martin; Gill, Tony**. *The Cidoc conceptual reference model: a standard for communicating cultural contents*. February 2003.  
<http://www.cultivate-int.org/issue9/chios>

11. <http://travesia.mcu.es/documentos/requisitos.pdf>

12. [http://www.ifla.org/VII/s13/wgfrbr/FRBR-CRMdialogue\\_wg.htm](http://www.ifla.org/VII/s13/wgfrbr/FRBR-CRMdialogue_wg.htm)

13. *EAG (Encoded archival guide)* es una DTD para la elaboración del Censo-Guía de los archivos españoles.

[http://aer.mcu.es/sgae/jsp/censo\\_guia/seg\\_nivel/presentacion\\_proyecto.html](http://aer.mcu.es/sgae/jsp/censo_guia/seg_nivel/presentacion_proyecto.html)

14. Resource Description and Access (RDA).  
<http://www.collectionscanada.ca/jsc/rda.html>

15. <http://www.w3.org/2005/rules/>

16. El *Rule Interchange Format Working Group* del W3C trabaja actualmente en una definición de un lenguaje común de reglas, su utilización e intercambio entre sistemas, y su aplicación a OWL y RDF.  
<http://www.w3.org/TR/rdf-sparql-query/>

**Francisca Hernández-Carrascal**,  
Fundación Marcelino Botín  
[hdz@fundacionmbotin.org](mailto:hdz@fundacionmbotin.org)



ANUARIO

# Think

## EPI

### Artículos en la sección CONTENIDOS

#### **Pero, ¿qué es Google?**

*Francisco Tosete*

#### **Web y buscadores: ¿una medida de la realidad?**

*Isidro F. Aguillo*

#### **Función de los portales temáticos en la era de la búsqueda posmoderna**

*Ernest Abadal y Lluís Codina*

#### **OCLC: 21.985 miembros ¿cuántos españoles?**

*Adela d'Alòs-Moner y Eva Méndez*

#### **Prensa digital en 2007**

*Javier Guallar*

Más información en <http://www.thinkepi.net/repositorio>

Ya puedes pasarnos tu pedido: [epi@sarenet.es](mailto:epi@sarenet.es)

**Anuario ThinkEPI 2008**

89,42 € + IVA = 93 €

**Anuario ThinkEPI 2007 + 2008**

115,39 € + IVA = 120 €

**Anuario ThinkEPI**

<http://www.thinkepi.net>

# Cómo realizar tests de usabilidad con personas ciegas

Por Mireia Ribera, Miquel Térmens y Maika García-Martín

**Resumen:** Se presentan una serie de recomendaciones prácticas para realizar tests de usabilidad con personas ciegas. Tanto la planificación, como la ejecución y el análisis posterior son comentados en detalle para aconsejar un conjunto de buenas prácticas. El artículo finaliza con algunas reflexiones críticas sobre la pretendida discapacidad de los usuarios ciegos.

**Palabras:** Test de usabilidad, Test de accesibilidad, JAWS, Sintetizadores de voz, Usuarios ciegos.

**Title:** Usability testing for blind users

**Abstract:** Practical recommendations are offered for usability testing to meet the needs of users who are blind, including a detailed explanation of best practices in planning, implementation, and evaluation. The article concludes with critical reflections on the supposed limitations of blind users.

**Keywords:** Usability testing, Accessibility testing, JAWS, Screen readers, Blind users

**Ribera, Mireia; Térmens, Miquel; García-Martín, Maika.** "Cómo realizar tests de usabilidad con personas ciegas". En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, pp. 99-105.

DOI: 10.3145/epi.2008.ene.12



**Mireia Ribera Turró**, diplomada en biblioteconomía y documentación e ingeniera informática. Es profesora del Departamento de Biblioteconomía y Documentación de la Universitat de Barcelona. Su área de interés es la accesibilidad digital y ha coordinado la primera traducción autorizada del Consorcio W3C, las WCAG 1.0 al catalán.

**Miquel Térmens Graells**, doctor en documentación, licenciado en historia y diplomado en biblioteconomía y documentación. Es profesor del Departamento de Biblioteconomía y Documentación de la Universitat de Barcelona. Sus líneas de investigación se centran en el uso y la preservación de los documentos digitales y la gestión de bibliotecas digitales.

**Maika García Martín**, licenciada en biblioteconomía y documentación y máster en gestión de contenidos digitales. Es profesora del Departamento de Biblioteconomía y Documentación de la Universitat de Barcelona y consultora de usabilidad en Xperience Consulting. Sus áreas de interés son la usabilidad, la accesibilidad digital y la arquitectura de la información.

## 1. Introducción

LA ACCESIBILIDAD DIGITAL EN LOS ÚLTIMOS AÑOS ha pasado de ser un valor añadido a ser un requisito. En el ámbito legal, en el caso de España, hemos pasado de tener leyes que eran declaraciones de buenas intenciones, como la *Lssice* (Lssice, 2002), a una legislación con cada vez más concreciones, como la *Liondau* (Liondau, 2003), hasta llegar a la exigencia actual de garantizar la accesibilidad de los discapacitados en sus relaciones con la administración (Real Decreto, 2007).

En el ámbito normativo las directrices WCAG (*web content ac-*

*cessibility guidelines*), publicadas como pautas de seguimiento recomendable para los creadores de contenido, se han incluido en legislaciones (*United States Access Board*) y se han convertido en normas (Aenor, 2004). El diseño accesible, y que sigue los estándares, actualmente forma parte integral del pliego de requisitos habituales de una web.

En este entorno de creciente demanda son diversas las voces que claman por una redefinición de la accesibilidad (Mankoff; Fait; Tran, 2005; Brajnik, 2006). La visión simplista de equiparar accesibilidad con el cumplimiento de unas

pautas se pone en cuestión (Sloan, 2006) y se recupera la visión más completa de la accesibilidad dentro del marco de la usabilidad (Henry, 2007).

Que una web sea accesible, desde este nuevo punto de vista, significa que sea usable para personas con discapacidades. Y las pautas son tan sólo un punto de partida en el camino hacia este objetivo. Para garantizar una accesibilidad completa se deben incorporar otros métodos del diseño centrado en el usuario; uno de ellos, clave para entender a los usuarios, es el test de usabilidad.

---

**“La idea simplista de equiparar accesibilidad con el cumplimiento de unas pautas se pone en cuestión y se recupera la visión más completa de la accesibilidad dentro del marco de la usabilidad”**

---

Este artículo quiere contribuir a facilitar este camino dando algunos consejos prácticos para realizarlos con invidentes, a partir de la experiencia de los autores en un caso real realizado en Barcelona con 30 usuarios ciegos (Termens, et al., 2006). Como complemento se adjunta un pequeño listado comentado de recursos para ayudar a los diseñadores web o creadores de recursos digitales.

## **2. Fase de definición del test de usuarios**

El primer paso para realizar un test de usuarios es la planificación. El tiempo de los usuarios es valioso y caro, por lo que hay que aprovecharlo al máximo. Algunas de las especificidades de los tests con personas ciegas que afectan la fase de planificación son las siguientes:

– Test ecológico/test con formación previa. Una de las grandes decisiones en la etapa de planificación es saber si se desea realizar un test ecológico, es decir, sin alterar las condiciones naturales aunque haya más variabilidad, o un test alrededor de unas condiciones dadas, que se prevé que se van a cumplir en el entorno a diseñar. En el caso de las personas ciegas, una de las opciones principales a alterar es el conocimiento de las posibilidades del programa JAWS, el lector de pantalla más usado en España y que ha ido incrementando sus funcionalidades en las últimas versiones (la más reciente versión comercial es la 6.20), incorporando teclas rápidas

para desplazarse a saltos, navegar más rápidamente, etc. Ahora bien, dado su coste, muchos usuarios aún utilizan versiones antiguas (3.7 es la más habitual), que no ofrecen estas posibilidades. Incluso en el caso de las personas ciegas que por motivos laborales pueden acceder a las últimas actualizaciones, a menudo no conocen todas sus posibilidades debido a la formación recibida. Si lo que se desea es testear alguna aplicación con las versiones recientes de JAWS habrá que llevar a cabo sesiones previas de formación de los usuarios. En cualquier caso, se deberá tener constancia de la versión del programa utilizada a nivel particular o laboral para tener datos de las plataformas informáticas habituales de consulta.

– Familiarización con las herramientas. Antes de empezar a trabajar con los usuarios es necesario que el equipo de investigación se habitúe con la tecnología de asistencia correspondiente para entender mejor las acciones de estos y los posibles problemas que se puedan encontrar. En este punto, es casi imprescindible contar con una persona especializada en la tecnología que explique al equipo su uso real, las dificultades y los detalles de configuración. También será necesario consultar los manuales de las tecnologías de asistencia y de las aplicaciones a utilizar para conocer las posibilidades. La consulta de los manuales no suele ser suficiente y se deben revisar foros o webs especializados (como por ejemplo *Jawsp*, *Rnib*, *WebAIM*) para conocer *bugs*, o *workarounds* frecuentes en las tareas a realizar. Otros tests similares ya realizados también son una buena fuente de información, si bien no son muy numerosos.

<http://ar.groups.yahoo.com/group/jawsp/>

<http://www.rnib.org.uk>

<http://www.webaim.org/>

– *Screening*. Es una técnica que consiste en que el equipo de

investigación simula la realización de un test usando las tecnologías de asistencia. Presenta serias dificultades pues es difícil usar la misma lógica que los usuarios en el aprendizaje de las herramientas. Por ejemplo, en el caso de JAWS los usuarios reales deben memorizar todas las teclas de accionamiento para usarlo y a menudo sólo recuerdan un número muy reducido de ellas; en cambio, la tendencia del equipo de investigación será la de aprender las funciones más avanzadas y anotarlas para recordarlas. La recomendación es usar esta técnica con un dominio mínimo de los comandos básicos necesarios para la tarea y probarlos realizando el test en las mismas condiciones que los usuarios (por ejemplo con la pantalla apagada y sin ratón).

---

**“El tiempo de los usuarios es valioso y caro, por lo que hay que aprovecharlo al máximo”**

---

– Prueba piloto. Como en todos los tests de usabilidad, idealmente se debe realizar con unos pocos usuarios para detectar situaciones no previstas, necesidades no atendidas o para comprender mejor el funcionamiento de la interacción. Dado que tanto la modalidad de la interacción, las herramientas usadas, como el modelo mental de los usuarios ciegos difieren en gran medida del común entre los facilitadores, en este colectivo esta prueba tiene más importancia. Para sacarle mayor provecho, se deberá cerrar la prueba piloto con una revisión posterior en la que se haga participar al usuario y se le pregunten algunos aspectos de la interacción, problemas surgidos, etc.

– Elección de la ubicación del test. Espacio de los usuarios. Para

## Recursos para el diseño accesible dirigido a personas ciegas

### 1. Manuales

**Clark, Joe.** *Building accessible websites*. 2002. Uno de los libros clásicos sobre accesibilidad, disponible en la web. Su autor está ahora involucrado en el comité PDF/UA, la subversión del formato pdf enfocada a usabilidad y accesibilidad. Algunos de sus escritos recientes han sido muy críticos con las WCAG 2.0.

<http://joelclark.org/book>

**Letourneau, C.; Freed, G.** *Welcome to the curriculum for web content accessibility guidelines 1.0*. 2000. Manual para entender y aplicar las WCAG 1.0.

<http://www.w3.org/WAI/wcag-curric>

**Thatcher, Jim.** *Jim Thatcher.com*. El web personal de un experto en accesibilidad y autor de uno de los clásicos en el área, ofrece una guía sobre cómo hacer una web accesible con consejos prácticos basados en la experiencia.

[http://www.jimthatcher.com/site\\_resources.htm](http://www.jimthatcher.com/site_resources.htm)

### 2. Webs con recursos

*Disability Rights Commission. Website accessibility guidance*. London: The Stationery Office. La comisión para los derechos de los discapacitados es una agencia independiente creada por el parlamento de Inglaterra para promover y vigilar el cumplimiento de los derechos de este colectivo. Su web, además de ser ejemplar, incluye un código de buenas prácticas para crear webs accesibles en el marco de la DDA, la legislación vigente en el Reino Unido.

[http://www.drc-gb.org/library/website\\_accessibility\\_guidance.aspx](http://www.drc-gb.org/library/website_accessibility_guidance.aspx)

*IBM. Web accessibility*. Lista de 16 directrices creadas por IBM para garantizar la accesibilidad web con explicaciones de su motivación, funcionamiento, etc.

<http://www-306.ibm.com/able/guidelines/web/accessweb.html>

*University of Washington. DO-IT Streaming videos with supporting publications*. DO-IT es una iniciativa de la Universidad de Washington para facilitar la inclusión de estudiantes con discapacidades. Su web dispone de interesantes vídeos en los que se muestra diferentes situaciones de acceso a internet con tecnologías de asistencia.

<http://www.washington.edu/doi/Video/>

*WebAim. Web Accessibility in Mind* es una organización sin fines de lucro de EUA que promueve la accesibilidad en la web. Reune un conjunto de recursos y consejos para crear webs y documentos electrónicos accesibles.

<http://www.webaim.org>

### 3. Accesibilidad pdf

**Johnson, Duff.** *Planet pdf. Accessible pdf*. El principal recurso para estar al día en accesibilidad para el formato pdf.

<http://www.planetpdf.com/enterprise/learning-center.asp?ContainerID=1505>

un test ecológico es recomendable realizarlo en su domicilio particular o en su lugar de trabajo (Clark, 2007; Coyne; Nielsen, 2001a). Este enfoque es más informativo (conoceremos más detalles del contexto real de los usuarios donde hacen sus tareas; ellos trabajarán con su configuración personal, etc.) pero menos controlado. En cualquier caso, para garantizar el éxito deberemos:

- a. Saber qué dispositivos utilizan y conocerlos.
- b. Garantizar que dispongan de monitor.
- c. Asegurar que tengan acceso

a internet (si se realiza un test de una web) y/o los programas necesarios para las tareas a ejecutar.

– Espacio controlado. Si lo que se desea es prever el uso de un punto de consulta adaptado, no está fuera de lugar realizar el test en un laboratorio con unas condiciones más pautadas. Este enfoque es más controlado (por ejemplo, es ideal para tests cuantitativos) pero menos realista. Se debe tener en cuenta que muchos usuarios no dominan la configuración de las herramientas que usan, y que en un entorno nuevo no trabajan en las mismas condiciones que en casa o en el

trabajo, donde otra persona les ha personalizado las herramientas. En cualquier caso, para garantizar el éxito deberemos:

- a. Saber cuáles son los dispositivos exactos que utiliza el participante y adquirirlos tanto a nivel de hardware (teclados específicos) como a nivel de software (JAWS u otros).
- b. Proporcionar un medio de transporte que les facilite la llegada al laboratorio o, si se da el caso, que esté cercano a un transporte público conocido por el usuario y ofrecer acompañarle. Algunas personas ciegas son bastante autóno-

#### 4. Artículos

**Lazar, Jonathan; Allen, Aaron; Kleinman, Jason; Malarkey, Chris.** "What frustrates screen reader users on the web: a study of 100 blind users". En: *International journal of human-computer interaction*, 2007, v. 22, n. 3, pp. 247-269.

Conclusiones de la observación de 100 usuarios. Se analizan las causas más frecuentes de problemas en la interacción con sintetizadores de pantalla.

"Special issue: accessibility of web-based information for people with disabilities". En: *Library hi tech*, 2002, v. 2, n. 2.

"Special issue: accessibility of web-based information resources: part 2". En: *Library hi tech*, 2002, v. 2, n. 4.

Monográficos de la revista dedicados a accesibilidad y coordinados por **Axel Schmetzke**.

**Theofanos, M. F.; Redish, J.** "Guidelines for accessible –and usable– web sites: observing users who work with screenreaders". En: *Interacions*, 2003, v. 10, n. 6, pp. 38-51.

Directrices específicas para usuarios ciegos basadas en la observación de 16 usuarios de *JAWS* y *WindowsEyes*.

**Theofanos, M. F.; Redish, J.** "Helping low-vision and other users with web sites that meet their needs: Is one site for all feasible?". En: *Technical communication*, 2004, n. 52, pp. 9-20.

Conclusiones de la observación de 10 usuarios de *Zoomtext* (amplificador de pantalla) y consejos para la creación de contenido accesible.

#### 5. Herramientas

*Ai Squared. Zoomtext 9.1.* El amplificador de pantallas *Zoomtext* es el más usado en España. La mayoría de usuarios amplía entre 1,75x y 3x. Se puede descargar una versión de demostración con una duración de 30 días.

<http://www.aisquared.com/index.cfm>

*Freedom Scientific. JAWS 7.10.*

El lector de pantallas *JAWS* es el más usado en España. Se puede usar en modo demostración, sin pagar la licencia, con la única restricción de un uso limitado a 40 minutos, tras los cuales hay que reiniciar el ordenador para volverlo a utilizar. Su utilización requiere un poco de familiarización con las teclas de comandos.

[http://www.freedomscientific.com/fs\\_downloads/jaws.asp](http://www.freedomscientific.com/fs_downloads/jaws.asp)

*Vision Australia. "Web accessibility toolbar".* Una empresa dedicada a la distribución de productos para personas ciegas y con problemas de visión difunde en su web interesantes recursos para el webmaster. En particular la barra de herramientas es instalable en *Internet Explorer* y *Opera* y es muy útil para verificar el cumplimiento de las WCAG de forma semiautomática; además incluye una sección de simulaciones muy interesante. También en su web podemos encontrar una clara descripción de las principales tecnologías de asistencia para personas ciegas y con baja visión (*Adaptive technology handbook*).

<http://www.visionaustralia.org.au/info.aspx?page=614>

<http://www.visionaustralia.org.au/info.aspx?page=1230>

mas en sus desplazamientos, pero problemas típicos de obras en la calle, o el cambio de recorrido de un transporte público les representan serias barreras para la llegada. En el caso que el desplazamiento se realice en taxi se deberá avisar para que se admitan los perros lazarillo.

c En el cálculo de los costes de transporte y en la adecuación de las instalaciones deberemos tener en cuenta que muchos invidentes a menudo se desplazan con acompañantes.

d. Finalmente, el espacio idealmente debería ser accesible a nivel

arquitectónico, aunque la mayoría de los casos las barreras son superables para las personas ciegas con acompañamiento personal. Algunos aspectos a tener en cuenta son la inexistencia de peldaños en la entrada, la presencia de suelos llanos y regulares y la disponibilidad de ascensor.

– Agradecimiento. Para agradecer su participación es habitual dar una pequeña compensación económica a los participantes en el test. Aunque algunas directrices (**Coyne; Nielsen, 2001a**) recomiendan dar los honorarios en metálico, por cuestiones de control adminis-

trativo los autores optaron por dar tarjetas-regalo de compra, que no presentaron ningún problema.

– Planificación de la duración. Por último, debemos tener en cuenta que un ciego tarda entre dos y tres veces más que un usuario no-ciego para completar la misma tarea, y los requerimientos cognitivos de mantener la atención concentrada por el oído son superiores a los de la vista. Estos hechos provocan que la fatiga, en los tests realizados por este colectivo, sea superior y haya que reducirla con tareas de menor duración y paradas de descanso durante la realización.

### 3. Fase de reclutamiento de usuarios

Como en todos los tests de usabilidad, se debe disponer de un guión preestablecido con las explicaciones e instrucciones a dar a los potenciales participantes desde el primer contacto. En el caso de las personas ciegas se les comunicará por vía oral preferiblemente, mediante una llamada telefónica. Durante los breves minutos de esta llamada hay que informar de todos los aspectos que puedan influir su decisión a formar parte del test y es importante detectar si se cumplen los criterios de inclusión para ser incluido en el estudio.

Finalmente, no es necesario disponer de una versión en braille del resumen de la investigación que se está llevando a cabo ni del documento de consentimiento informado, pues basta con enviar el fichero de texto por correo electrónico con suficiente antelación a los usuarios para que se lo lean en casa con la ayuda del ordenador o de algún amigo o familiar.

El proceso de captación se puede hacer a partir de empresas de reclutamiento de usuarios o lo pueden hacer los propios investigadores del equipo. Si se opta por la primera opción hay que tener en cuenta que el usuario ciego es considerado como un perfil “alto” y de difícil captación, al mismo nivel que un alto ejecutivo. Esto viene justificado por dos motivos. En primer lugar porque se trata de un porcentaje de población reducida. Además, porque no todas las personas ciegas, al igual que no toda la población en general, hacen el mismo uso de las tecnologías de la información. El número de usuarios potenciales es por tanto muy reducido. Por todo ello, el precio de captación de un usuario ciego es entre un 50% a un 100% más caro que el de un perfil habitual.

Si se reclutan los participantes de forma directa, se recomienda

contactar con las asociaciones de discapacitados locales y pedir su colaboración. En el caso de *JAWS*, el equipo investigador intentó conseguir usuarios a través de la lista *Jawsesp*, pero los únicos voluntarios fueron personas del otro lado del Atlántico. A partir de unos pocos usuarios identificados se puede contar un poco con el boca-oreja de los participantes que conocen otras personas que también podrían participar.

### 4. Fase de ejecución

En general las personas ciegas son muy consideradas con las necesidades de las personas no-ciegas. En el test habrá que hacer especial hincapié en que son ellos los que se deben sentir cómodos en esa situación –tanto si se realiza en su domicilio como en un laboratorio especializado– y adaptar el entorno a su gusto.

Habitualmente los invidentes están más acostumbrados a tratar con personas no-ciegas que viceversa. Por tanto, se recomienda a los investigadores que actúen con espontaneidad tanto en la manera de hablar como en el trato.

Recomendamos preguntar de forma natural directamente a la persona ciega cualquier duda en cuanto a comportamiento o expresiones. Según nuestra experiencia, los participantes ciegos no se sienten incómodos en estas situaciones de vacilación de los investigadores, sino todo lo contrario, acostumbran a mostrarle cómo debe actuar y dan buenos consejos. De esta manera es mucho más fácil saber la forma de ofrecerle el brazo para hacerle de guía a una persona invidente o preguntarle si ha “leído” o “visto” una sección de la web sin que esta terminología y/o comportamiento sea ofensiva:

– “Después de cenar me gusta sentarme en el sofá para mirar la tele con mi familia”, comentaba una participante.

– “Cuando ofreces el brazo, debes caminar siempre un paso por delante de la persona ciega, ello nos ayuda a saber si hay una escalera o un desnivel. Eso sí, no permitas nunca que un ciego se cuelgue de tu brazo, debes hacer de guía pero no tirar de nosotros”, aconsejaba otro participante.

---

**“Recomendamos preguntar de forma natural directamente a la persona ciega cualquier duda en cuanto a comportamiento o expresiones”**

---

Algunas personas ciegas llevan un perro lazarillo como acompañante. Se debe conocer el código de comportamiento respecto al animal y asegurar también su comodidad proporcionando un recipiente con agua. Si se le dan galletas, que sean sin calorías, pues algunos de los perros siguen dietas estrictas. Cuando lleva puesto el arnés está trabajando; en estas circunstancias está entrenado para no responder a ningún estímulo exterior puesto que se podría desconcentrar y provocar un incidente a su amo, por lo que se debe obviar su presencia. En el momento que la persona ciega se lo quita, ya puede ser tratado como cualquier animal de compañía.

Si el test se realiza en un laboratorio especializado se recomienda informar al participante de la distribución de la habitación, de los ordenadores, de las cámaras de vídeo u otros aparatos que puedan formar el conjunto tecnológico de la prueba. También es importante que saluden las personas que se encuentran en la sala y conozcan previamente el rol de cada una ellas.

Antes de iniciar la sesión deberemos explicarle que se irán toman-

do notas en papel (u ordenador). Y durante los tiempos muertos que se den en las sesiones, recordarle al usuario que la parada se debe a las notas que se están tomando.

Asimismo hay que comunicarle si se mueve un objeto de lugar (sus objetos personales como abrigo, bolso, bastón o alguna bebida que se le ofrezca), especialmente si el test se realiza en su casa.

En cuanto a la comunicación con el participante, se debe tener muy presente que el único canal es el oído, el mismo con el que él interactúa con el ordenador. Con tal de evitar interferencias y garantizar una buena comunicación se recomienda:

a. Pedirle que silencie el sintetizador de voz durante la lectura de las tareas para facilitar su comprensión.

b. Escribir el comportamiento de los usuarios durante la ejecución de la prueba y realizar las preguntas oportunas al final de las tareas.

c. El protocolo de pensamiento explícito presenta dificultades con el uso de sintetizadores de voz (Chandrasheka et al., 2006). En cualquier caso, se pueden hacer preguntas o pedir explicaciones tras haber realizado el test con resultados más que aceptables, puesto que las personas ciegas suelen tener muy buena memoria y retienen los detalles para explicarlos después.

d. Durante la sesión con ellos hay que dar señales de comunicación verbal para mantener el contacto. Se recomienda hacer uso de expresiones cortas y con poca carga semántica, como por ejemplo: “de acuerdo”, “sí”, “ajá”. En situaciones con personas no-ciegas este tipo de comunicación se resolvería con la comunicación no verbal: miradas de asentimiento y lenguaje corporal.

Si el facilitador se queda en la sala del test, al inicio de la sesión

debe sentarse cerca del participante y evitar cambiar de lugar durante la ejecución de la prueba. Se recomienda escoger un lugar desde donde pueda ver las acciones que realiza el usuario y oír bien el sintetizador de voz.

En el momento de iniciar la prueba debe ofrecerse la posibilidad de ajustar la velocidad de *JAWS* en función de las preferencias del participante. Tiene especial relevancia si el tiempo se toma como indicador de eficiencia. Sin embargo, nuestra experiencia es que por deferencia al facilitador y por dificultades técnicas de configuración, la mayoría de usuarios no modifican la configuración por defecto del programa.

## 5. Infraestructura informática

El uso de tecnologías de asistencia presenta una dificultad adicional en la preparación de la plataforma informática del test. Con las personas ciegas debemos tener un teclado normativo con las teclas de función y controles en posición normalizada, pues tienen memorizada su posición. Debe disponer de relieves en las teclas “f” y “j” para facilitar el posicionamiento de las manos y en algunos casos se ofrecerán relieves adicionales a petición del usuario. Por otra parte, los usuarios ciegos no usan el ratón.

Para el seguimiento del test hay que tener en cuenta que *JAWS* trabaja con un cursor virtual que no se ve por pantalla y cuyos movimientos no desplazan el foco de la ventana. El observador no podrá en muchos casos hacer un seguimiento visual del foco de atención del usuario, especialmente cuando se consulten documentos más largos que la pantalla. En estos casos es recomendable disponer de una versión impresa de las páginas consultadas y hacer un seguimiento por oído, aunque para ello hace falta un poco de entrenamiento.

Si se usa software para registrar las sesiones es muy fácil que aparezcan problemas técnicos, pues tanto los programas de grabación de tests como *JAWS* trabajan con controladores de pantalla a un nivel muy cercano al hardware y a menudo se dan conflictos. En nuestra experiencia usamos *Morae* (versión 1.3), de *TechSmith*, y tras varios intentos conseguimos grabar las sesiones inhabilitando la opción de configuración de *Morae Recorder* “Disable display acceleration”; aún así no conseguimos grabar los accionamientos.

La revisión de las sesiones es más costosa que en otros entornos, lo que aconseja marcar en tiempo real los diferentes momentos de interacción para poder hacer un análisis pormenorizado a posteriori sin un trabajo excesivo. El lenguaje gestual de las personas ciegas difiere un poco del de los usuarios no ciegos y en caso de grabación, hay que aumentar el campo de visión de la cámara web.

## 6. Fase de despedida

Los usuarios invidentes mantienen durante toda la sesión un alto compromiso con las tareas que realizan. Es por ello que tienen un gran interés por conocer las respuestas correctas o saber cuál es la resolución adecuada de una tarea. La fase de despedida es el momento de resolver todas aquellas dudas que, en aras de la fiabilidad del estudio, no se habían podido comunicar durante la ejecución. De nuevo, habrá que recordar al usuario que la evaluación realizada está enfocada a la tecnología y no a los participantes.

Es importante agradecer su participación, su buena predisposición y ofrecerle vías de comunicación con el equipo para cualquier duda o comentario que pueda surgirle en los días posteriores a la realización de la prueba.

Se deben ofrecer al participante las mismas comodidades a su llegada que en su retorno. Es habitual

acompañarle a un punto conocido por él. El momento de acompañamiento suele ser un rato muy relajado, sin tensiones y un espacio de tiempo de comentarios y percepciones que aportan mucha información cualitativa al estudio.

## 6. Fase de análisis

El análisis de los resultados del test realizado con *JAWS* requiere un esfuerzo adicional respecto de uno habitual. Se debe hacer un análisis por oído, al que no estamos habituados y se debe tener en cuenta la existencia, ya mencionada, de los dos cursores, el del Pc y el de *JAWS*. En algunas situaciones el usuario queda atrapado por algún error y salta de un cursor a otro sin saber donde está, y la interacción se vuelve muy caótica. Finalmente, igual como se ha previsto más tiempo en la realización de tareas respecto al necesario en pruebas para no-ciegos, también se destinará más tiempo en su análisis, pues se seguirá el ritmo del usuario en la resolución de tareas.

## 7. Reflexiones

La visión egocéntrica de la mayoría de diseñadores es ver la discapacidad como una falta de habilidades para realizar tareas de forma productiva. Ahora bien, nuestra experiencia es que principalmente lo que se produce es un gran desajuste entre el diseño de las aplicaciones y el estilo cognitivo de las personas ciegas, en la línea de la visión más moderna de las discapacidades (**Palacios; Romanach, 2006**).

Algunas observaciones nos han hecho patentes las capacidades extraordinarias de este colectivo frente a los usuarios no-ciegos, como por ejemplo:

– La capacidad de memorización superior demostrada en la respuesta a las preguntas de satisfacción y en la asimilación de la estructura del documento.

## “Algunas observaciones nos han hecho patentes las capacidades extraordinarias de este colectivo”

– La comprensión lingüística avanzada que les permitía detectar de forma muy aguda incoherencias e inconsistencias tanto en las preguntas, como en el texto, como en la estructura. Una comprensión que les permitía seguir el texto a pesar de la falta de prosodia de la síntesis automática de voz.

– La tolerancia a la frustración y a la fatiga: a pesar de las diversas dificultades experimentadas en el test, los usuarios demostraron una gran capacidad de adaptación a las posibilidades y fueron muy perseverantes en la realización de las tareas.

## 8. Bibliografía

Aenor UNE 139803:2004. *Aplicaciones informáticas para personas con discapacidad. Requisitos de accesibilidad para contenidos en la web*.

**Brajnik, Giorgio**. “Web accessibility testing: when the method is the culprit”. En: *Lecture notes in computer science*, 2006, n. 4.061, pp. 156-163.

**Chandrashekar, Sambhavi**, [et al]. “Using think aloud protocol with blind users: a case for inclusive usability evaluation methods”. En: *8th International ACM sigaccess conference on computers and accessibility*, 2006, pp. 251-252.

**Clark, Joe**. *Screen-reader usability study*. Consultado en: 24-04-07. <http://joelclark.org/access/research/E-commerce-plus-screenreader/>

**Coyne, Kara Pernice; Nielsen, Jacob**. *How to conduct usability evaluations for accessibility: methodology guidelines for testing websites and intranets with users who use assistive technology*. Fremont, CA: Nielsen Norman Group, 2001.

**Coyne, Kara Pernice; Nielsen, Jacob**. *Beyond ALT text: making the web easy to use for users with disabilities: design guidelines for websites and intranets based on usability studies with people using assistive technology*. Fremont, CA: Nielsen Norman Group, 2001.

**Henry, Shawn Lawton**. *Just ask: integrating accessibility throughout design*. Consultado en: 24-04-07. <http://www.uiaccess.com/accessud>

**Liondau, Ley 51/2003**, de 2 de diciembre, de igualdad de oportunidades, no discriminación y accesibilidad universal de las personas con discapacidad. (BOE, n. 289, de 3 diciembre 2003).

**Lssice, Ley 34/2002**, de 11 de julio, de servicios de la sociedad de la información y de comercio electrónico. (BOE, n. 166, de 12 julio 2002).

**Mankoff, Jennifer; Fait, Holly; Tran, Tu**. “Is your web page accessible? A comparative study of methods for assessing web page accessibility for the blind”. En: *Sigchi conference on human factors in computing systems*, 2005, pp. 41-50.

**Real Decreto 366/2007**, de 16 de marzo, por el que se establecen las condiciones de accesibilidad y no discriminación de las personas con discapacidad en sus relaciones con la Administración General del Estado. (BOE, n. 72, de 24-marzo-2007).

**Palacios, Agustina; Romanach, Javier**. *El modelo de la diversidad. La bioética y los derechos humanos como herramientas para alcanzar la plena dignidad en la diversidad funcional*. Consultado en: 24-04-07. <http://www.asoc-ies.org/docs/modelo%20diversidad.pdf>

**Sloan, David**, et al. “Contextual web accessibility-maximizing the benefit of accessibility guidelines”. En: *International cross-disciplinary workshop on web accessibility (W4A)*, 2006, pp. 121-131. Consultado en: 24-04-07. <http://www.ukoln.ac.uk/web-focus/papers/w4a-2006/>

**Termens, Miquel**, et al. *Estudio de la accesibilidad de la documentación científica en soporte digital*. Barcelona: 2006. Estudio subvencionado por: Programa de Estudios y Análisis, del Ministerio de Educación, Cultura y Deporte, convocatoria 2006, ref: EA2006-0049. Consultado en: 24-04-07. <http://www.centrorecursos.com/mec/ayudas/repositorio/20061215001338EA2006-0049.%20Informe%20final.pdf>

**United States Access Board. Section 508 homepage: electronic and information technology**. Consultado en: 24-04-07. <http://www.access-board.gov/508.htm>

**W3C. Web Content Accessibility Guidelines 1.0. May 1999**. Consultado en: 24-04-07. <http://www.w3.org/TR/WAI-WEBCONTENT/>

**W3C. Web Content Accessibility Guidelines 2.0. W3C Working Draft 27 April 2006**. Consultado en: 24-04-07. <http://www.w3.org/TR/WCAG20/>

**Mireia Ribera, Miquel Tèrmens, Maika García Martín**, Departamento de Biblioteconomía y Documentación de la Universitat de Barcelona, Melcior de Palau, 140, 08014 Barcelona. [ribera@ub.edu](mailto:ribera@ub.edu) [termens@ub.edu](mailto:termens@ub.edu) [maika.garcia@gmail.com](mailto:maika.garcia@gmail.com)

# Administración electrónica y gestión documental. Consideraciones a la luz de la Ley para el Acceso Electrónico de los Ciudadanos a los Servicios Públicos

Por Carlota Bustelo y Elisa García-Morales

**Resumen:** La entrada en vigor de la Ley para el Acceso Electrónico de los Ciudadanos a los Servicios Públicos va a tener un gran impacto sobre las administraciones públicas en España. Esta ley afecta muy directamente a la gestión de documentos en el ámbito de las administraciones, eliminando las posibles incertidumbres sobre la validez de los documentos electrónicos. Se argumenta porque la implantación de un sistema de gestión de documentos electrónicos es necesaria para poder cumplir con esta ley. Como herramienta práctica, se enumeran en forma de preguntas las cuestiones a resolver sobre gestión documental cuando se produce una implantación por procesos o trámites de la e-administración.

**Palabras clave:** E-administración, Legislación, Gestión de documentos electrónicos, España

**Title: Electronic administration, records management, and Spain's Electronic Access to Public Services legislation**

**Abstract:** Implementation of the e-government law guaranteeing Spanish citizens electronic access to public services (Ley para el Acceso Electrónico de los Ciudadanos a los Servicios Públicos) will have great impact on public administration at every level. This law directly affects government records and document management, removing possible uncertainties about the validity of electronic records. Arguments are offered concerning the need for a document and records management system to support compliance with the law. As a practical tool, a list of questions is provided to help resolve records and document management issues when e-government procedures or business processes are implemented.

**Keywords:** E-government, Legislation, Electronic records management, Spain

**Bustelo, Carlota; García-Morales, Elisa.** "Administración electrónica y gestión documental. Consideraciones a la luz de la Ley para el Acceso Electrónico de los Ciudadanos a los Servicios Públicos". En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, pp. 106-111.

DOI: 10.3145/epi.2008.ene.13



**Carlota Bustelo Ruesta y Elisa García-Morales Huidobro** son socias-directoras de Inforárea SL, empresa consultora especializada en la organización de la documentación y la información. Han dirigido numerosos proyectos de gestión documental, organización de contenidos y diseño de sistemas de información en España y Latinoamérica, temas sobre los que han publicado artículos en congresos y revistas profesionales. Una de sus especialidades es la gestión de documentos electrónicos en las organizaciones, terreno en el que actualmente dirigen varios proyectos de evaluación, diseño e implantación. Inforárea es miembro activo del CTN50 (Información y Documentación) de Aenor y de su Subcomité Gestión de Documentos y Aplicaciones. Recientemente Carlota ha sido elegida "convenor" de grupo de trabajo internacional de ISO que se encargará de la redacción de la nueva norma de requerimientos para la gestión de documentos.

## 1. Una nueva ley de gran impacto

EL DÍA 23 DE JUNIO DE 2007 se publicó en el *Boletín Oficial del Estado (BOE)* la *Ley para el acceso electrónico de los ciudadanos a los servicios públicos* (conocida como la ley de la e-administración) que reconoce el derecho de los ciudadanos a exigir a las administraciones públicas –en cualquiera de sus niveles- la realización de los trámites por medios electrónicos.

Las administraciones deberán poner en marcha las medidas necesarias para que ello sea una realidad a partir del 31 de diciembre de 2009. Con este nuevo texto legal se pretende fomentar la realización de tramitaciones administrativas por internet, teléfono, televisión o cualquier otro medio electrónico disponible en el futuro.

En este texto queremos llamar la atención sobre el impacto que esta disposición tiene en la gestión

documental de las administraciones públicas, pues respalda jurídicamente la validez de los documentos electrónicos.

La administración electrónica es una realidad imparable y muchas instituciones y organismos públicos están inmersos desde hace años en proyectos para hacerla posible. En España, gran parte de estas iniciativas se han focalizado en el "front-office" y han tenido poco impacto en la manera de trabajar con los do-

cumentos por parte de las administraciones. La inseguridad respecto a la validez de los documentos electrónicos y –hasta fecha reciente- la escasa utilización de la firma electrónica, son los motivos aducidos para que los documentos que dan soporte a los trámites y procedimientos se continúen gestionando en papel y en paralelo a los procesos automatizados y documentos electrónicos cuando éstos existen. La mayoría de los expedientes y documentos se siguen conservando de forma tradicional.

Algunas iniciativas de digitalización llevadas a cabo han servido para crear copias de originales existentes en papel, con el propósito de facilitar el acceso y uso; pero en ningún caso han suplantado a éstos pues nadie se ha atrevido a eliminar el papel.

Sí hay casos en que las aplicaciones informáticas han hecho “desmaterializarse” determinados documentos que tradicionalmente se venían produciendo y conservando en soporte en papel: véase por ejemplo el padrón municipal o los libros de cuentas. Esto presenta otro tipo de problemática –a la que se ha prestado escasa atención- que es la conservación de esta información, que antes se plasmaba en documentos y ahora constituyen evidencias electrónicas contenidas en bases de datos.

Con la introducción del registro y las notificaciones telemáticas (Ley 24/2001) y el uso la firma electrónica (Ley 59/2003) muchas organizaciones ya se han encontrado con documentos electrónicos originales y auténticos, pero que forman a su vez parte de trámites y expedientes en los que el resto de la documentación está en papel, lo cual prácticamente obliga a su impresión para poderlos gestionar.

## 2. Qué cambia con la Laecsp<sup>1</sup> para la gestión documental

La Ley consagra el derecho de los ciudadanos a relacionarse con

las administraciones por medios electrónicos y obliga a las administraciones a poner los medios para hacer posible entre otras cosas:

*“presentar solicitudes y recursos; realizar el trámite de audiencia cuando proceda; efectuar pagos o acceder a las notificaciones y comunicaciones que les remitan la Administración Pública” (Exposición de motivos I)*

Esto lleva implícito un crecimiento exponencial de masa de documentación que existirá ya en su forma original en soporte electrónico. La gestión documental electrónica se hace imprescindible para hacer frente a este problema.

La Laecsp acaba con la indefinición en cuanto a la validez de los documentos electrónicos, consagrando la existencia tanto del documento como del expediente electrónico (Capítulo IV, art. 29, 32), y el acceso de los ciudadanos al mismo:

*“...los interesados en un procedimiento tienen derecho de acceso al mismo y ver los documentos. Lo mismo debe suceder, como mínimo, en un expediente iniciado electrónicamente o tramitado de esta forma. Dicho expediente debe poder permitir el acceso en línea a los interesados para verificar la situación del expediente, sin mengua de todas las garantías de la privacidad.” (Exposición de motivos III)*

La Ley establece las modalidades de copias de los documentos electrónicos: copias digitalizadas de documentos que originalmente existían en soporte papel y viceversa, copias en papel de documentos cuya existencia original es en soporte electrónico (Capítulo IV, art. 30. Copias electrónicas).

La Ley respalda el archivo electrónico de documentos siempre que esté sujeto a unas medidas de segu-

ridad que garanticen la integridad, autenticidad, confidencialidad, calidad, protección y conservación de los documentos almacenados.

*“Artículo 31. Archivo electrónico de documentos.*

*Podrán almacenarse por medios electrónicos todos los documentos utilizados en las actuaciones administrativas.*

*Los documentos electrónicos que contengan actos administrativos que afecten a derechos o intereses de los particulares podrán conservarse en soportes de esta naturaleza, ya sea en el mismo formato a partir del que se originó el documento o en otro cualquiera que asegure la identidad e integridad de la información necesaria para reproducirlo. Se asegurará en todo caso la posibilidad de trasladar los datos a otros formatos y soportes que garanticen el acceso desde diferentes aplicaciones.*

*Los medios o soportes en que se almacenen documentos, deberán contar con medidas de seguridad que garanticen la integridad, autenticidad, confidencialidad, calidad, protección y conservación de los documentos almacenados. En particular, asegurarán la identificación de los usuarios y el control de accesos, así como el cumplimiento de las garantías previstas en la legislación de protección de datos”.*

## 3. Obligación de las administraciones

La Ley no determina cómo una organización debe o puede abordar la manera de hacer efectiva la relación con los ciudadanos por medios electrónicos; pero pone muchísimo énfasis en la utilización de las nuevas tecnologías en el funcionamiento interno, especialmente en la consecución de la tramitación electrónica:

*En cuanto al funcionamiento interno de la Administración, las nuevas tecnologías permiten oportunidades de mejora (eficiencia y reducción de costes) que hacen ineludible la consideración de las formas de tramitación electrónica, tanto para la tramitación electrónica de expedientes, como para cualquier otra actuación interna de la Administración, expandiéndolas gradualmente con el objetivo del año 2009. (Exposición de motivos III)*

En definitiva, la estrategia de gestión de documentos electrónicos es un elemento esencial para garantizar el cumplimiento de la futura ley:

La tramitación electrónica conlleva la gestión de documentos electrónicos. La propia ley insiste en este aspecto:

*“Las Administraciones Públicas utilizarán las tecnologías de la información de acuerdo con lo dispuesto en la presente Ley, asegurando la disponibilidad, el acceso, la integridad, la autenticidad, la confidencialidad y la conservación de los datos, informaciones y servicios que gestionen en el ejercicio de sus competencias. (Título preliminar, art. 1.2)*

¡Aunque puedan quedar dudas de que las administraciones públicas puedan conseguirlo íntegramente a corto o medio plazo, la aplicación de la ley nos empuja hacia el establecimiento del sistema de gestión de documentos electrónicos como soporte de las tramitaciones ya sean las comunicaciones con el ciudadano vía papel o por medios electrónicos. Esto equivale a establecer que los documentos “oficiales” de la Administración sean los electrónicos. Supone que los recibidos en papel deben digitalizarse adecuadamente para formar parte de los

expedientes y que de la gestión y tramitación desaparecerían los documentos en papel.

Es posible que el sistema de gestión de documentos electrónicos requiera más adelante desarrollos reglamentarios. Mientras esto sucede las opciones tecnológicas existentes nos permiten que los documentos en los que se refleja la actividad de las administraciones públicas:

- a) Se archiven (o guarden) dispersos en múltiples aplicaciones de gestión de datos o de expedientes.
- b) Se archiven en un repositorio único y organizado.
- c) Desaparezcan como “objetos documento” y se conviertan en datos de las aplicaciones de gestión.

En cada Administración estas decisiones estratégicas deben tomarse de acuerdo con las capacidades y medios de cada una de ellas, de forma que se puedan obtener los máximos beneficios, evitando los riesgos.

#### **4. Por qué se hace más necesaria una estrategia de gestión de documentos**

La gestión de documentos siempre ha sido necesaria, tanto en el mundo en papel como para los documentos electrónicos. Pero si bien en la gestión de los documentos en papel existe un conocimiento colectivo de buenas prácticas, que viene del uso del mismo desde hace algunos siglos, cuando hablamos de gestión de documentos electrónicos existen muchos procedimientos y prácticas menos conocidos y un montón de dudas. Es más necesario que nunca un modelo global que se aplique a toda la organización y que no puede ser una copia del modelo en papel. Cualquier modelo debe tener:

– un componente conceptual que implica la definición de cómo se quiere trabajar.

– un componente organizativo que establece las políticas, procedimientos y sistema de responsabilidades.

– un componente tecnológico que incluye las herramientas a utilizar, sus características, los nuevos procesos documentales (como la migración de soportes o la conversión de formatos), las interrelaciones con otras aplicaciones y la interoperabilidad.

Cualquier estrategia de gestión de documentos electrónicos debe basarse en los siguientes principios:

#### **a) Conservar las evidencias y pruebas de las acciones de la organización en el corto, medio y largo plazo.**

En casi todas las administraciones públicas existe una práctica administrativa con siglos de tradición respecto a la formación, organización y conservación de expedientes y documentos en papel. Aunque con bastantes dificultades, que en parte provienen de la propia aparición de las nuevas tecnologías en los hábitos de trabajo, de cierto deterioro de la importancia concedida a los archivos en las administraciones públicas y de la escasez de recursos, los sistemas de archivo tradicionales se han mantenido a lo largo del tiempo. En el entorno electrónico, las instituciones deberán poder conservar sus documentos a largo plazo con las mismas garantías de autenticidad, integridad y accesibilidad que se conservan en papel y esto supone una redefinición importante de funciones y competencias.

#### **b) Facilitar el acceso y el entendimiento de los documentos tanto a los ciudadanos en los trámites que les afectan, como a los organismos o instituciones que los requieran como prueba de la acción administrativa e internamente dentro de las propias dependencias de las instancias gestoras.**

En los documentos en papel esta labor se ejerce mediante el establecimiento de procedimientos y políticas, y el cumplimiento de las mismas por el personal que custodia los documentos y expedientes. En los documentos electrónicos esto no es suficiente. Si los documentos y expedientes se encuentran en una tecnología a la que no se tiene acceso o que requiere de unas habilidades especiales para su comprensión se estarían limitando los derechos legítimos. Es necesario por lo tanto establecer los mecanismos por los que los documentos electrónicos se hacen accesibles y comprensibles por cualquier persona que tenga derecho a acceder a los mismos, sin necesidad de que dichas personas tengan que saber manejar o estar autorizadas a entrar en las aplicaciones en las que se gestionan los trámites administrativos.

### c) **Garantizar el cumplimiento de los requerimientos de autenticidad, integridad y confidencialidad de los documentos.**

Estos requerimientos que en el mundo en papel son evidentes, son más complejos y novedosos cuando hablamos de documentos electrónicos. La estrategia de gestión documental debe establecer y velar por el cumplimiento de los requisitos de fiabilidad e integridad de los documentos. Tiene que incluir nuevos procedimientos y prácticas informáticas que los sistemas de validación y firma electrónica, la auditoría de los sistemas informáticos o la captura de metadatos asociados a los documentos. La complejidad tecnológica que pueden suponer no debe nunca comprometer los otros principios de la gestión de documentos, encontrando el equilibrio necesario entre todos ellos.

### d) **Garantizar la accesibilidad a medio y largo plazo de las evidencias electrónicas.**

La necesidad de acceso a documentos y expedientes no se limita en el tiempo a la duración del trá-

mite. Las administraciones deben conservar sus documentos durante el tiempo que de ellos se deriven derechos y obligaciones, para la consulta de antecedentes, para respaldar la responsabilidad durante un período determinado y para la investigación histórica.

### 5. **Qué hay que plantearse ante la sustitución del papel por documentos electrónicos**

Las administraciones públicas deben ser conscientes de la necesidad de plantearse una estrategia global de gestión de documentos electrónicos para no tener que dar soluciones individualizadas a cada situación particular.

Sin embargo lo habitual es que la implantación de la administración electrónica se haga por trámites o procesos concretos, empezando por los que son más demandados por la sociedad. En muchos casos esta automatización de procesos se hace sin reparar en la necesidad de abordarlo conjuntamente con la gestión de los documentos electrónicos que se van a producir.

### **Problemática de la documentación**

Cuando el objetivo es que en esos procesos los documentos electrónicos sustituyan definitivamente al papel se deben plantear las siguientes cuestiones:

#### 1. **Qué documentos o información constituyen la evidencia del proceso**

– ¿Qué documentos se producen en el proceso?

– ¿Cómo es el expediente en papel que se quiere sustituir?

– Una vez acabado el trámite ¿que información y documentos se necesitan para que constituya el expediente electrónico?

– ¿En que momento los documentos o la información se puede considerar definitiva y auténtica?

– ¿Qué documentos deberíamos convertir en papel (copia certificada) si el ciudadano exige el documento en papel?

– ¿A que datos y documentos debería acceder un ciudadano que tiene derecho a consultar el expediente electrónico?

La determinación de qué es documentación corporativa o evidencias es el paso previo imprescindible para delimitar el ámbito sobre el que actuar. Las evidencias pueden guardarse en forma de objetos-documentos (por ejemplo un pdf) o en forma de conjuntos de datos que representan la evidencia documental, siempre que se puedan cumplir las condiciones de disponibilidad, acceso, integridad, autenticidad, confidencialidad y conservación.

#### 2. **Cómo se puede asegurar la integridad y autenticidad de los documentos**

– ¿La información o los documentos pueden eliminarse o modificarse?

– ¿Cuáles son los elementos de autenticación que se utilizan?

– ¿Cómo puede comprobar la autenticidad una persona con derechos que accede al expediente electrónico?

– ¿Pueden machacarse los datos de una aplicación una vez terminado el trámite?

– ¿Se guardan en formatos que impidan la escritura?

Asegurar estas características para la documentación de cada proceso que se haya determinado como evidencia es imprescindible para poder asegurar una buena gestión. Las formas para asegurarlas pueden ser diversas y pueden estar determinadas por los formatos, los lugares de almacenamiento, las firmas electrónicas o el blindaje de los datos.

#### 3. **Qué información asociada se guarda junto con los documen-**

## tos para hacerlos comprensibles y localizables

– ¿Se guarda información con cada documento del proceso al que corresponde?

– ¿Se guarda información sobre el documento que nos permita identificar la fecha en que se realiza, quien lo hace y las características básicas del mismo?

– ¿Se podrían identificar las personas que aprueban los documentos y los cargos que ejercen en ese momento?

– ¿Los tipos de documentos se han definido específicamente o se han alineado con otros sistemas?

– ¿Se han dado reglas para la cumplimentación de datos que responden a algún estándar (Nombres, DNI, Callejeros, Unidades administrativas, etc.)?

– ¿La información o datos que describen el documento son legibles por sí mismos sin necesidad de uso de la aplicación en la que se ha gestionado el documento o están generados a base de códigos?

– ¿Los documentos mantienen su vinculación, por ejemplo la pertenencia al mismo expediente? ¿Se guarda un índice del expediente?

– ¿Se podrá vincular automáticamente este documento con otros relacionados (por ejemplo dos expedientes de un mismo ciudadano, de la misma calle)?

Para manejar y gestionar un sistema documental es imprescindible contar con una estructura de clasificación, que es la base conceptual para tener la documentación identificada y organizada en categorías que representan las funciones y actividades de la organización a la que pertenecen.

Para hacer los documentos recuperables, legibles y fáciles de buscar –estén donde estén-, será necesario contar con elementos de descripción de los mismos (meta-

datos) que sean lo más homogéneos posible. Cuando se analiza un proceso de forma aislada es difícil saber cuál es el camino a seguir, pero al menos se debería documentar cuáles decisiones se han tomado y por qué.

## 4. Cómo se plantea el acceso a los documentos en el presente y a lo largo de todo el ciclo de vida en que deban conservarse

– ¿Es posible que personas que tienen que tener acceso a los documentos –internas o ciudadanos- puedan acceder a los mismos sin tener que usar las aplicaciones de tramitación?

– ¿Pueden las oficinas gestoras determinar a su criterio quién tiene que tener o no acceso a los documentos electrónicos?

– ¿los documentos pueden ser legibles, e inteligible su contexto fuera de la aplicación en la que se gestionan?

– ¿Cómo se define el nivel de seguridad que tienen que tener los documentos: cuáles están sujetos a *Ley de Protección de Datos*, cuáles son públicos, cuáles son confidenciales, a cuáles acceden los interesados?

– ¿Cómo se van a facilitar los accesos de los ciudadanos a los expedientes cuando lo requieran..., por internet, teniéndose que personar físicamente?

– ¿Quién y dónde puede emitir duplicados de los documentos digitales?

– ¿Quién y cómo se envían documentos electrónicos o copias de los mismos a instancias judiciales u organizaciones fiscalizadoras?

Un requerimiento esencial de la gestión documental es que los documentos sean accesibles cuando se requieran para distintos fines. Las decisiones sobre el nivel de acceso y seguridad que tienen que tener los documentos hay que establecerlas

no sólo pensando en las necesidades del día a día de las oficinas, sino desde la necesidad de acceso para la organización, el ciudadano y la sociedad a medio y largo plazo. Hay que determinar cómo se va a garantizar que la documentación sea accesible (localizable, legible y disponible según los criterios de seguridad) a lo largo del tiempo. El acceso a los documentos tiene distintos componentes que se entremezclan: los formatos de conservación (los formatos abiertos como PDFa o OAIS parece que son ya un acuerdo internacional) y las estrategias de conservación (ver siguiente punto) por un lado; la capacidad para dar acceso electrónico a los que tienen derecho sin necesidad de conocimientos especiales; y la seguridad y protección de los datos confidenciales.

## 5. Cómo se plantea la conservación y almacenamiento de los documentos

– ¿Se ha planteado la eliminación de información cuando ya no se necesite?

– ¿Se plantea el crecimiento infinito de los datos?

– ¿Se plantean estrategias de migración de la información a otros soportes y/o sistemas cuando se necesite menos?

– ¿Se podrá acceder a los documentos y datos cuando ya no exista la aplicación?

– ¿Cómo se garantiza que los documentos a conservar no se pierden, o no han sido manipulados y son auténticos, íntegros y fiables?

– ¿Existen estrategias planificadas de conversión de datos a formatos de conservación?

Pensar en todos estos temas requiere un esfuerzo de proyección a futuro, aunque no sea muy lejano. Este es un punto muy relacionado con el anterior, y por lo tanto es difícil hacerlo de forma parcial. En este caso al menos la Admi-

nistración se debería encaminar a minimizar los riesgos: los formatos abiertos, las tecnologías abiertas y los planteamientos de crecimiento no infinito de las aplicaciones son estrategias internacionalmente aceptadas. Es importante en este punto no confundir las copias de seguridad de las aplicaciones con la conservación de la información y los documentos.

#### 6. Qué rastro queda de las operaciones realizadas sobre los documentos

– ¿Queda constancia del uso de la información o los documentos?

– ¿Queda constancia de los procesos de importación, exportación o traspaso a otro almacenamiento?

Con los documentos electrónicos la trazabilidad de las acciones realizadas sobre los documentos es posible técnicamente. Estos datos permiten hacer frente a posibles auditorías que demuestren las características que pedimos a los documentos.

#### Nota:

1. El proyecto en un principio se llamó *Ley para el acceso electrónico de los ciudadanos a las administraciones públicas (Laecap)*

**Carlota Bustelo; Elisa García-Morales**, Inforárea SL, Núñez de Balboa, 118, 1-D. 28006 Madrid.

Tel.: +34-915 971 383; fax: +34-915 970 582

cbustelo@inforarea.es

correo@inforarea.es

# E-DOCPA

**[edocpa@congresosasturias.com](mailto:edocpa@congresosasturias.com)**  
**<http://edocpa.com/>**

El *Parlamento Europeo* considera que a partir de las bibliotecas digitales y los repositorios institucionales se

puede construir una red para el intercambio de conocimientos y de cultura, próxima a todos los ciudadanos, y una de cuyas tareas básicas es poner al alcance y servicio de la ciudadanía la información necesaria para el desenvolvimiento social, con independencia del lugar de residencia, el nivel de formación o la situación económica.

**E-DOCPA 2008**  
Oviedo - 11, 12 y 13 de junio de 2008

La infoesfera en la innovación de la e-Administración: el conocimiento como proceso modernizador de la gestión documental y de las relaciones con la ciudadanía

The infosphere in e-Government Innovation: the knowledge as a way to modernize records management and relationships with citizens

» Web E-DOCPA 2006

COBRADO DEL PRINCIPADO DE ASTURIAS  
www.asturias.es

Proyecto cofinanciado por la UE-FEDER

Entrar Enter

:: Aviso Legal :: Desarrollado por: eDise ::

## Isis para bibliotecas

Por Ricardo Eíto-Brun

**Resumen:** *Isis –el popular programa informático creado por la Unesco– es un clásico para la gestión de bases de datos textuales. Sin embargo, su carácter gratuito ha hecho que en ciertos contextos se le considere una opción menor frente a otras alternativas comerciales. Si bien su modelo de desarrollo no se corresponde con el de software código abierto (pese a su distribución gratuita), algunos módulos y aplicaciones adicionales sí se han realizado según este modelo, como el editor IsisMarc. Éste ofrece una interfaz fácilmente adaptable a distintos formatos Marc para gestionar un catálogo, incorpora un cliente Z39.50 y permite el control de autoridades. Es posible su integración con otras herramientas Isis para la publicación de bases de datos vía web. Isis también constituye la base de distintos programas de código abierto para gestión de bibliotecas (Catalis y OpenMarcoPolo).*

**Palabras clave:** *CDS-Isis, IsisMarc, Marc, Bases de datos documentales, Control de autoridades.*

**Title:** *Isis for libraries*

**Abstract:** *ISIS –the popular software developed at UNESCO– is a classic tool for database management. Nonetheless, in some cases the fact that it is freeware has made it seem less desirable than commercial alternatives. Although not designed as open software (despite being distributed free of charge), some modules and additional applications, such as the IsisMarc editor, have been developed using open code. IsisMarc provides an interface that is easily adapted to various MARC formats for catalog management, incorporates a Z39.50 client, and permits authority control. It can be integrated with other ISIS tools to publish databases to the web. ISIS also provides the foundation for various open software programs for library management (Catalis and OpenMarcoPolo).*

**Keywords:** *CDS-Isis, IsisMarc, Marc, Document databases, Authority control.*

**Eíto-Brun, Ricardo.** “Isis para bibliotecas”. En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, pp. 112-117.

DOI: 10.3145/epi.2008.ene.14



**Ricardo Eíto-Brun** (Huesca, 1973), es profesor asociado en la Universidad Carlos III de Madrid y desarrolla su actividad profesional en el área de la calidad y gestión documental en el sector de las TIC. Es autor de dos libros sobre XML y de varios artículos sobre lenguajes de marcas, gestión documental y de contenidos.

**EL SOFTWARE CDS-ISIS** (*Computerized Documentation System - Integrated Set for Information Systems*), inicialmente *MicroIsis*, viene siendo distribuido por la Unesco desde la década de los ochenta. Permite la gestión de bases de datos textuales y se ha utilizado con frecuencia como ejemplo de las llamadas bases de datos documentales. Hoy en día es difícil mantener la distinción tradicional entre bases de datos documentales y relacionales, ya que éstas últimas han añadido progresivamente funciones adicionales para indexar y recuperar campos de longitud variable y documentos

en distintos formatos. Pero esto no significa que las tradicionales bases de datos documentales hayan perdido su utilidad.

Son numerosos los profesionales que han sabido explotar las ventajas de *CDS-Isis*, que podríamos resumir con dos términos: flexibilidad y economía (el sistema puede descargarse y usarse gratuitamente, sin ningún tipo de coste para la organización). Frente a esto, otros grupos de usuarios no han mostrado interés por esta solución, en parte debido a su carácter gratuito, que lo situaba en una posición de inferioridad en las preferencias de muchas organizaciones.

### Características de CDS-Isis

Permite almacenar y recuperar registros formados por un número prácticamente ilimitado de campos de longitud fija o variable, para los

---

**“Son numerosos los profesionales que han sabido explotar las ventajas de CDS-Isis, que podríamos resumir con dos términos: flexibilidad y economía”**

---

que se puede establecer su obligatoriedad o carácter opcional, si son repetibles o no, etc. También facilita la búsqueda de combinaciones de términos en el texto completo de los registros, acotar búsquedas a campos específicos, consultar índices con las ocurrencias de las palabras usadas en los registros de la base de datos, etc.

Junto a estas características básicas, incorpora distintas utilidades integradas dentro del programa que hacen posible:

- Definir la estructura y el modelo de registro de la base de datos, los campos y subcampos que la forman.

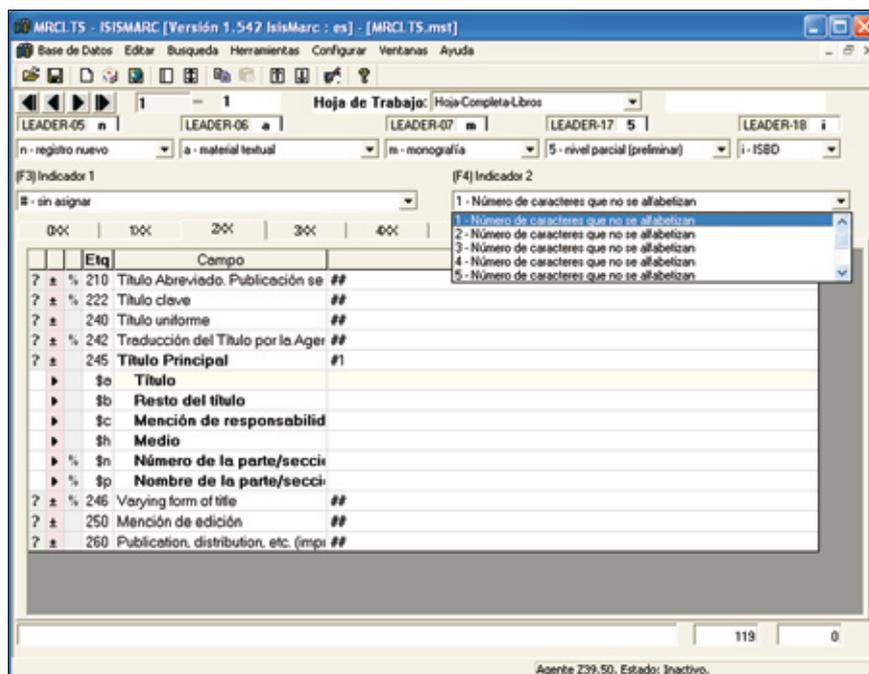
- Crear formularios de entrada y edición de datos. Reciben el nombre de “hojas de entrada de datos”, y en ellas se puede indicar qué campos van a figurar en el formulario, en qué orden, y asociarles mensajes de ayuda y reglas de validación para que su valor se escriba en un formato específico.

- Establecer la presentación impresa y en línea para la lectura de los registros mediante formatos de presentación.

- Establecer los campos que deben indexarse y las reglas que se aplicarán para ello.

Mediante las utilidades anteriores, el programa ofrece numerosas opciones para adaptar la estructura de las bases de datos y sus formatos de presentación a las necesidades de distintos grupos de usuarios.

Además, su funcionalidad inicial se ha ido completando con desarrollos adicionales también disponibles en el portal de la *Unesco*.



Interfaz de usuario de IsisMarc para la edición de registros Marc

Entre ellos podemos citar *GENIsis*, para publicar bases de datos a través de un navegador web, o *WebLIS*, adaptación para bibliotecas. Contamos también con otras aplicaciones de código abierto para bibliotecas basados en *Isis* como *Catalis* y *OpenMarcoPolo*, ambos desarrollados en Argentina.

<http://inmabb.criba.edu.ar/catalis>

<http://marcopolo.uner.edu.ar/>

### ISISMarc

Es uno de los desarrollos complementarios que resulta especialmente interesante para la comunidad bibliotecaria. Este proyecto ha sido liderado por **Ernesto Spinak**, **Alexandra Fernández** y **Emiliano Marmontí**. Ofrece una base de datos adaptada a la compleja estructura de Marc21, formato utilizado

**“IsisMarc ofrece una base de datos adaptada a la compleja estructura de Marc21, formato utilizado ampliamente en bibliotecas para la creación y mantenimiento de registros bibliográficos y de autoridades”**

ampliamente en bibliotecas para la creación y mantenimiento de registros bibliográficos y de autoridades. Para lograr esto, los autores de *IsisMarc* han establecido las correspondencias entre la estructura del registro Marc y los campos y subcampos de una base de datos *CDS-Isis*. El resultado es un sistema con una interfaz de usuario muy próxima a las de los módulos de catalogación de los sistemas integrados de gestión bibliotecaria, en la que se respeta con fidelidad la estructura del formato Marc.

La documentación publicada para *IsisMarc* ofrece numerosos detalles sobre la estructura interna de esta base de datos, lo que puede

Una base de datos *CDS-Isis* consta de una serie de archivos con distintas extensiones: *.mst* guarda los datos; *.fdt* la relación de campos que conforman el modelo de registro; *.fmt* son hojas de entrada de datos; *.pft* formatos de impresión, mientras que los que recogen los índices inversos para realizar búsquedas tienen las extensiones *.ipf*, *.n01*, etc.

llevarnos a pensar que se trata de un programa complejo. Sin embargo, desde el punto de vista de usuario, la interfaz resulta sumamente intuitiva, gracias en gran parte a las numerosas ayudas que se han incorporado. Entre estas destacan los mensajes y las páginas de ayuda para los campos Marc y la posibilidad de seleccionar los valores de los indicadores y de las distintas posiciones de los campos de longitud fija usando listas predefinidas. A modo de ejemplo, para completar las posiciones del campo 008 correspondientes al código del idioma de la publicación, *IsisMarc* nos muestra una lista con los códigos y el nombre de los idiomas según la norma aplicable; el usuario únicamente debe seleccionar el valor correspondiente de esta lista.

La interfaz de usuario consta de una tabla en la que cada campo y sus subcampos se presentan en filas separadas. Junto a cada fila se señala -mediante iconos- si se trata de un campo o subcampo repetible. Los subcampos definidos para cada campo pueden ocultarse y mostrarse para facilitar la edición.

Por defecto, los campos Marc se agrupan en pestañas según su primer dígito (recordemos que los campos que comienzan por el mis-

Field	Icon	Value
Fecha ingresada en el archivo	📅	200712
Tipo de fecha/estado de la publicación	📅	s
Fecha 1	▶	2007
Fecha 2	▶	
Lugar de publicación	📍	sp
Filler	📄	
Forma del ítem	📄	#
Filler	📄	
Idioma	📄	afr
Registro modificado	📄	#
Fuente de la catalogación	▶	d

Asistente para completar el campo de longitud fija 008

mo dígito suelen tener una función similar: por ejemplo, los que comienzan por 6 recogen las materias asignadas; por 5 contienen notas, etc.) Esta disposición de los campos agrupados en pestañas se puede cambiar mediante las opciones de configuración.

El programa también ofrece la posibilidad de mostrar los registros bibliográficos en modo lectura usando distintas presentaciones o formatos, así como utilizar las funciones de búsqueda estándar de *CDS-Isis* (en este caso, las consultas se pueden acotar a los campos de Marc).

## Modos de edición y control de autoridades

Otra característica interesante son los modos de edición. Hace mención a las formas de completar la información de los campos Marc. En *IsisMarc* disponemos de campos cuyo valor:

- se asigna automáticamente tras realizar algún tipo de cálculo o con un valor por defecto; sucede así con el 005 (Fecha de la última modificación del registro) o con las posiciones 00-05 del campo 008, que se completan con la Fecha de creación;
- debe ser completado manualmente por el usuario, y que son la mayoría de los casos;
- se toma de una lista predefinida o bien,
- se debe seleccionar de una lista de autoridades.

Este último es uno de los aspectos más atractivos de *IsisMarc*. En la práctica catalográfica, es aconsejable (por no decir que obligatorio) disponer de un listado de términos y nombres autorizados para referirnos a las personas, entidades, títulos de series, materias o títulos uniformes que podemos usar como puntos de acceso a los registros bibliográficos: constituyen el fichero o lista de autoridades. Este proce-

El formato Marc es una de las principales normas bibliotecarias. Ideado a finales de los años sesenta en la *Library of Congress* para facilitar la transferencia e intercambio de registros bibliográficos legibles por ordenador, sigue siendo la base para la catalogación automatizada y el intercambio de datos bibliográficos.

El registro Marc se estructura en tres partes: cabecera, directorio y campos (de longitud fija y variable). Cada campo se identifica mediante una etiqueta de tres dígitos y se divide en uno o más subcampos representados normalmente por una letra minúscula. Se trata de un registro estructurado que permite codificar con un gran nivel de detalle la descripción de los registros bibliográficos y sus puntos de acceso. También se utiliza para codificar registros de autoridad.

so de normalización garantiza que distintos catalogadores usen una forma común para hacer referencia a las mismas personas, entidades, materias, etc. Los listados de autoridades también recogen diversos nombres usados para referirse a una misma persona, entidad, título, etc., y diferencian entre todas ellas cuál es la forma autorizada (es decir, la que debemos usar en la catalogación y en la búsqueda) y cuáles las no autorizadas. También permiten registrar relaciones del tipo “véase además” entre encabezamientos autorizados, que remiten al usuario a otros que –al estar relacionados con uno de los términos que él ha propuesto en su búsqueda– pueden resultar también relevantes en la recuperación.

El uso de autoridades dota a los registros del catálogo de una consistencia que se traduce, necesariamente, en una recuperación más efectiva, y garantiza que tanto el catalogador como el usuario utilicen los mismos términos para referirse a las mismas entidades en los procesos de catalogación y búsqueda.

**“El uso de autoridades dota a los registros del catálogo de una consistencia que se traduce, necesariamente, en una recuperación más efectiva”**

El formato Marc21 no sólo permite crear registros bibliográficos; también se ha publicado el formato para la codificación de registros de autoridad en los que constarán datos como la forma autorizada y las no autorizadas, referencias cruzadas del tipo “véase además”, distintas notas y las fuentes documentales donde se han identificado los diferentes nombres.

*IsisMarc* ofrece además la capacidad de trabajar con registros de autoridad. Podemos usar una base de datos independiente para mantener registros de autoridad Marc. Por otra parte, es posible vincular campos de una base de datos de registros bibliográficos con la de autoridades; de esta forma, cuando el usuario complete el valor de estos campos, el sistema le obligará a seleccionar uno de los valores ya disponibles en autoridades. *IsisMarc* permite mantener distintas bases de datos abiertas, por lo que la navegación entre ellas no resulta compleja para el usuario, si bien éste es uno de los aspectos mejorables en la interfaz de usuario del programa.

**Posibilidades de adaptación**

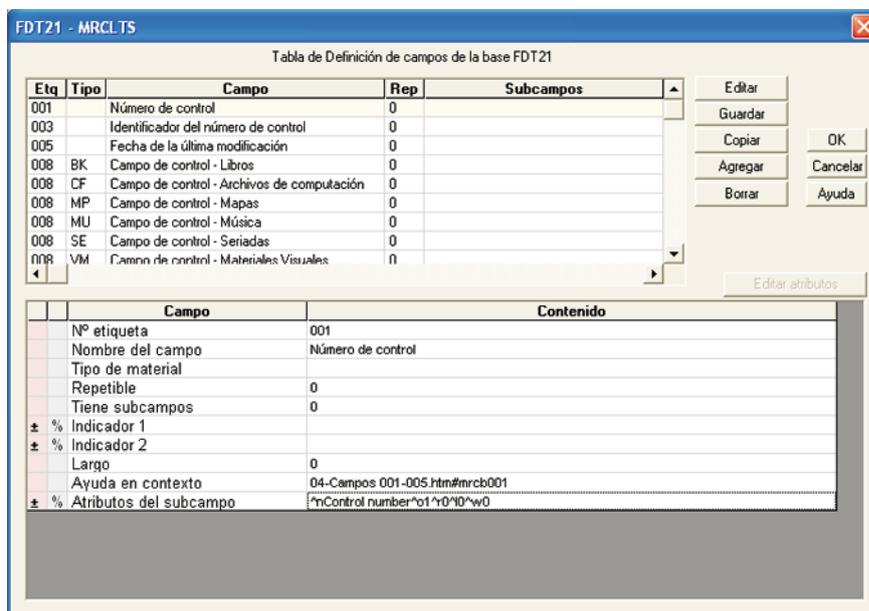
La distribución de *IsisMarc* incluye bases de datos de ejemplo para registros bibliográficos y de autoridad en inglés y español, así como formatos de presentación y de entrada de datos. Pueden tomarse como referencia para crear otras (bastaría con copiar los ficheros correspondientes a otra carpeta, renombrarlos y hacer unas modificaciones menores). Para facilitar esto, *IsisMarc* permite modificar la estructura de la base de datos y

crear nuevas hojas de entrada de datos y formatos de visualización. Los literales (menús y mensajes) que muestra la aplicación pueden traducirse a otros idiomas.

**“La distribución de IsisMarc incluye bases de datos de ejemplo para registros bibliográficos y de autoridad en inglés y español, formatos de presentación y de entrada de datos”**

**Cliente Z39.50**

La norma Z39.50 es uno de los logros más significativos en la aplicación de las tecnologías de la información en bibliotecas. Esta norma establece un modelo de recuperación abstracto y normaliza la forma en la que un sistema informático cliente puede recuperar información de un servidor remoto. Para ello normaliza los mensajes de petición y respuesta que deben intercambiar los dos ordenadores en las distintas etapas del proceso de recuperación (establecer la conexión,



Ventana para editar la estructura de la base de datos

formular y enviar la petición de búsqueda, transmitir información sobre los resultados, descargar los registros recuperados, etc.).

Esta norma -ampliamente difundida y soportada por los sistemas integrados de gestión bibliotecaria y por los programas de gestión de bibliografías tipo *EndNote*- permite a los usuarios consultar simultáneamente catálogos que ejecutan programas informáticos de distintos fabricantes desde un único programa, y obtener una única lista de resultados (de hecho, Z39.50 es uno de los pilares en los que se basan numerosos metabuscadores o *buscadores federados* populares en los últimos años). *IsisMarc* incorpora un cliente Z39.50 basado en el módulo YAZ de la empresa *Index Data*.

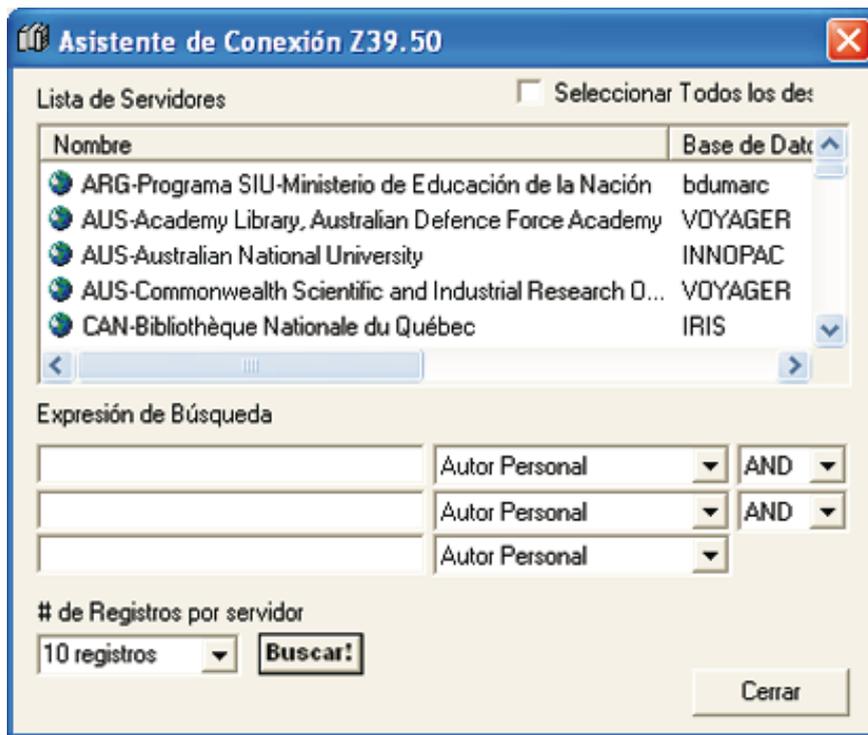
---

**“*IsisMarc* incorpora un cliente Z39.50 basado en el módulo YAZ de la empresa *Index Data*”**

---

Su funcionamiento es el habitual en este tipo de programas: tras seleccionar los catálogos donde queremos buscar (hay una serie de servidores ya configurados) e indicar las condiciones (podemos combinar criterios mediante operadores booleanos y acotar criterios a campos específicos), lanzaremos la búsqueda. Conforme se recuperen registros de los servidores se mostrarán en pantalla en formato de lista, desde donde podremos abrir el registro y volcarlo a nuestra base de datos de registros bibliográficos.

A parte del cliente Z39.50, *IsisMarc* posee funciones para importar registros Marc que hayamos obtenido por cualquier otro medio (por ejemplo, mediante una descarga desde opacs disponibles en la web) y exportar los registros de la



Asistente para búsquedas a través de Z39.50

base de datos en formato Marc para intercambiarlos con otros programas o centros. Soporta el formato Marcxml (adaptación a xml) para facilitar el intercambio en otros entornos que usen este formato.

### Conclusiones

Desde su lanzamiento, la familia de programas *CDS-Isis* se ha completado con aplicaciones adicionales que han ampliado las funciones iniciales de este gestor de bases de datos. El soporte al formato Marc en *IsisMarc* constituye una ventaja añadida a las que vienen disfrutando los usuarios de *CDS-Isis* y sirve como muestra de la evolución continua y del compromiso por la mejora de un software desarrollado al margen de los modelos de distribución comerciales.

Obviamente, sus características no pueden equipararse a las que ofrecen los sistemas comerciales, pero en los últimos años se han ido despejando incógnitas sobre las capacidades reales de las aplicaciones desarrolladas siguiendo modelos alternativos (debemos recordar que

---

**“Negar que el código abierto puede alcanzar amplios niveles de calidad y aceptación sería negar la evidencia”**

---

*CDS-Isis* es un sistema creado con anterioridad a la generalización del modelo basado en el código abierto y que a pesar de su carácter gratuito no encaja con este tipo de aplicaciones, mientras que *IsisMarc* sí). Negar que el código abierto puede alcanzar amplios niveles de calidad y aceptación sería negar la evidencia.

Pero no debemos obviar que los proyectos de código abierto que han obtenido un impacto significativo se han desarrollado en áreas donde existían:

- posibilidades de obtener beneficios o ahorros sustanciales (como sucede con los sistemas operativos, sistemas gestores de bases de datos relacionales, aplicaciones ofimáticas o entornos de desarrollo de software), y

– en las que se ha contado con el apoyo de empresas que, por motivos competitivos, han realizado inversiones importantes en la implementación de aplicaciones alternativas a las distribuidas comercialmente.

En otras áreas donde no se cumplan estas condiciones es difícil prever una evolución exitosa y una adopción destacada de soluciones basadas en código abierto. El caso de *Isis* puede constituir una excepción: disponemos de la experiencia adquirida por una amplia comunidad de usuarios (gracias sobre todo

a los compañeros de Hispanoamérica) que garantiza su futura evolución, a pesar de los rumores sobre su posible discontinuación publicados en algunos foros internet en las últimas fechas.

### Bibliografía y referencias

Blog *CDS-Isis*, gestionado por **Hugo Rueda-Vildoso**.

[http://hrueda-isis.blogspot.com/2007\\_12\\_01\\_archive.html](http://hrueda-isis.blogspot.com/2007_12_01_archive.html)

Lista de distribución de correo de usuarios de *CDS-Isis*

<http://infolac.ucol.mx/microisis/lista.html>

**Martínez, Jerónimo.** *MicroIsis para bibliotecarios*. Madrid: Arco-Libros, 1996.

**Perpinyà, Remei.** "Heurisko, un buen programa sin terminar". En: *El profesional de la información*, julio 1995, pp. 15-16.

[http://www.elprofesionalde lainformacion.com/contenidos/1995/julio/heurisko\\_un\\_buen\\_programa\\_sin\\_terminar.html](http://www.elprofesionalde lainformacion.com/contenidos/1995/julio/heurisko_un_buen_programa_sin_terminar.html)

**Tramullas, Jesús; Garrido-Picazo, Piedad** (coords.). *Software libre para servicios de información digital*. Madrid: Pearson Prentice-Hall, 2006.

*Unesco CDS-Isis*.

[http://portal.unesco.org/ci/en/ev.php-URL\\_ID=2071&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/ci/en/ev.php-URL_ID=2071&URL_DO=DO_TOPIC&URL_SECTION=201.html)

**Ricardo Eúto-Brun**, *Universidad Carlos III de Madrid*.  
[reito@bib.uc3m.es](mailto:reito@bib.uc3m.es)



ANUARIO

# Think

# EPI

## Artículos en la sección SISTEMAS DE INFORMACIÓN

### Nuevas formas de vida en la Web: mashups bibliotecarios

*Jorge Serrano-Cobos*

### Usabilidad de lápiz y papel para servicios de información

*Jorge Serrano-Cobos*

### Reflexión abierta sobre las reuniones virtuales en internet

*Jordi Grau*

### Web móvil

*Francisco Tosete*

### Software libre, clave para los sistemas de información en Hispanoamérica

*Ricardo Eúto-Brun*

Más información en <http://www.thinkepi.net/repositorio>

Ya puedes pasarnos tu pedido: [epi@sarenet.es](mailto:epi@sarenet.es)

Anuario ThinkEPI 2008

89,42 € + IVA = 93 €

Anuario ThinkEPI 2007 + 2008

115,39 € + IVA = 120 €

Anuario ThinkEPI

<http://www.thinkepi.net>



baratz

gestionando el conocimiento

**+ Absys**

Sistema Integrado de  
Gestión de Bibliotecas

**+ Albalá**

Sistema Integrado de  
Gestión de Archivos

**+ BKM**

Solución Integrada para  
la Gestión Documental



baratz

gestionando el conocimiento

+ Tel 91 456 0360  
+ [informa@baratz.es](mailto:informa@baratz.es)  
+ [www.baratz.es](http://www.baratz.es)

## E-LIS: una comunidad internacional de profesionales

Por **Fernanda Peset**

**Peset, Fernanda.** "E-LIS: una comunidad internacional de profesionales" (reseña). En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, p. 119.

DOI: 10.3145/epi.2008.ene.15

**Y CELEBRAMOS EL E-LIS Meeting: "Enhancing the access to LIS literature", Valencia, 13-14 de noviembre de 2007.**

Dado que *E-LIS* es un proyecto sobradamente conocido, ahora con más de 7.000 documentos a texto completo, con cuatro millones de visitas anuales y picos mensuales de 21.000, no insistiré en su importancia para nuestro sector...

Tras este encuentro creo que el término *espíritu de comunidad* define las verdaderas columnas que sustentan el repositorio: el equipo de trabajo (administradores, informáticos, editores...) y los autores que introducen sus datos.



Destacaré cuatro cuestiones que lo avalan:

1. Los estudios estadísticos presentados por **Tim Brody**, que permiten conocer los patrones de trabajo de cada archivo, indican qué repositorios se mantienen por la comunidad que los apoya: en sus palabras, "*size isn't everything*". El patrón temporal de inserción de documentos en *E-LIS* –pocos artí-

culos por parte de muchos autores, de una forma continuada– demuestra el apoyo que recibe por parte de una amplia comunidad profesional.

2. Esto mismo, los científicos en la red, fue destacado en las intervenciones de **Tomás Baiget** (*The Networking Science*) y **Thomas Krichel** (*RePEC*, el repositorio de documentos de Economía que sirvió de modelo a *E-LIS*).

3. Por último, el apoyo de las instituciones y su presencia en el evento: la propia *Universidad Politécnica*, la *Subdirección General de Coordinación Bibliotecaria del Ministerio de Cultura*, la *Dirección General del Libro y Archivos de la Generalitat Valenciana*, el *Servicio de Bibliotecas de la Universitat de València*, el grupo de investigación *Tecnodoc* de la *Universidad Carlos III de Madrid*, *Florida Universidad*, *Ciepi* o *El Profesional de la Información*.

Y esperamos seguir aumentando y que las nuevas generaciones que colaboraron en la organización se involucren activamente en el proyecto. Gracias a todos.

[http://personales.upv.es/mpesetm/elis/info\\_gen\\_es.html](http://personales.upv.es/mpesetm/elis/info_gen_es.html)

<http://eprints.rclis.org>

**Fernanda Peset**, *Universidad Politécnica de Valencia*  
mpesetm@upv.es



Un coffe-break: Remedios Melero, Paula Sequeiros, José Manuel Barrueco, Antonia Ferrer, Fernanda Peset, Jessie Hey, Dennis Nicholson, Tim Brody y Lucía Sanjaime

# Scientific publishing in the European research area

Por Fernanda Peset

**Peset, Fernanda.** "Scientific publishing in the European research area" (reseña). En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, p. 120.

DOI: 10.3145/epi.2008.ene.16

RECIENTEMENTE LA **COMISIÓN EUROPEA** ha distribuido la edición impresa de las actas de la conferencia sobre edición científica celebrada en Bruselas, 15-16 de febrero de 2007<sup>1</sup>.

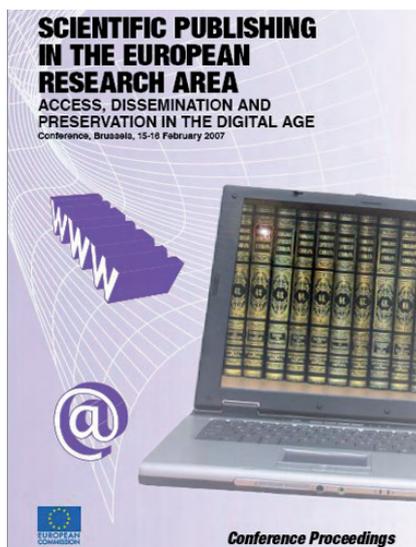
Puede bajar gratuitamente de:

[http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/conference-proceeding-022007\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/conference-proceeding-022007_en.pdf)

Cabe destacar dos aspectos de la conferencia. Lo primero que me resultó sorprendente fue la orientación de la mayoría de intervenciones –que pueden verse desarrolladas en el texto completo del libro de actas– hacia el Acceso Abierto, cuando en realidad el título de la conferencia se refería a la publicación científica en general.

Recordemos que depositar un trabajo en un archivo abierto no es sinónimo de publicarlo en el sentido estricto del término. Por tanto, aunque las revistas científicas siguen detentando su papel de certificadoras del conocimiento a través del sistema de evaluación por pares, parece que el movimiento de Acceso Abierto les preocupa.

De hecho, el esloveno **Janez Potočnik**, comisario para ciencia e investigación en la UE, mostró el interés de la *Comisión* en conciliar los intereses entre los editores científicos y el modelo de distribución abierto. El borrador de la directiva posterior de la *Comisión Europea* se fraguó en esa fecha y recogía este nuevo paradigma de difusión de la información científica<sup>2</sup>. También en este sentido, **Norbert Kroó**, de la



*Academia Húngara de las Ciencias*, recordó la Declaración del *Consejo Europeo de Investigación* sobre el Acceso Abierto<sup>3</sup>. Los organismos financiadores no pueden soslayar los beneficios de las políticas que obligan a depositar los resultados de proyectos financiados con fondos públicos.

Resulta reconfortante que las más altas autoridades europeas y las grandes editoriales transnacionales acusen el impacto del movimiento de Acceso Abierto. El éxito de los múltiples proyectos que poco a poco conforman una malla tupida de contenidos científicos es el éxito de todas las personas –técnicos, autores, autoridades...– interesadas e involucradas en ellos.

El segundo aspecto que cabe destacar fue la cantidad de personas interesadas en el tema y su distinta procedencia. La representación española fue destacable, con **Manuela Palafox** (UCM), **Alice Keefer** (UB), **Alicia López Me-**

**dina** (UNED), **Remedios Mele-ro** (IATA), **Ernest Abadal** (UB), **Francisca Abad** (UV), **Antonia Ferrer** y yo misma (UPV).

## Notas

1. *Scientific publishing in the European research area: access, dissemination and preservation in the digital age: Conference, Brussels, 15-16 February 2007*. Luxembourg: Office for Official Publications of the European Communities, 2007. ISBN 978-92-79-05476-1.

[http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/conference-proceeding-022007\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/conference-proceeding-022007_en.pdf)

2. Comisión de las Comunidades Europeas. *Comunicación de la Comisión al Parlamento Europeo, al Consejo y al Comité Económico y Social Europeo sobre la información científica en la era digital: acceso, difusión y preservación: COM(2007) 56 final [SEC(2007)181]*. Bruselas: CE, 14.2.2007

[http://eur-lex.europa.eu/LexUriServ/site/es/com/2007/com2007\\_0056es01.pdf](http://eur-lex.europa.eu/LexUriServ/site/es/com/2007/com2007_0056es01.pdf)

3. European Research Council. *Scientific Council statement on open access*. Bruselas: ERC, December 2006.

<http://erc.europa.eu/pdf/open-access.pdf>

**Fernanda Peset**, Universidad Politécnica de Valencia  
[mpesetm@upv.es](mailto:mpesetm@upv.es)



En la foto de la p. 19 de las actas oficiales distinguimos estos españoles: **M. Palafox**, **R. Mele-ro**, **E. Abadal** y **M. F. Abad** (de espaldas)

## BOLETÍN DE SUSCRIPCIÓN

Deseo recibir todos los números de la revista EPI a partir del mes de enero del año

Suscripción:  Institucional  Personal

Nombre:  Institución:

(Los suscriptores individuales no han de escribir ningún nombre de institución, sólo indicar la dirección particular)

Departamento:  NIF institucional:

Dirección:

Código postal:  Ciudad:  País:

Teléfono:  Fax:  Correo-e:

Método de pago:

Tarjeta de crédito:  VISA  Master Card  American Express

Titular de la tarjeta:

Número de tarjeta:

Caducidad (mm/aaaa):

Cheque nominativo en euros a nombre de El profesional de la información

Transferencia bancaria a la cuenta de La Caixa 2100 0818 93 0200745544

Enviar, fotocopiado o escaneado, el resguardo de la transferencia.

Las transferencias desde fuera de España deben hacerse a:

IBAN ES95 2100 0818 9302 0074 5544

BIC/Código Swift CAIXESBXXX

Giro postal al apartado de correos 32.280 de Barcelona

Enviar, fotocopiado o escaneado, el resguardo del giro.

Domiciliación en cuenta bancaria

Entidad:  Oficina:  DC:  Núm:

Titular de la cuenta:

Los precios para el año 2008 son los siguientes:

Suscripción anual  
Institucional:  
147,2 € + 4% IVA  
= 153 €

Suscripción sólo online:  
85 € + 4% IVA  
= 88,4 €

Individual:  
75 € + 4% IVA  
= 78,00 €

Número suelto:  
25 € + 4% IVA  
= 26 €

(gastos de envío fuera de España: 7,00 €)

Coste adicional de  
correo aéreo:  
España: 00,00 €

Europa (menos  
España): 30,00 €

Américas y resto del  
mundo: 45,00 €

La suscripción a la revista se realiza por años naturales completos, es decir, desde el mes de enero del año que usted desee que comience su nueva suscripción

Enviar el boletín  
relleno, por correo  
postal o electrónico, a:

Apartado 32.280

08080 Barcelona

España

suscripciones@  
elprofesionaldelainfor  
macion.com

Teléfono de atención al  
suscriptor:

+34 609 352 954

Boletín para comenzar o renovar online la suscripción: <http://www.elprofesionaldelainformacion.com/suscripciones.html>

## Información para los autores

Todos los profesionales que lo deseen pueden remitir a la redacción de la revista **El profesional de la información** sus colaboraciones en forma de:

- Notas breves

- Trabajos más amplios sobre temas de fondo para la sección "Artículos".

El texto ha de enviarse en formato electrónico. Aparte, los materiales gráficos, en papel o ficheros gif, jpeg o tiff con unos anchos de entre 12 y 5,7 cm. y una resolución de 300 ppp.

El tamaño ideal de un estudio para la sección "Artículos" es de 4.000 palabras. En casos excepcionales pueden publicarse artículos de mayor extensión. Los trabajos de esta sección son aprobados según el sistema tradicional "peer review": al menos dos expertos en el tema, del Consejo Asesor de la revista y/o externos, deben dar el visto bueno antes de su publicación.

Los textos deben enviarse sin formatos especiales (títulos, secciones, subsecciones, pies de página, sangrías, tabulaciones, colores, etc.).

Los trabajos para la sección "Artículos" deben incluir: a) título en castellano, b) resumen en castellano de 100-150 palabras, c) 5-10 palabras clave en castellano, d) título en inglés, e) resumen en inglés de 100-150 palabras, f) 5-10 palabras clave en inglés, g) texto completo en castellano y h) nombre de los autores, lugar de trabajo y dirección de correo electrónico. **Han de ser inéditos.**

Se valorará especialmente que los trabajos sean concisos y precisos. Se ruega a los autores que eviten una excesiva retórica.

Las citas bibliográficas en el texto se realizarán de la forma: (Apellido, año). Las referencias bibliográficas, que se limitarán a las obras citadas en el texto, han de prepararse de acuerdo con el siguiente esquema:

Artículos de una publicación periódica:

Apellido, Nombre; Apellido2, Nombre2. "Título del artículo". En: Título de la publicación periódica, año, mes, v. [volumen], n. [número del ejemplar], pp. [págs. comienzo-final].

Ponencia presentada en un congreso:

Apellido, Nombre; Apellido2, Nombre2. "Título de ponencia". En: nombre del congreso, año, pp. [págs. comienzo-final].

Capítulo de una monografía:

Apellido, Nombre; Apellido2, Nombre2. Título del capítulo. En: Apellido, Nombre; Apellido2, Nombre2. Título de la monografía. Lugar de publicación: editor, fecha. ISBN [número].

Monografías:

Apellido, Nombre; Apellido2, Nombre2. Título del trabajo. Lugar de publicación: Editor, fecha. ISBN [número]

Recurso en línea:

Apellido, Nombre; Apellido2, Nombre2. Título del recurso. Consultado en: día-mes-año.

Dirección:

Las contribuciones se pueden enviar a la redacción de la revista o a cualquiera de los miembros del consejo de redacción.

El hecho de que un trabajo sea publicado en EPI no implica que la redacción se adhiera a las opiniones expresadas en él.

Redacción EPI:

Apartado 32.280

08080 Barcelona.

[epi@elprofesionaldelainformacion.com](mailto:epi@elprofesionaldelainformacion.com)

La redacción se reserva el derecho de adaptar los textos al estilo gramatical y literario de la revista.