



Observatorio

Madurez de la gestión de contenidos, ¿sinónimo de desgaste o de oportunidades?

Ricardo Eito-Brun

Artículos

Marcado semántico automático en gestores de contenidos: integración y cuantificación

J. A. Pastor-Sánchez, E. Orduña-Malea y T. Saorín

La gestión de contenidos como actividad estratégica en empresas de radiodifusión

M. Fernández-Sande, D. Rodríguez-Barba y M. Rodríguez-Pallares

Gestión de los fondos documentales en RNE

Carmen Marta-Lazo y Miguel-Ángel Ortiz-Sobrino

Estrategias de difusión de una revista científica. Un experimento con *El profesional de la información*

Sergi Cortiñas-Rovira y Xavier Ramon-Vegas

Gestión de datos de investigación: infraestructuras para su difusión

L. M. González, T. Saorín, A. Ferrer-Sapena, R. Aleixandre-Benavent y F. Peset

Análisis

Gestión de contenidos con *Drupal*: revisión de módulos específicos para bibliotecas, archivos y museos

Jesús Tramullas

Evolución de los repositorios documentales. El caso *SocialNet*

P. Lara, E. Serradell y D. Maniega

Percepción museográfica de la biblioteca histórica o patrimonial

Manuel-José Pedraza-Gracia

Nuevas tecnologías en análisis de inteligencia competitiva. Casos prácticos

P. Ramírez-Calvo, A. C. Triviño, A. Berges-García, J. M. Meneses-Chaus y J. Fernán-Martínez

UNE-ISO 16175 sobre gestión de documentos en oficinas electrónicas

Vicent Giménez-Chornet

Indicadores

The need for end-user customization of the journal-sets of the subject categories in the *SCImago Journal Rank*

Péter Jacsó

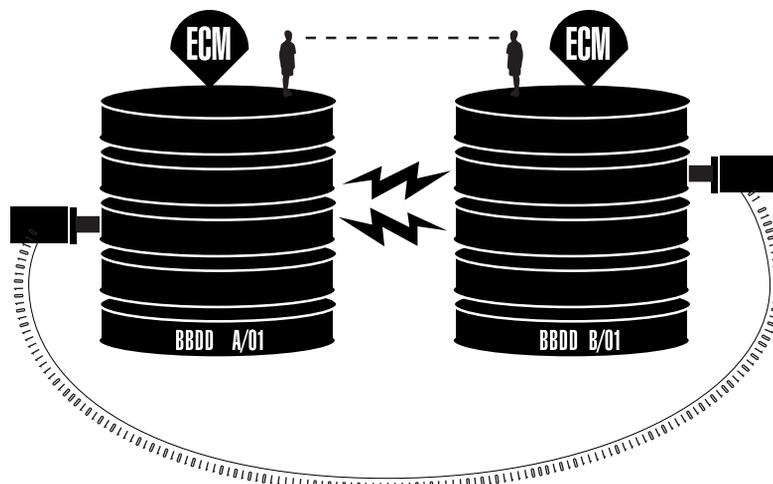
Excellence with leadership: the crown indicator of *SCImago Institutions Rankings Iber report*

V. Jeremić, M. Jovanović, Z. Radojičić, and M. Martić



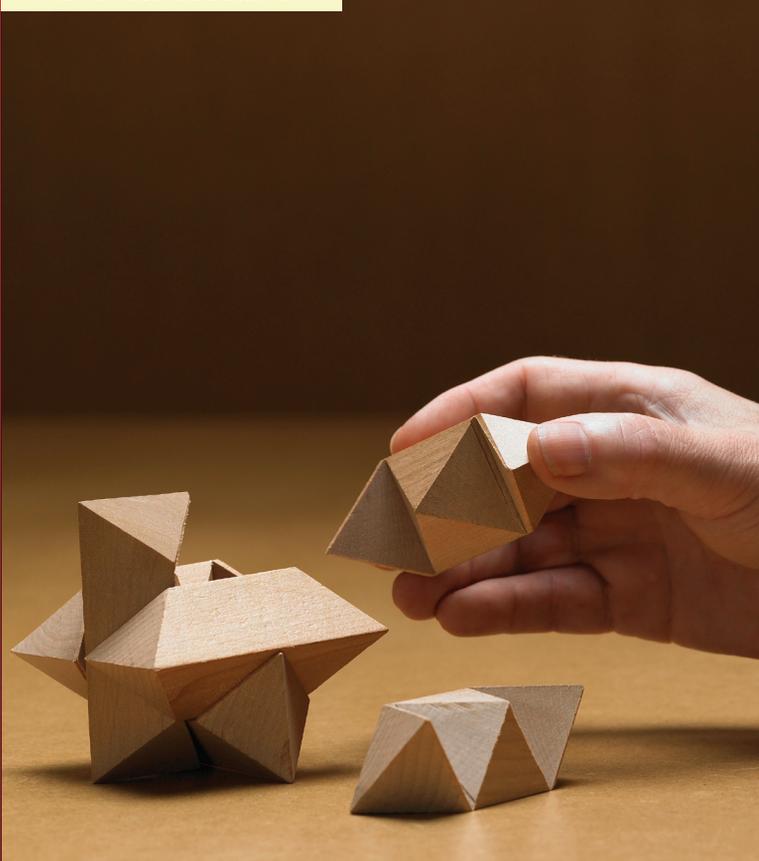
EL PROFESIONAL DE LA INFORMACIÓN

Revista internacional científica y profesional sobre documentación, comunicación, bibliotecas, sistemas y tecnologías de la información.



Gestión de CONTENIDOS

11:24 AM - ECM ORDERING DATA



El profesional de la

información

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

El profesional de la información es una revista de
la editorial EPI SCP NIF: J63664544

Apartado 32.280 - 08080 Barcelona

Tel.: +34 - 609 352 954

<http://www.elprofesionaldelainformacion.com>

Redacción

El profesional de la información

Apartado 32.280

08080 Barcelona

Tel.: +34 - 934 250 029

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>

Servicios online

María T. Moreno

mt.moreno@ono.com

Diseño

MASmedios, <http://www.masmedios.com>

Director artístico: Moisés Mañas

Maquetación

Isabel Olea

isabel.iolea@gmail.com

Producción e Impresión

Sanvergrafic

Pol. Ind. Mascaró

C/ Ponent, nave 6

08756 La Palma de Cervelló (Barcelona)

Tel.: +34 - 936 720 099

Distribución online

MetaPress, Birmingham, Alabama, EUA

<http://elprofesionaldelainformacion.metapress.com>

Depósito legal: B. 12.303-1997

Los trabajos publicados en *EPI* son aprobados según el sistema tradicional "peer review" en doble ciego: son revisados al menos por dos expertos en el tema, del Consejo Asesor de la revista y/o externos.

Para conseguir que los trabajos no pierdan 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

EPI SCP

<http://www.baiget.com>

SUBDIRECTOR

Javier Guallar

Univ. de Barcelona / Univ. Ramon Llull / Univ. Oberta de Catalunya

<http://sites.google.com/site/sitiodejavierguallar>

COORDINADOR EDITORIAL

Carlos Tejada-Artigas

Universidad Complutense de Madrid

tejada@ccdoc.ucm.es

REDACTORA JEFE

Isabel Olea

Universidad de León

isabel.iolea@gmail.com

REDACCIÓN

Natalia Arroyo-Vázquez

Fundación Germán Sánchez Ruipérez

narroyo@fundaciongsr.es

Ricardo Eito-Brun

Grupo GMV

reito@gmv.es

Javier Leiva-Aguilera

Catorze.com

<http://www.javierleiva.info>

Toon Lowette

Grid Electronic Publishing

toon@grid.be

Roser Lozano

CRAI Universitat Rovira i Virgili

roser.lozano@urv.cat

Jorge Serrano-Cobos

MASmedios

jorgeserrano@gmail.com

REVISIÓN DE LENGUA INGLESA

Elaine M. Lilly

Writer's First Aid

elaine@writersfirstaid.com

CONSEJO ASESOR

Ernest Abadal

Universitat de Barcelona, Barcelona.

Isidro F. Aguillo

Centro de CC Humanas y Sociales, CSIC, Madrid.

Adela d'Alòs-Moner

Consultora, Barcelona.

Carlos B. Amat

Inst. Agroquím. y Tecn. Alimentos, CSIC, Valencia.

Jesús Bustamante

Biblioteca, Cedefop, Salónica, Grecia.

Carlota Bustelo-Ruesta

Consultora, Madrid.

Atilio Bustos

Pontificia Univ. Católica de Valparaíso, Chile.

Lluís Codina

Universitat Pompeu Fabra, Barcelona.

Emilio Delgado-López-Cózar

Universidad de Granada, Granada.

Javier Díaz-Noci

Universitat Pompeu Fabra, Barcelona.

Assumpció Estivill

Universitat de Barcelona, Barcelona.

Antonia Ferrer-Sapena

Universidad Politécnica de Valencia, Valencia.

António Fidalgo

Universidade da Beira Interior, Portugal.

Francisco-Javier García-Marco

Universidad de Zaragoza, Zaragoza.

Johannes Keizer

Food and Agriculture Org. (FAO), Roma, Italia.

Jesús Lau

Universidad Veracruzana, Veracruz, México.

Mari-Carmen Marcos

Universitat Pompeu Fabra, Barcelona.

Pere Masip

Blanquerna, Univ. Ramon Llull, Barcelona.

Charles McCathieNeville

Yandex, Moscow, Rusia.

José-Antonio Moreira-González

Universidad Carlos III de Madrid.

Juan-Antonio Pastor

Universidad de Murcia, Murcia.

Fernanda Peset

Universidad Politécnica de Valencia, Valencia.

Joan Roca

Minnesota State University, Mankato, USA.

Luis Rodríguez-Yunta

CCHS, CSIC, Madrid.

Ramón Salaverria

Universidad de Navarra, Pamplona.

Tomás Saorín

Universidad de Murcia, Murcia.

Robert Seal

Loyola Univ. Chicago, Evanston, Illinois, USA.

Ernesto Spinak

Consultor, Montevideo, Uruguay.

Emir-José Suaidén

IBICT, Brasilia, Brasil.

Daniel Torres-Salinas

Universidad de Navarra, Pamplona.

Jesús Tramullas

Universidad de Zaragoza, Zaragoza.

BASES DE DATOS

Academic search premier (Ebsco)

<http://www.ebscohost.com/academic/academic-search-premier>

Dialnet (Universidad de La Rioja)

http://dialnet.unirioja.es/servlet/revista?&clave_revista=469

Economía y negocios (Ebsco)

<http://www.ebscohost.com/academic/economia-y-negocios>

Francis (Inist)

<http://www.inist.fr/spip.php?article170>

ISI Social science citation index, Social SCI, WoS (Thomson Reuters)

http://wokinfo.com/products_tools/multidisciplinary/webofscience/ssci/

Impact Factor 2012 = 0,439

Inspec, Information services in physics, electronics and computing (IET, The Institution of Engineering and Technology)

<http://www.theiet.org/publishing/inspec/>

ISOC, Índice español de ciencias sociales y humanidades (Iedcyt)

<http://bddoc.csic.es:8080/ver/ISOC/revi/0721.html>

ISTA, Information science and technology abstracts (Ebsco)

<http://www.ebscohost.com/public/information-science-technology-abstracts>

Lisa, Library and information science abstracts (CSA)

<http://www.csa.com/factsheets/lisa-set-c.php>

Lista, Library, information science & technology abstracts (Ebsco)

<http://www.libraryresearch.com>

Pascal (Inist)

<http://www.inist.fr/spip.php?article22>

Scopus (Elsevier) Scimago Journal Rank 2012 = 0,279

<http://www.scopus.com>

CATÁLOGOS Y SERVICIOS BIBLIOTECARIOS

Catálogo de Latindex

<http://www.latindex.unam.mx/buscador/ficRev.html?folio=6772>

In-Recs, Revistas españolas de ciencias sociales Grupo EC3, Evaluación de la Ciencia y de la Comunicación Científica, Universidad de Granada

<http://ececubo.ugr.es/ec3/Documentacion.html>

Registros bibliográficos para bibliotecas públicas españolas (Rebeca)

<http://www.mcu.es/bibliotecas/MC/Rebeca/>

ACCESO A LOS TEXTOS COMPLETOS

MetaPress (2000-)

<http://elprofesionaldelainformacion.metapress.com/>

ALPSP Learned journals collection (ALJC) (2011-)

<http://aljc.swets.com>

Ebscohost Electronic Journals Service (2000-embargo 1 año)

<http://ejournals.ebsco.com/direct.asp?JournalID=105302>

Library, information science & technology abstracts with full text (2000-embargo 1 año)

<http://www.ebscohost.com/thisTopic.php?marketID=1&topicID=584>

Academic search complete (2000-embargo 1 año)

<http://www.ebscohost.com/thisTopic.php?marketID=1&topicID=633>

Business source complete (2000-embargo 1 año)

<http://www.ebscohost.com/academic/business-source-complete>

El profesional de la información (1992-embargo 2 años)

<http://elprofesionaldelainformacion.com/contenidos.html>

SwetsWise (2000-)

<https://www.swetswise.com>

EPI EN REDES SOCIALES

<http://www.facebook.com/elprofesionaldelainformacion>

http://twitter.com/revista_EPI

<http://www.linkedin.com/company/el-profesional-de-la-informacion-epi->

<http://pinterest.com/source/elprofesionaldelainformacion.com>

PLATAFORMA DE PRODUCCIÓN OJS

Recyt, Repositorio español de ciencia y tecnología (Fecyt)

<http://recyt.fecyt.es/index.php/EPI>

Tema central: Gestión de contenidos

OBSERVATORIO

- 377 **Madurez de la gestión de contenidos, ¿sinónimo de desgaste o de oportunidades?**
Ricardo Eito-Brun

ARTÍCULOS

- 381 **Marcado semántico automático en gestores de contenidos: integración y cuantificación**
Juan-Antonio Pastor-Sánchez, Enrique Orduña-Malea y Tomás Saorín
- 392 **La gestión de contenidos como actividad estratégica en empresas de radiodifusión. Estudio de casos en la radio comercial española**
Manuel Fernández-Sande, Dolores Rodríguez-Barba y Miriam Rodríguez-Pallares
- 399 **Gestión de los fondos documentales en *Radio Nacional de España***
Carmen Marta-Lazo y Miguel-Ángel Ortiz-Sobrino
- 405 **Estrategias de difusión de una revista científica. Un experimento con *El profesional de la información***
Sergi Cortiñas-Rovira y Xavier Ramon-Vegas
- 415 **Gestión de datos de investigación: infraestructuras para su difusión**
Luis-Millán González, Tomás Saorín, Antonia Ferrer-Sapena, Rafael Aleixandre-Benavent y Fernanda Peset

ANÁLISIS

- 425 **Gestión de contenidos con *Drupal*: revisión de módulos específicos para bibliotecas, archivos y museos**
Jesús Tramullas
- 432 **Evolución de los repositorios documentales. El caso *Socialnet***
Pablo Lara-Navarra, Enric Serradell-López y David Maniega-Legarda
- 440 **Percepción museográfica de la biblioteca histórica o patrimonial: perspectivas y reflexiones en torno a los fondos y libros antiguos**
Manuel-José Pedraza-Gracia
- 448 **Nuevas tecnologías en análisis de inteligencia competitiva. Casos prácticos**
Pilar Ramírez-Calvo, Ana-Cristina Triviño, Aurelio Berges-García, Juan-Manuel Meneses-Chaus y José-Fernán Martínez
- 455 ***UNE-ISO 16175* sobre gestión de documentos en oficinas electrónicas: alcance y limitaciones**
Vicente Giménez-Chornet

INDICADORES

- 459 **The need for end-user customization of the journal-sets of the subject categories in the *Scimago Journal Ranking* database for more appropriate league lists. A case study for the library & information science field**
Péter Jacsó
- 474 **Excellence with leadership: the crown indicator of *Scimago Institutions Rankings Iber* report**
Veljko Jeremić, Marina Jovanović-Milenković, Zoran Radojičić, and Milan Martić

481 AGENDA

484 INFORMACIÓN PARA LOS AUTORES



Creación de Archivos, Museos y Bibliotecas Virtuales

Desde la digitalización de materiales bibliográficos hasta la asignación de metadatos y su implementación en la red, conforme a la normativa internacional.

Productos para crear Bibliotecas Digitales y Virtuales

DIGIBIB 8.0

Solución avanzada para la creación de Bibliotecas Digitales y la Gestión Bibliotecaria Multilingüe.

DIGIARCH 2.0

Sistema digital de descripción y gestión archivística. Descripción en ISAD(G) y EAD 2.0.

DIGIMUS 2.0

Sistema digital de gestión museológica. Actividad subvencionada por el Ministerio de Educación, Cultura y Deporte.



OAsIs-PMH 2.0

Sistema integrado de recolección de diversos esquemas de metadatos:

- DCMI sin cualificar
- MARC 21
- EAD
- SWAP
- mod_OAI
- Linked Open Data

ADAPTACIÓN A EUROPEANA

Implementación del esquema ESE 3.4.1 (Europeana Semantic Elements) y EDM 5.2.3 (Europeana Data Model) Adaptado a la Agenda Digital Europea 2020.

DIGITALIZACIÓN AVANZADA

Con asignación dinámica de metadatos.

- Recolección en la Web para Entidades e Instituciones de Memoria en OAI-PMH y Dublin Core cualificado con ESE 3.4.1
- Consultoría y mappings a EDM 5.2.3 (Europeana Data Model)
- Bibliotecas digitales que permiten la creación, recuperación y recolección de metadatos (MARCXML, DCMI y RDF y RDFs)
- Archivos Web que facilitan la creación, recuperación y recolección de metadatos (EAD 2.0 y EAC 2010)
- Implementación de la Europeaana OpenSearch API
- Adaptación del repositorio OAI para la transmisión de instancias RDF según ORE
- Repositorios Institucionales DIGIPRESV para Preservación Digital a largo plazo mediante PREMIS 2.2 y OAIS ISO 14721
- Intercambio de metadatos en METS 1.9.1 (diferentes Profiles) integrando todos los esquemas de metadatos
- Creación de METSRights para el control de los derechos de autor
- Reconocimiento Óptico de Caracteres OCR y generación dinámica de ALTO (Analyzed Layout and Text Object)
- Generación e integración de registros SKOS mediante MARC 21(Up.16)/RDA
- Creación de eBooks o libros digitales en formatos: ePub y Mobipocket.
- Adaptación de DIGIBIB a Linked Open Data

ORACLE PARTNERNETWORK



Incubator Activity



Validación en el Data Providers de la Open Archives Initiative. Genera un Sitemap para Google.



ePUB



<EAC-GPF>

<ead>



Nº ES042816-1



MADUREZ DE LA GESTIÓN DE CONTENIDOS, ¿SINÓNIMO DE DESGASTE O DE OPORTUNIDADES?

Ricardo Éito-Brun



Ricardo Éito-Brun es doctor en documentación por la *Universidad de Zaragoza* y licenciado en la misma disciplina por la *Universidad de Granada*. Desempeña su actividad profesional desde 1996 en el sector de las tecnologías de la información, lo que ha compaginado con la docencia universitaria en la *Universidad Carlos III de Madrid*. Es autor de varias monografías sobre lenguajes de marcas y gestión de contenidos, y de artículos y contribuciones a congresos y reuniones profesionales. <http://orcid.org/0000-0003-1219-0510>

Grupo GMV. Departamento de Calidad
Isaac Newton 11, P.T.M.
28760 Tres Cantos, Madrid, España

Resumen

Las soluciones técnicas para gestionar contenidos han alcanzado un alto nivel de madurez. Bajo la denominación ECM (*enterprise content management*) se han agrupado varias aproximaciones técnicas para la gestión integral de la información no estructurada. La aceptación de la gestión de contenidos (GC) en distintos contextos es indicador del valor que estas prácticas aportan a las organizaciones. Pero cabe el riesgo de interpretar esta madurez como sinónimo de desgaste: se debe considerar si todavía hay oportunidades de mejora y nichos de actividad en los que centrar esfuerzos, o si por el contrario queda poco espacio para la innovación. El autor sostiene que una vez se han sentado las bases para el despliegue de este tipo de soluciones y la gestión de contenidos se ha convertido en un término familiar para los usuarios de las tecnologías de la información (TI), existe aún un gran número de oportunidades relacionadas con: a) el análisis del consumo de datos y la experiencia de los usuarios, incluyendo temas complejos como la minería de opiniones, y b) el diseño de soluciones GC que combinen las buenas prácticas de las TI y del área de la gestión de información.

Palabras clave

Gestión de contenidos, Tecnologías de la información, Gestión de la información, Profesionales de la información.

Title: Content management maturity as a synonym for obsolescence or opportunity?

Abstract

Technical solutions for content management have reached a high level of maturity. Under the heading of ECM (enterprise content management), several technical approaches for the management of unstructured information have been grouped. The acceptance of content management in different contexts is indicative of the value that these practices contribute to organizations. There is, however, a risk of interpreting this maturity as a synonym for being past its prime: we should consider whether there are still some opportunities for improvement and niches of activity on which to focus efforts, or if instead there is little room for innovation. The author contends that once we have laid the foundation for the deployment of such solutions and content management has become a familiar term for information technology (IT) users, there remain many opportunities related to: a) the analysis of data usage and user experience, including complex issues such as opinion mining, and b) the design of ECM solutions that combine best practices in the areas of IT and information management.

Keywords

Content management, Information technologies, Information management, Information professionals.

Éito-Brun, Ricardo (2013). "Madurez de la gestión de contenidos, ¿sinónimo de desgaste o de oportunidades?". *El profesional de la información*, septiembre-octubre, v. 22, n. 5, pp. 377-380.

<http://dx.doi.org/10.3145/epi.2013.sep.01>

Mayor alcance de la gestión de contenidos

Una de las preguntas más frecuentes al revisar los informes de los analistas y estudiar las nuevas características de las aplicaciones informáticas para la gestión de contenidos es si ésta ha alcanzado realmente su madurez. Atendiendo únicamente a los movimientos del mercado, se aprecia dicha madurez, aunque siempre haya incógnitas en torno a las estrategias comerciales de algunas compañías (Wood, 2013). Desde una perspectiva más amplia, se puede apuntar que la madurez alcanzada en aspectos técnicos y la incorporación de la gestión de contenidos en el portfolio de tecnologías de la información de las organizaciones –aún con un alcance limitado– es la cadena de transmisión idónea para identificar nuevas oportunidades profesionales y dotar de un mayor dinamismo a las disciplinas enfocadas en la gestión de la información.

La gestión de contenidos ha evolucionado hacia la integración progresiva de las tecnologías para gestionar información y datos no estructurados.

Inicialmente el término tuvo un significado acotado a un tipo muy concreto de programas informáticos enfocados a la edición colaborativa de sitios web y a la coordinación de los actores que participaban en su diseño y desarrollo.

Posteriormente pasó a englobar herramientas, métodos y procesos de análisis característicos de la gestión documental tradicional. Esto dio al término un significado mucho más amplio. La incorporación de métodos para el análisis de flujos de trabajo, tipos de contenidos, su organización lógica, etc., hizo que la gestión de contenidos se comenzase a entender como una disciplina soportada por soluciones técnicas, y dejó de estar limitada al despliegue de aplicaciones informáticas puntuales. Esta progresión la dotó de un mayor contenido y de una mayor complejidad, y obligó a pensar en la forma de sincronizar y reconciliar la gestión de la información no estructurada con su difusión y publicación a través de diferentes canales, incluyendo los sitios web que han caracterizado la web social.

“ El foco de la gestión de contenidos ha pasado a ser el repositorio ”

Si inicialmente el principal elemento en el que se centraba la atención en cualquier estrategia de gestión de contenidos era el sitio web corporativo, en la actualidad los sitios web son canales a través de los cuales se puede difundir y dar visibilidad de forma selectiva a la información que las organizaciones generan y recopilan. El foco de la gestión de contenidos ha pasado a ser el repositorio: que no está limitado a aquellos datos a los que se ha pensado dar visibilidad a través de internet, sino que engloba toda la información que genera y recibe la organización y es susceptible de ser utilizada –y reutilizada– en la ejecución de procesos de trabajo y en la prestación de servicios a usuarios y clientes.

Este cambio de filosofía se refleja fielmente en los acrónimos que se utilizan al hablar de gestión de contenidos. Se ha establecido una diferencia entre WCM (*web content management*) y ECM (*enterprise content management*). El

primero para hacer referencia a la gestión de los contenidos para ser difundidos en la Web, y el segundo como un súper-conjunto que englobaría a la WCM junto a la gestión documental tradicional en sus múltiples variantes: gestión de imágenes digitalizadas o *imaging*, gestión de registros o *records management*, gestión de informes y listados generados desde aplicaciones de gestión, etc.

Desde el punto de vista del profesional de la información, esta evolución añade una mayor complejidad a nuestras actividades y obliga a reformular la función y competencias que se precisan para desempeñarlas en este nuevo contexto.

“ La evolución de la gestión de contenidos añade una mayor complejidad a las actividades del profesional de la información y obliga a reformular su función y competencias ”

La gestión de contenidos no se ocupa únicamente de la edición de páginas usando plantillas y de la coordinación de equipos de editores, diseñadores gráficos y programadores conocedores de ciertas aplicaciones software. Su alcance se ha ampliado de forma significativa. Un gestor de contenidos debe ser capaz de contextualizar las actividades de la organización, entender y analizar de qué procesos se nutre el repositorio, y qué datos e información se deben exponer y facilitar a los usuarios internos, externos, y a las aplicaciones (incluyendo sitios web).

De todo esto se puede deducir que esta evolución ha tenido como consecuencia:

- un alcance más amplio para la gestión de contenidos, que ha pasado a convertirse en una disciplina de gestión global de información;
- una mayor independencia de la gestión de contenidos respecto a aplicaciones informáticas particulares. Ya no se trata de gestionar procesos de edición y publicación en la Web, sino de gestionar un repositorio que interactúe, mediante protocolos estándares, con múltiples aplicaciones que producen y consumen contenidos;
- la disociación entre *gestión de contenidos e información no estructurada*. Resulta difícil mantener la validez de esta asociación tradicional;
- la necesidad, desde un punto de vista profesional y académico, de formalizar métodos que permitan analizar y documentar las necesidades de los usuarios y desplegar soluciones para gestionar contenidos basadas en tecnologías.

Otro aspecto que requiere la atención de la comunidad profesional es el diseño de métodos para evaluar la madurez de las prácticas de gestión de contenidos de las organizaciones. Las posibilidades reales de aplicar y poner en práctica esta aproximación generalista se pueden ver afectadas por las necesidades más o menos inmediatas de solventar problemas particulares. Es complejo que una organización considere la gestión de contenidos como una gestión global de sus activos de información. Se están dando pasos en esta línea, y la gestión de contenidos web ha pasado a englobar actividades como el comercio-e, el marketing digital multi-

Bibliografía

Antion, Daniel (2012). "ECM is dead". *AIIM: The global community of information professionals*, 19 June.

<http://www.aiim.org/community/blogs/expert/ecm-is-dead>

Davenport, Thomas H.; Harris, Jeanne G.; Morison, Robert (2010). *Analytics at work: smarter decisions, better results*. Harvard Business Review Press. ISBN: 978 1422177693

Davenport, Thomas H.; Kim, Jinho (2013). *Keeping up with the quants: your guide to understanding and using analytics*. Harvard Business Review Press. ISBN: 978 1422187258

MacComascaigh, Mick; Gilbert, Mark R.; Murphy, Jim; Tay, Gavin (2013). "Magic quadrant for web content management". *Gartner*, 31 July.

Gartner, 31 July.

<http://www.gartner.com/technology/reprints.do?id=1-113UYPS&ct=130802&st=sb>

Myers, Anthony (2013). "Gartner MQ for web content management 2013: Adobe, SDL, SiteCore Lead + Market overview". *CMSWire*, 1 August.

<http://www.cmswire.com/cms/customer-experience/gartner-mq-for-web-content-management-2013-adobe-sdl-sitecore-lead-market-overview-021950.php>

Wood, Rich (2013). "Reports of SharePoint's death are greatly exaggerated". *CMSWire*, 22 August.

<http://www.cmswire.com/cms/social-business/reports-of-sharepoints-death-are-greatly-exaggerated-022206.php>

Anuario ThinkEPI 2013



272 páginas de análisis de tendencias en información, documentación y comunicación

Formulario de compra:

<http://www.thinkepi.net/solicitud>

Información y pedidos:

Isabel Olea

epi.iolea@gmail.com

+34-608 491 521



MARCADO SEMÁNTICO AUTOMÁTICO EN GESTORES DE CONTENIDOS: INTEGRACIÓN Y CUANTIFICACIÓN



Juan-Antonio Pastor-Sánchez, Enrique Orduña-Malea y Tomás Saorín



Juan-Antonio Pastor-Sánchez es doctor en documentación y profesor de la *Facultad de Comunicación y Documentación* de la *Universidad de Murcia*, en el área de la construcción de servicios y sistemas de información digital. También ha desempeñado su carrera profesional como documentalista y diseñador de sistemas de información, creación de entornos de enseñanza en red y diseño web. Colabora con el *W3C* en la traducción de material de referencia de *SKOS* y otras tecnologías de la web semántica. Investiga la aplicación de las tecnologías de la web semántica, *linked open data*, diseño de ontologías, gestión de contenidos digitales y arquitectura de la información.

<http://orcid.org/0000-0002-1677-1059>

*Universidad de Murcia. Facultad de Comunicación y Documentación
Campus de Espinardo, 30071 Murcia, España
pastor@um.es*



Enrique Orduña-Malea es doctor en documentación por la *Universidad Politécnica de Valencia (UPV)*. Investigador posdoctoral en el *Instituto de Diseño y Fabricación (IDF)* y profesor externo en el *Departamento de Comunicación Audiovisual, Documentación e Historia del Arte (DCADHA)* en la *UPV*. Desde 2012 es miembro del grupo de investigación *EC3* de la *Universidad de Granada*. Entre otras actividades, es subdirector del *Anuario ThinkEPI* y miembro del equipo del ranking de profesores *H Index Scholar*. Su área de investigación es la cibermetría (tanto descriptiva, instrumental como aplicada), orientada especialmente a la creación, difusión y consumo de información académica en internet y a la cuantificación y visualización de estas actividades.

<http://orcid.org/0000-0002-1989-8477>

*Universidad Politécnica de Valencia, Escuela Técnica Superior de Informática
Edificio 1H. Camino de Vera, s/n. 46022 Valencia, España
enorma@upv.es*



Tomás Saorín es profesor asociado en la *Facultad de Comunicación y Documentación* de la *Universidad de Murcia* y documentalista de la *Comunidad Autónoma de Murcia* donde ha participado en la puesta en marcha de proyectos de gestión de contenidos e información institucional en las áreas de servicios sociales, trabajo, empleo y bibliotecas. Participa en el capítulo español de *Wikimedia* para el conocimiento libre y en acciones de divulgación del movimiento *GLAM-Wiki* para bibliotecas, archivos y museos. Ha investigado sobre estrategia digital y de edición electrónica, así como sistemas de gestión de contenidos y proyectos colaborativos.

<http://orcid.org/0000-0001-9448-0866>

*Universidad de Murcia. Facultad de Comunicación y Documentación
Campus de Espinardo, 30071 Murcia, España
tsp@um.es*

Resumen

Se ofrece en primer lugar una visión general de los diferentes formatos de marcado semántico así como de las tecnologías existentes para incorporar información semántica explícita (microformatos, microdatos y *RDFa*). Posteriormente se describen algunos servicios que permiten automatizar los procesos de anotación semántica (*Sindice*, *Calais*, *AlchemyAPI* y *DBpedia Spotlight*) al tiempo que se caracteriza el ciclo completo de este proceso en un CMS concreto (*Wordpress*) mediante un *plugin* especializado (*RDFaCE-Lite*). Finalmente, con el propósito de cuantificar la creación y la conectividad del contenido marcado semánticamente en la Web, se analiza el conjunto de universidades españolas (y una selección de 25 universidades internacionales) mediante *Sindice*. Para ello se calculan indicadores de tamaño semántico y de enlaces entrantes, salientes, internos y de terceros (*third party links*) en los *datasets* de las universidades de la muestra. Los resultados indican la todavía escasa presencia de contenido marcado semánticamente en las universidades, así como el alto aislamiento en visibilidad web de estos contenidos.

Palabras clave

Web semántica, *Linked data*, Marcado semántico, Gestores de contenidos, *Wordpress*, *RDFaCE-Lite*, *Sindice*, Universidades, Webometría.

Artículo recibido el 15-06-2013
Aceptación definitiva: 18-09-2013

Title: Automatic semantic markup in content management systems: integration and quantification

Abstract

A general overview of the different semantic markup formats and the existing technologies to incorporate explicit semantic information (microformats, microdata and RDFa) is provided. Services are described that automate, to some extent, semantic annotation processes (*Sindice*, *Calais*, *AlchemyAPI* and *DBpedia Spotlight*), while characterizing the complete cycle of this process in a particular CMS (*Wordpress*) using a specialized plugin (*RDFaCE-Lite*). Finally, in order to quantify the creation and connectivity of semantically marked content on the Web, the space formed by all Spanish universities (and a selection of 25 international institutions) is analysed with *Sindice*. Semantic page count and visibility indicators (inlinks, outlinks, internal and third party) are calculated for the sample. The results indicate limited presence of semantically marked content in the universities and highly isolated web visibility of this content.

Keywords

Semantic web, Linked data, Semantic markup, Content management systems, *Wordpress*, *RDFaCE-Lite*, *Sindice*, Universities, Webometrics.

Pastor-Sánchez, Juan-Antonio; Orduña-Malea, Enrique; Saorín, Tomás (2013). "Marcado semántico automático en gestores de contenidos: integración y cuantificación". *El profesional de la información*, septiembre-octubre, v. 22, n. 5, pp. 381-391.

<http://dx.doi.org/10.3145/epi.2013.sep.02>

1. Introducción

El ecosistema que conforma la web semántica suele tratarse como una propuesta a largo plazo. Actualmente existe un intenso foco de interés en conectar conjuntos de datos (*datasets*), dando lugar a lo que se denomina *linked open data* (datos enlazados). Este fenómeno sigue unas pautas ampliamente conocidas y sugeridas inicialmente por **Tim Berners-Lee**. En la Web se publican grandes conjuntos de datos de una manera estructurada siguiendo dichas pautas, al tiempo que se definen de forma explícita las relaciones entre recursos mediante uno o varios vocabularios, como por ejemplo RDF, OWL (*web ontology language*) o SKOS (*simple knowledge organization system*) entre otros (**Pastor-Sánchez**, 2011). *DBpedia* es el *dataset* más utilizado en el universo *linked open data*.

<http://www.w3.org/wiki/LinkedData>

<http://dbpedia.org>

El funcionamiento eficiente de *datasets* interconectados ofrece enormes posibilidades de aplicación práctica, desde la mejora de la recuperación de información en motores de búsqueda generalistas al diseño de servicios de valor añadido en ámbitos de uso intensivo de información, como la documentación científica (**García-García**, 2012) o las bibliotecas (**Sellés-Carot**; **Orduña-Malea**; **Serrano-Cobos**, 2013) entre otros, aunque sus aplicaciones aún están en fase incipiente, pendientes de masa crítica y de *killer apps* (**Saorín**; **Peset**; **Ferrer-Sapena**, 2013).

Es preciso conceptualizar una dualidad, en la que *datasets* y contenidos web deberían dar forma a una realidad en la que los procesos de publicación partieran de la interrelación y la consiguiente integración de grandes cantidades de datos y contenidos informativos.

Los sistemas de gestión de contenidos (CMS) han alcanzado durante la última década un alto grado de desarrollo conceptual desde el punto de vista de la arquitectura de

la información, la interconexión con bases de datos corporativas, la modularidad y la reutilización de contenidos. La disponibilidad en abierto de muchas de estas aplicaciones y su facilidad de instalación y manejo han propiciado un uso masivo por parte de los cibernautas.

Según *Builtwith* -a mayo de 2013- los tres CMS más utilizados son *Wordpress*, *Joomla!* y *Drupal*, sumando entre los tres casi 10 millones de instalaciones. Por su parte *W3Tech* apunta que el 58% de los sitios web gestionados mediante CMS utilizan *Wordpress*, mientras que los sitios que no utilizan ningún CMS se han reducido en un 10% durante los últimos 3 años.

<http://trends.builtwith.com/cms/top>

<http://w3techs.com>

Esto no hace más que poner de manifiesto el enorme peso que, en la creación y difusión de contenidos web a nivel global, tienen actualmente los CMS, que han ido a su vez evolucionando con el tiempo:

1ª fase: contenido y presentación mezclados;

2ª fase: contenido y presentación separados;

3ª fase: contenido, presentación y significado separados.

<http://www.webnodes.com/why-semantic-cms>

Aun así, los CMS tienen todavía una escasa participación en el ciclo de vida (creación, publicación y reutilización) de los *datasets* enlazados y, en consecuencia, existe un bajo nivel de integración de éstos con los contenidos web (**Pastor-Sánchez**, 2012). Dicha integración supone un reto y una necesidad, puesto que actualmente el uso cada vez mayor conlleva una "larga cola" en cuanto a la producción de contenidos web, que aporta una diversidad informativa insustituible.

Dadas las facilidades que los CMS aportan en la gestión de sitios web (diseño, usuarios, contenidos, servicios, etc.), resulta fundamental que éstos se expandan y participen igual-

mente en el ciclo de vida de la gestión de conjuntos de datos a partir del marcado semántico de contenidos, analizado más adelante. La clásica idea de una web para máquinas y una web para personas encuentra un punto de convergencia en este enfoque. Se trata de un flujo en el que contenidos y conjuntos de datos intercambian información.

La extracción de información semántica de los propios contenidos web, publicados de forma extremadamente distribuida, permitiría mejorar la eficiencia de los motores de búsqueda y ofrecería un mecanismo para la creación automática de conjuntos de datos semánticos. Por otro lado, el marcado semántico abre nuevas posibilidades para la reutilización de *datasets* como fuente de información semántica que sería incorporada a los contenidos.

Este trabajo muestra y analiza las posibilidades que aportarían los procesos de marcado semántico en la gestión de contenidos mediante CMS. Asimismo se utilizan ciertos indicadores web para determinar el nivel de desarrollo y aplicación del marcado semántico.

El marcado semántico de contenidos supone la incorporación explícita de información a documentos web, de forma que éstos sean inteligibles y procesables por parte de aplicaciones informáticas

2. ¿Qué es el marcado semántico de contenidos web?

El marcado semántico de contenidos supone la incorporación explícita de información a documentos web, de forma que éstos sean inteligibles y procesables por parte de aplicaciones informáticas. Básicamente se trata de añadir determinados atributos (de un modo transparente al lector) al marcado (x)html, con la finalidad de identificar tanto objetos como propiedades de los mismos, así como las relaciones entre ellos.

2.1. Formatos de marcado semántico

Con independencia del esquema de descripción utilizado, los principales formatos usados en el marcado semántico son microformatos, *RDFa* y microdatos.

Microformatos

La técnica de los microformatos (Khare; Çelik, 2006) se basa en el uso de xhtml (versiones 1 y 5) y html (versiones 4 y 5) de forma combinada con css. En esencia se utiliza el atributo "class" para identificar vocabularios, clases y propiedades. La clave de su éxito inmediato y rápida difusión se encuentra en el uso de tecnologías ya existentes, y en la adopción de una serie de convenciones muy sencillas para referirse a los vocabularios y elementos de descripción utilizados. Sin embargo, uno de sus principales problemas es la ambigüedad producida por la combinación, en un mismo atributo "class", de la información relativa a los estilos de formato visual y a la descripción del contenido semántico de un elemento. Tampoco es posible utilizar URIs para identificar los

elementos sobre los que se realiza la descripción y no existe una correspondencia clara con RDF, que entorpece la extracción de información interoperable.

RDFa

La solución del W3C para el marcado semántico es *RDFa*, fundamentada en una total compatibilidad con la tecnología clave de la web semántica: RDF. Su última versión (*RDFa* 1.1) utiliza un conjunto de atributos específicos y la reutilización de otros ya existentes para la inclusión de información semántica.

<http://www.w3.org/TR/rdfa-syntax>

La especificación contempla su uso en diferentes lenguajes "anfitrión" como xhtml (versiones 1 y 5), html (versiones 4 y 5), xml, SVG (*scalable vector graphics*), ePUB y *OpenDocument*. La potencia de *RDFa* contrasta con cierta complejidad en su aplicación. Su creación ha precisado de la definición de nuevos atributos y la reutilización de otros. Por este motivo, el W3C desarrolló la recomendación *RDFa Lite*, que indica cómo utilizar *RDFa* de un modo más simplificado, cubriendo las necesidades más comunes de marcado semántico.

<http://www.w3.org/TR/rdfa-primer>

<http://www.w3.org/TR/rdfa-lite>

Podría decirse que *RDFa* es una alternativa flexible al tiempo que potente para la inclusión en una página web de cualquier tipo de sentencia RDF.

Microdatos

La tercera opción para el marcado semántico son los microdatos, cuyo desarrollo está ligado al de html5. El grupo de trabajo *Whatwg* se creó en 2004 cuando el W3C decidió apostar por las tecnologías de marcado basadas en xml, como xhtml, abandonando html. El objetivo de dicho grupo era la creación de un lenguaje de marcado con nuevos elementos de marcado estructural, de fácil aplicación en dispositivos con baja disponibilidad de recursos y susceptible de integrar capacidades de reproducción de contenido multimedia.

<http://www.whatwg.org>

Actualmente los trabajos de estandarización de html5 están coordinados por el W3C, puesto que este organismo decidió no continuar con el desarrollo de xhtml2. Se decidió que la especificación de html5 podría utilizarse con dos serializaciones: una basada en una DTD (html5) y otra basada en xml (xhtml5). Los microdatos son la solución nativa de html5 para incorporar contenido semántico. Además, el proyecto *Schema.org*, liderado por Google¹, ha supuesto un espaldarazo a este formato de marcado semántico. La propuesta define propiedades nuevas (aunque menos que en el caso de *RDFa*) y reutiliza otras ya existentes.

<http://schema.org>

<http://www.w3.org/TR/microdata>

El más potente de los tres formatos es sin duda *RDFa*, pero también es la opción más compleja: su aplicación implica cierto dominio y comprensión de RDF, e incluso de OWL si se desea una estructuración formal de los datos semánticos incorporados en el marcado. En cuanto a la formalización semántica, los microformatos representan la opción con mayores carencias. Los microdatos ofrecen una solución in-

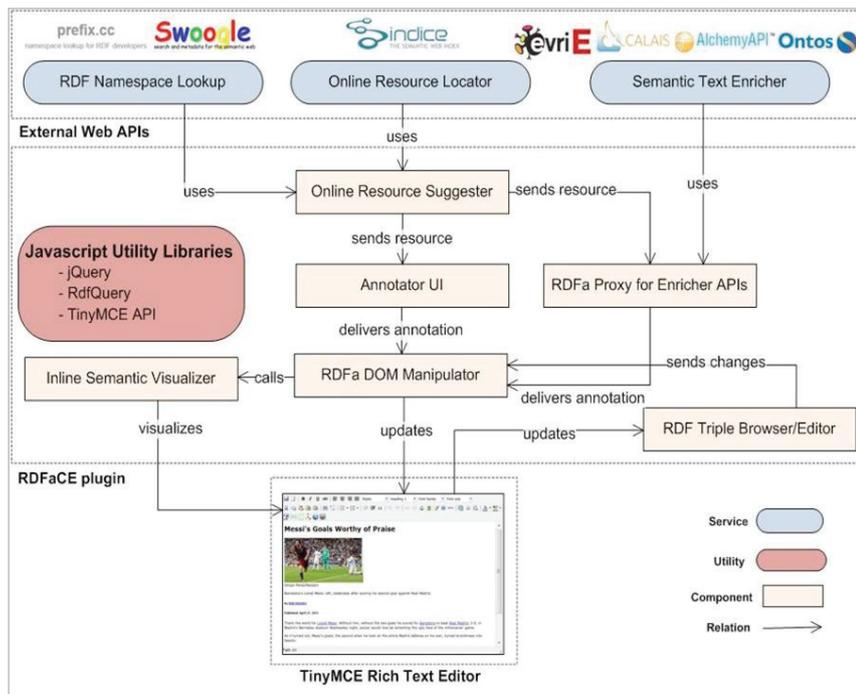


Figura 1. Arquitectura funcional de *RDFaCE* (fuente: Khalili; Auer; Hladky, 2012)

termedia que pueden utilizarse con la mayoría de vocabularios RDF, además de *Schema.org*.

2.2. Marcado semántico y CMS

Los CMS son herramientas de gran ayuda en el proceso de marcado semántico en cualquiera de los tres formatos. Incluso es posible combinar los tres formatos dentro de un mismo documento, aumentando así las posibilidades de procesamiento automático de tales contenidos. Puesto que el *backend* de un CMS genera el código (x)html, el administrador de un sitio web únicamente deberá definir las equivalencias entre las estructuras de contenido con los elementos de uno o varios vocabularios de descripción. A partir de ahí, el propio CMS se encarga de realizar el marcado adecuado. Son numerosas las extensiones en distintos CMS que implementan esta funcionalidad de un modo sencillo y muy escalable. Aquí cabe destacar la potencia de *Drupal* (a partir de su versión 7) que incorpora el marcado semántico mediante *RDFa*, tanto en su núcleo como mediante extensiones. El tipo de información susceptible de este enfoque de marcado semántico es información muy estructurada: datos de personas, productos, libros, películas e incluso recetas de cocina.

http://www.w3.org/wiki/Mixing_HTML_Data_Formats
<https://drupal.org/project/rdfx>

Pero no hay que olvidar que la mayor parte de los contenidos web no están sujetos a estructuras tan definidas. La mayor parte de los CMS siguen organizando los contenidos usando un conjunto de metadatos descriptivos (título, autor, fecha) asociados a un campo de texto principal sobre el que se pueden aplicar distintos tipos de formatos, definir enlaces o insertar imágenes. Sin embargo, es precisamente en el cuerpo del contenido web donde se encuentra realmente el mensaje en sí mismo. Por tanto, la cuestión es ¿qué sucede con el texto ubicado en el clásico campo “body” desde el punto de vista del marcado semántico?

2.3. Servicios de datos semánticos para el marcado semántico automático

Existen algunas herramientas que amplían las funciones de los CMS para el marcado semántico manual del cuerpo de los contenidos. Pero realizarlo manualmente resulta tedioso, poco operativo e incluso puede ocasionar un marcado deficiente y con errores. Existen proveedores de servicios que facilitan este trabajo, sugiriendo e incluso realizando directamente el marcado. Su uso suele realizarse a través de APIs (*application programming interfaces*) que permiten su portabilidad a múltiples plataformas. Los CMS se conectan con estos servicios mediante extensiones que hacen de puente entre dichas APIs y el motor de marcado de contenido, generalmente basado en un editor *wysiwyg*. Las APIs devuelven los resultados en formatos como xml, RDF o json (*javascript object notation*).

AlchemiAPI

Se trata de un *saas*² basado en técnicas de procesamiento del lenguaje natural (PLN). Se utiliza mediante una API que interactúa con un servicio web o un *kit* de desarrollo disponible en múltiples lenguajes de programación. Esta API es en realidad una interfaz que permite la explotación de bases de datos estadísticas y lingüísticas mediante un conjunto de complejos algoritmos. Tras suministrar un texto, en alguno de los idiomas soportados, es posible identificar personas, organizaciones, etiquetas, conceptos y relaciones entre distintos tipos de entidades incluso pertenecientes a conjuntos de datos LOD externos.

<http://www.alchemyapi.com>

“ Con independencia del esquema de descripción utilizado, los principales formatos usados en el marcado semántico son: microformatos, *RDFa* y microdatos ”

Sindice

Este servicio en línea básicamente es un motor de búsqueda que recopila documentos RDF y los indiza con el objeto de facilitar la búsqueda de recursos (Oren *et al.*, 2008). Localiza documentos que contengan sentencias sobre un determinado recurso. En el marcado semántico resulta útil para definir sentencias sobre dichos recursos a partir de las entidades identificadas en el contenido web. *Sindice* indiza documentos web que utilicen cualquier tipo de formato para el marcado semántico y dispone de APIs para la búsqueda y la consulta de los índices, así como de herramientas para la extracción de tripletas RDF y la validación del marcado.

<http://www.sindice.com>

OpenCalais

Es un servicio web de *Thomson-Reuters* que también utiliza técnicas de PLN para el enriquecimiento semántico de contenidos web. El funcionamiento sigue la misma pauta que en *AlchemiAPI*: se envía un texto al servicio web y éste, tras analizarlo, localiza entidades, hechos y eventos. Ésta es una diferencia significativa con *AlchemiAPI*, que únicamente se limita a la identificación de entidades. En principio este servicio gratuito está limitado a 50.000 transacciones diarias. Su uso de la API puede realizarse mediante invocaciones a servicios web soap (*simple object access protocol*) o rest (*representational state transfer*). El servidor responde devolviendo código html marcado con microformatos o declaraciones RDF en los formatos json, N3 (notation3) o *simple format*. <http://www.opencalais.com>

DBpedia Spotlight

Este servicio permite la aplicación de la ontología de *DBpedia* en tareas de marcado semántico, basándose en técnicas enfocadas a la desambiguación semántica. Permite identificar entidades de una de las 272 clases de la ontología (*Mendes et al.*, 2011). *Spotlight* puede usarse a través de un servicio web restful/soap o utilizando el archivo java ofrecido por este proyecto para desarrollar una aplicación web. Mediante *Spotlight* es posible reconocer entidades, anotar un texto o desambiguar un término. <http://spotlight.dbpedia.org>

2.4. Caso de uso: marcado semántico en Wordpress mediante RDFaCE

Como se ha expuesto anteriormente, los CMS comienzan a permitir, generalmente mediante extensiones, el marcado semántico de estructuras y metadatos, aunque el marcado del cuerpo de texto está todavía menos desarrollado.

En *Drupal*, por ejemplo, se dispone del módulo *Semantic markup editor*. Se trata de un proyecto en plena fase de desarrollo inicial para el marcado semántico del contenido dentro del editor *wysiwyg*. Otros módulos para este CMS son *OpenCalais* y *Alchemy*, que hacen uso de las correspondientes APIs para el marcado automático de los contenidos. Otro CMS muy extendido, *MediaWiki*, dispone de la extensión *Semantic MediaWiki* para definir propiedades y atributos, incorporando semántica a los "wikienlaces". Su uso con otros módulos para formularios o *RDFa*, permite trabajar una wiki como si fuera una base de datos, utilizando un lenguaje de interrogación y generando contenidos dinámicos. Un ejemplo de este tipo de funciones es *Referata*, un servicio de alojamiento de wikis semánticas cuyos datos pueden gestionarse y reutilizarse de forma colaborativa. https://drupal.org/project/semantic_markup_editor

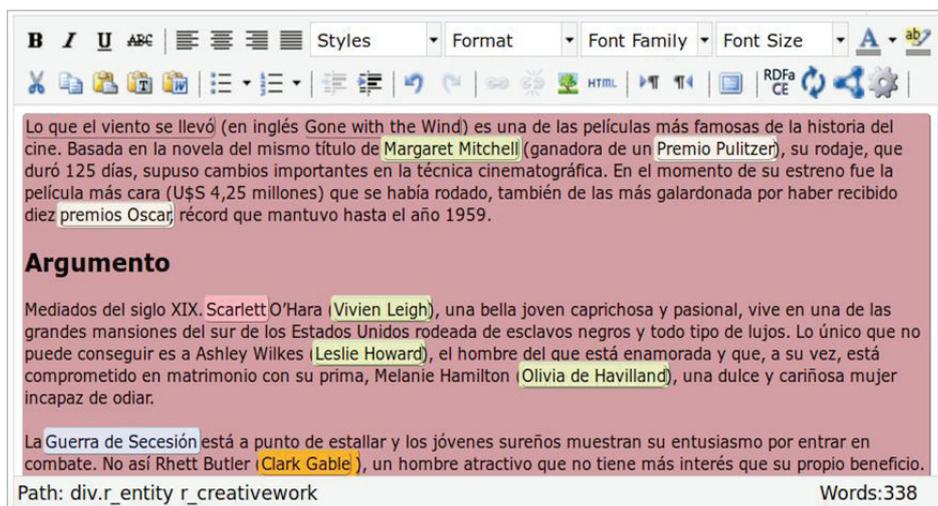


Figura 2. Ejemplo de marcado semántico (manual y automático) de un fragmento de texto usando la instalación demo de *RDFaCE* <http://rdface.aksw.org/new/tinymce/examples/rdface.html> (fuente: <http://rdface.aksw.org>)

- <http://drupal.org/project/opencalais>
- <http://drupal.org/project/alchemy>
- <http://semantic-mediawiki.org>
- <http://www.referata.com>

Por su parte, *Wordpress* dispone de un *plugin* denominado *RDFaCE* que combina la edición manual y el marcado automático en el propio cuerpo del contenido de forma interactiva. Dado el predominio de *WordPress* en el mercado web, este *plugin* se analiza a continuación con mayor detalle (figura 1).

RDFaCE realiza el marcado semántico utilizando *RDFa* y microdatos y aplicando los siguientes vocabularios:

- *rNews*: esquema de metadatos desarrollado por *IPTC* (*International Press Telecommunications Council*) para la descripción de recursos relacionados con los medios de comunicación. http://www.iptc.org/std/rNews/1.0/specification/rNews_1.0-diagram.pdf
- *foaf* (*friend of a friend*): permite describir personas, sus actividades y las relaciones con otras personas, organizaciones y objetos. <http://www.foaf-project.org>
- *DBpedia*: ontología para la representación de información extraída de *Wikipedia*. <http://dbpedia.org>
- *Schema*: esquema de ámbito general que permite representar una amplia variedad de recursos: personas, trabajos creativos, organizaciones, eventos, etc. <http://schema.org>

El marcado manual se efectúa seleccionando parte o la totalidad del texto del cuerpo del contenido e identificando el tipo de entidad al que se hace referencia: persona, evento, organización, noticia, lugar, objeto audiovisual, etc. En el texto correspondiente a la entidad identificada, se seleccionan fragmentos más específicos para definir sus propiedades: nombre, URL, fechas, etc. También es posible establecer relaciones entre diferentes entidades (figura 2).

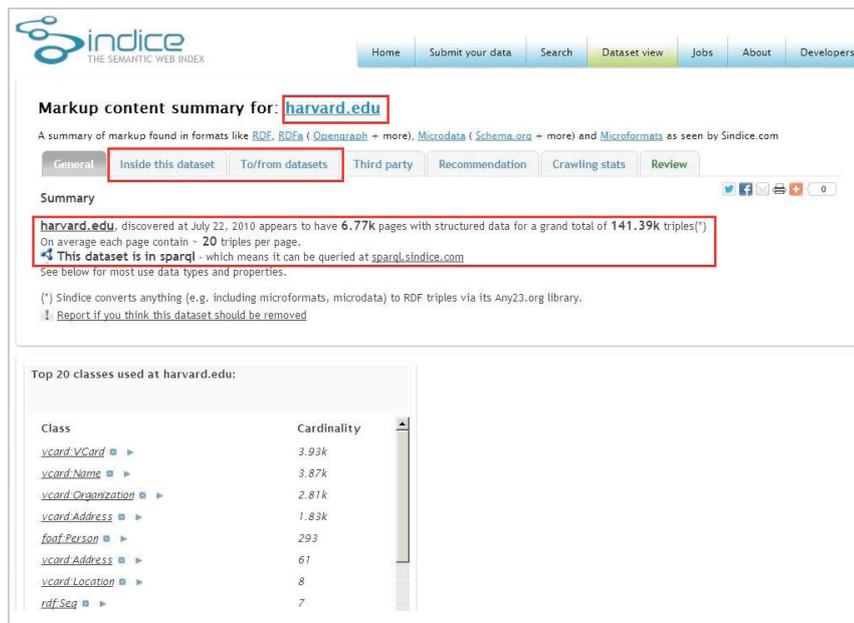


Figura 3. Análisis de datasets en dominios web mediante Sindice <http://sindice.com>

La auténtica potencialidad de *RDFaCE* puede verse en el uso que hace de tres tipos de servicios semánticos:

- Buscador de espacios de nombres RDF: utiliza *Swoogle* y *Prefix.CC* para identificar los espacios de nombres más adecuados para hacer referencia a las entidades identificadas.
- Localizador de recursos online: busca recursos web a partir de los espacios de nombres identificados previamente, y utiliza los URLs encontrados para marcar como enlaces el texto de las entidades identificadas.
- Enriquecedor semántico: realiza el marcado semántico automático utilizando alguno de los siguientes servicios: *DBpedia Spotlight*, *Ontos*, *OpenCalais*, *AlchemyAPI*, *Extractiv*, *Evri*, *Lupedia* o *Saplo*.

Es necesario puntualizar que *RDFaCE* debe ser configurado para poder utilizar algunos de estos servicios mediante una clave para el uso de la API correspondiente (por ejemplo en el caso de *OpenCalais*).

3. Análisis métrico de la información semántica web

La creación, evolución y extensión del marcado semántico del contenido web, dada su estructuración estandarizada, puede cuantificarse mediante el uso de ciertos indicadores. Esto supone un campo incipiente de análisis de la cibermetría, orientada al análisis métrico de información semántica en la Web.

Hasta la fecha los estudios relativos a la aplicación de técnicas ciber métricas a *linked data* son escasos. **Longqing y Qingfeng** (2011) estudian el uso de metadatos *Dublin Core* (DC) como fuente de datos para la cibermetría, analizando especialmente la existencia de los elementos "source" o "relation" (que permiten explicitar el tipo de relación entre recursos enlazados mediante URLs). Sin embargo, los autores concluyen que estos metadatos no son suficientes por sí mismos y reclaman la necesidad de publicar mediante pa-

trones de expresión estándar así como una mayor gestión de la organización de recursos web. La existencia de contenidos no estructurados provoca una situación caótica, en la que los robots de los motores de búsqueda no pueden proporcionar información en las condiciones que precisa la cibermetría: la búsqueda directa en etiquetas genera un enorme ruido documental que anula la validez de cualquier análisis estadístico.

El creciente uso de los diversos formatos de marcado semántico explicados anteriormente permite, de una manera más precisa, estructurar las relaciones entre recursos web. Esto ha permitido desarrollar diversas iniciativas para añadir información semántica a los enlaces cuando se aplican en diversos indicadores. Así como el *PageRank* y otros algoritmos similares como *Hits*

(**Kleinberg**, 1999) o *SimRank* (**Jeh; Widom**, 2002) no tienen en cuenta los tipos de relaciones que los enlaces están proporcionando, otras iniciativas como *ObjectRank* (**Balmin; Hristidis; Papakonstantinou**, 2004) o *TripleRank* (**Franz et al.**, 2009) sí permiten explicitar las relaciones semánticas proporcionadas por los enlaces.

Por otro lado, se han llevado a cabo estudios pioneros en la elaboración de rankings de universidades basados en la información existente sobre éstas en determinados datasets. Por ejemplo, **Meymandpour y Davis** (2013) utilizan la información estructurada depositada en *DBpedia* para diseñar un ranking mundial de universidades a partir de los datos contenidos en las relaciones "dbo:", logrando correlaciones significativas ($r = 0,85$; $p < 0,01$) con el *Ranking de Shanghai* (*Academic Ranking of World Universities*). <http://www.arwu.org>

La web semántica constituye una fuente fundamental de información en entornos académicos, desde un punto de vista métrico. **Stuart** (2012) analiza la funcionalidad de 5 fuentes de información semántica (seleccionando finalmente *Sindice*) para medir el uso de *foaf* en el seno del espacio académico británico. Para ello, recopila una serie de nuevas métricas web: número de documentos con al menos una determinada propiedad o relación (en este caso "foaf:name"), número total de declaraciones RDF o presencia de ciertos predicados dentro del dataset analizado. Estos indicadores se aplican a toda la cobertura de *Sindice* o a un determinado dominio web (por ejemplo una universidad), en el más puro estilo ciber métrico clásico.

No obstante, la aplicación de análisis métricos en el contexto de la web semántica es todavía escasa, así como el estudio y propuesta de nuevas fuentes e indicadores. Para suplir esta carencia, en esta última parte del artículo se propone la realización de un sencillo estudio con el que se pretende:

- Analizar las prestaciones de *Sindice* como fuente de información para estudios ciber métricos.

Tabla 1. Tamaño semántico y visibilidad web en *datasets* en universidades internacionales (*Sindice*)

Universidad	Dominio	Tamaño semántico	Enlaces internos	Enlaces de terceros	Enlaces entrantes	Enlaces salientes
Harvard University	harvard.edu	6.770	8.660	0	0	6
Stanford University	stanford.edu	13.350	3.020	4.970	111	569
Massachusetts Institute of Technology	mit.edu	10.840	2.820	350	140	204
University of Michigan	umich.edu	185	149	0	0	0
University of Pennsylvania	upenn.edu	206	184	6	0	1
University of California-Los Angeles	ucla.edu	2.170	1.610	0	0	0
University of California-Berkeley	berkeley.edu	266	2.020	0	0	17
Columbia University New York	columbia.edu	643	5.950	0	0	13
Cornell University	cornell.edu	290	2.220	0	0	1
University of Minnesota	umn.edu	312	179	0	0	0
Pennsylvania State University	psu.edu	3.100	2.460	0	0	0
University of Texas Austin	utexas.edu	690	651	1	0	3
Yale University	yale.edu	892	377	18.760	0	19
University of Cambridge	cam.ac.uk	505	1.980	0	0	29
Caltech	caltech.edu	2.510	963	0	0	0
University of Oxford	ox.ac.uk	8.310	13.030	2	1.730	99
Duke University	duke.edu	1.510	841	0	0	1
University of Wisconsin Madison	wisc.edu	293	1.150	0	0	0
Universidade de São Paulo (USP)	usp.br	1.260	1	0	0	1
University of Illinois-Urbana Champaign	illinois.edu	13	0	0	0	0
University of North Carolina-Chapel Hill	unc.edu	310	545	16	0	2
University of British Columbia	ubc.ca	9.020	13.570	0	0	2
University of Washington	washington.edu	3.710	2.750	0	0	1
Princeton University	princeton.edu	165	52	0	0	3
University of Utah	utah.edu	60	1.040	0	0	0

Las posiciones de las universidades responden a la posición obtenida en el ranking web, únicamente en tanto que fuente de la que se ha extraído la muestra de estudio

b) Describir indicadores de tamaño y enlazabilidad a nivel de *datasets*.

c) Aplicar los indicadores anteriores a los *datasets* de un conjunto de dominios web.

3.1. Método de trabajo

Tomando como referencia el *Ranking web of universities*, se recogió la totalidad de universidades del sistema español, tanto públicas como privadas (75), y las 25 primeras universidades del ranking mundial. Pese a que ciertas universidades utilizan varios dominios web, para cada universidad sólo se tomó el url recogido en el ranking web pues, según la metodología de este servicio, es el que mejores resultados ofrece para cada institución.

<http://www.webometrics.info>

La elección de universidades (en lugar de otras organizaciones o productos) se debe a que son instituciones analizadas extensamente en estudios cibernéticos y que, dado su tamaño y misiones fundacionales, son susceptibles de generar grandes cantidades de contenidos web estructurados, semiestructurados y no estructurados.

La muestra de universidades españolas comprende la totalidad del sistema universitario español, para conocer de

manera general y comparada su rendimiento. En el caso de las universidades internacionales, la muestra recoge únicamente las 25 instituciones con mayor cantidad de documentación en abierto (tamaño de su web) y con una mayor visibilidad (enlaces externos recibidos), de manera que los datos tengan la escala (en volumen) adecuada para ser representativos.

Los datos de las 100 universidades (75 españolas y 25 internacionales) se tomaron durante la última semana de mayo de 2013 directamente de *Sindice*, a partir del servicio *Dataset view*. Este servicio permite detectar, para cada dominio web analizado, la cantidad de páginas con marcado semántico en cualquier formato y que en este trabajo se denomina "tamaño semántico" (*semantic page count*). Se indica igualmente el número de tripletas existentes en el total de páginas (figura 3).

<http://demo.sindice.net>

Por otro lado, se recogen los datos de los siguientes indicadores de enlaces:

a) Enlaces internos en *datasets*: número de enlaces donde el sujeto y el objeto de la tripleta pertenecen al mismo dominio.

b) Enlaces de terceros (*third party*) en *datasets*: número de

enlaces, creados por el dominio web analizado, pero donde el sujeto y/o el objeto de la tripleta no pertenecen a este dominio.

c) Enlaces entrantes a *datasets*: número de enlaces dirigidos hacia algún recurso de un *dataset* albergado en un determinado dominio web.

d) Enlaces salientes de un *dataset*: número de enlaces que salen desde algún recurso de un *dataset* albergado en un determinado dominio web.

Los enlaces de terceros no se deben confundir con los *external links* usados en cibermetría clásica. Este indicador cuantifica si dentro del *dataset* de un dominio determinado se ha creado una declaración RDF donde el objeto o predicado (o ambos) es un URL que no pertenezca al dominio estudiado.

Los 5 indicadores utilizados (tamaño semántico, enlaces internos, de terceros, entrantes y salientes) no se han aplicado hasta la fecha en ningún estudio cibernético y, por tanto, suponen una novedad en este tipo de trabajos.

Los valores obtenidos son de utilidad para conocer directamente el rendimiento (en tamaño y visibilidad web) de los *datasets* analizados y, de manera indirecta, podrán ser utilizados para cuantificar el efecto de la adición de contenido semántico desde CMS específicos, debido a la posibilidad de conocer el origen de los enlaces a recursos alojados en determinados *datasets*.

3.2. Resultados del análisis cuantitativo

Los datos obtenidos no pretenden en ningún momento mostrar un ranking de universidades, sino ofrecer el tamaño y visibilidad (medida a través de hiperenlaces) de las universidades con mayor impacto en la Web. Los datos obtenidos para las 25 universidades internacionales se muestran en la tabla 1, mientras que las 75 universidades españolas se ofrecen en la tabla 2.

En cuanto a las universidades internacionales, los resultados relativos al tamaño semántico indican datos dispares y en general discretos. Sólo 5 universidades superan las 10.000 páginas con marcado semántico (*Harvard, Stanford, MIT, Oxford, British Columbia*). El resto muestra resultados prácticamente vacíos si se comparan con los tamaños documentales totales de estas universidades (*Harvard*, por ejemplo, supera los 16 millones de documentos indizados por *Google*).

Los indicadores de enlaces muestran asimismo resultados muy bajos. Los datos indican tasas de enlazado interno bajas (a excepción de algunas universidades concretas, como *Oxford, British Columbia* o *Harvard*) y tasas de enlazado de terceros prácticamente inexistentes (con la excepción de *Stanford* y, especialmente *Yale*). Respecto a los enlaces entrantes y salientes, la práctica totalidad de los resultados son muy bajos o directamente inexistentes, destacando únicamente *Oxford* en número de enlaces entrantes.

En las universidades españolas se observan patrones similares: cierta aleatoriedad en los datos, una lógica mayor presencia de enlaces internos respecto a los de terceros y, especialmente, los bajos niveles generales de enlaces entrantes y salientes. Hasta en 10 universidades (todas ellas

privadas) no se obtienen datos (es decir, no están siquiera recogidos en el catálogo de URLs de *Sindice*). Y en las 65 restantes, 24 no contienen ninguna página con contenido marcado semántico. Entre las 41 universidades con al menos 1 página con contenido marcado semántico, destacan especialmente los niveles logrados para la *Universidad Politécnica de Madrid* (58.530), reflejo directo de la actividad de determinados grupos de investigación, la *Universitat de Girona* (9.560) y la *Universidad de Oviedo* (4.630). <http://www.oeg-upm.net>

En el caso de los enlaces internos, destacan la *Universitat de Girona* (52.790) e, inesperadamente, la *Universitat Abat Oliba* (2.050), aunque los resultados son en general bajos o inexistentes (hasta 34 universidades no tienen ningún enlace interno, y 61 ningún enlace de terceros).

En cuanto a los enlaces salientes a *datasets*, los datos son prácticamente inexistentes: sólo 15 generan algún enlace saliente, destacando la *Politécnica de Valencia* (315) y la *Universidad de Valladolid* (152), mientras que ninguna universidad española recibe enlaces entrantes a *datasets*. Este resultado muestra el completo aislamiento en visibilidad del ya escaso contenido generado con marcado semántico.

3.3. Conclusiones del análisis cuantitativo

La asimetría entre el número de enlaces entrantes e internos es un indicio de cierta endogamia en la generación de enlaces, al primar relaciones entre recursos alojados en *datasets* de un mismo dominio. Esto se explica porque la mayoría de los enlaces generados sirven para explicitar relaciones verticales entre recursos de una misma colección.

Por otro lado, la ausencia de enlaces entrantes refleja una baja visibilidad web de dichos recursos. Es decir, los recursos web incluidos en los *datasets* no se enlazan entre ellos (ni se reciben enlaces ni se proporcionan), lo que resulta inesperado dado que las pautas de *linked data* promocionan precisamente el enlazado (en este caso con información semántica incluida) entre recursos web. Los resultados indican en cambio cierto aislamiento de los recursos web marcados semánticamente.

Estas bajas tasas de presencia (74,6% de las universidades españolas no alcanzan los 100 documentos marcados semánticamente) y visibilidad (ninguna universidad española recibe un solo enlace externo) de los recursos web pueden ser un reflejo de la situación descrita en la introducción: los *datasets* están siendo creados en determinados lugares y bajo determinados proyectos específicos. Sin embargo, el hecho de que no se estén generando contenidos marcados semánticamente desde CMS (las aplicaciones generalmente utilizadas para crear sitios web en la actualidad) puede provocar que la cantidad y enlazabilidad de los actuales recursos web alojados en *datasets* sean muy limitados. Este aspecto es de especial importancia en dominios web tan complejos como el de las universidades, donde conviven multitud de CMS diferentes, aunque esto resulte opaco para los usuarios.

No obstante, la muestra analizada es muy reducida y los resultados deberán complementarse en el futuro con muestras más amplias de instituciones, así como de otras organizaciones.

Tabla 2. Tamaño semántico y visibilidad web en *datasets* en las universidades españolas (*Sindice*)

Universidad	Url	Tamaño semántico	Enlaces internos	Enlaces de terceros	Enlaces entrantes	Enlaces salientes
Universidad Complutense de Madrid	uclm.es	1	374	0	0	6
Universitat Politècnica de Catalunya	upc.edu	600	1.150	0	0	1
Universidad Politécnica de Madrid	upm.es	58.530	22	0	0	0
Universitat Autònoma de Barcelona	uab.cat	137	57	0	0	0
Universidad de Granada	ugr.es	816	937	0	0	0
Universitat de Barcelona	ub.edu	3.780	2.840	0	0	8
Universidad de Valencia	uv.es	61	497	0	0	0
Universidad Politécnica de Valencia	upv.es	331	4.160	19	0	315
Universidad de Zaragoza	unizar.es	0	0	0	0	39
Universidad de Sevilla	us.es	871	660	3	0	18
Universidad Autónoma de Madrid	uam.es	3	28	0	0	5
Universidad del País Vasco	ehu.es	48	9	0	0	0
Universidad de Alicante	ua.es	111	247	0	0	0
Universidad de Vigo	uvigo.es	1	0	0	0	0
Universidad de Salamanca	usal.es	0	0	0	0	0
Universidad de Navarra	unav.es	18	0	0	0	0
Universidad de Santiago de Compostela	usc.es	0	0	0	0	0
Universitat Pompeu Fabra	upf.edu	2	5	0	0	1
Universidad de Murcia	um.es	139	433	0	0	4
Universidad de Málaga	uma.es	0	0	0	0	0
Universidad de Valladolid	uva.es	181	822	0	0	152
Universidad de Castilla la Mancha	uclm.es	1	0	0	0	0
Universitat Jaume I	uji.es	229	222	0	0	1
Universidad Carlos III de Madrid	uc3m.es	0	0	0	0	0
Universidad de Córdoba	uco.es	315	6	0	0	0
Universitat de Girona	udg.edu	9.560	52.790	3	0	0
Universidad de La Coruña	udc.es	78	108	0	0	0
Universidad de Cádiz	uca.es	324	1	0	0	0
Universitat de les Illes Balears	uib.es	2	4	0	0	0
Univ. Nacional de Educación a Distancia	uned.es	1	0	0	0	0
Universidad de Oviedo	uniovi.es	4.630	188	0	0	2
Universidad de León	unileon.es	0	0	0	0	0
Universidad de Las Palmas de Gran Canaria	ulpgc.es	1	3	0	0	0
Universitat Oberta de Catalunya	uoc.edu	52	72	2	0	2
Universidad de Huelva	uhu.es	0	0	0	0	0
Universidad de Alcalá	uah.es	1	4	0	0	1
Universidad de Cantabria	unican.es	0	0	0	0	0
Universidad de Jaén	ujaen.es	0	0	0	0	0
Universidad de Extremadura	unex.es	0	0	0	0	0
Universidad de La Laguna	ull.es	2	0	0	0	0
Universidad Rey Juan Carlos	urjc.es	0	0	0	0	0
Universitat Rovira i Virgili	urv.cat	55	49	0	0	0
Universidad Miguel Hernández	umh.es	0	0	0	0	0
Universidad Pablo Olavide	upo.es	0	0	0	0	0
Universidad de Almería	ual.es	0	0	0	0	0
Universidad Politécnica de Cartagena	upct.es	1	0	0	0	0
Universidad Pública de Navarra	unavarra.es	0	0	0	0	0
Universidad de La Rioja	unirioja.es	17	0	0	0	0
Universitat de Lleida	udl.es	0	0	0	0	0

Universidad	Url	Tamaño semántico	Enlaces internos	Enlaces de terceros	Enlaces entrantes	Enlaces salientes
Universidad de Deusto	deusto.es	594	1.390	0	0	1
Universidad Pontificia de Comillas	upcomillas.es	0	0	0	0	0
Universitat Ramon Llull	url.edu	0	0	0	0	0
Universidad de Burgos	ubu.es	0	0	0	0	0
Universitat de Vic	uvic.cat	4	4	0	0	0
Universidad Europea de Madrid	uem.es	Sin datos				
Universidad de Mondragón	mondragon.edu	152	0	0	0	0
Universidad Católica San Antonio de Murcia	ucam.edu	220	183	0	0	0
Universidad CEU Cardenal Herrera	uchceu.es	Sin datos				
Universidad San Pablo CEU	uspceu.es	0	0	0	0	0
Universidad Pontificia de Salamanca	upsa.es	2	0	0	0	0
Universitat Internacional de Catalunya	uic.es	0	0	0	0	0
Universidad Internacional de Andalucía	unia.es	299	0	0	0	0
Universidad Nebrija	nebrija.com	Sin datos				
Universidad Internacional de La Rioja	unir.net	Sin datos				
Universidad Camilo José Cela	ucjc.edu	0	0	0	0	0
Universidad Alfonso X El Sabio	uax.es	0	0	0	0	0
Universidad Católica de Valencia	ucv.es	Sin datos				
Universidad Francisco de Vitoria	ufv.es	Sin datos				
Universidad Internacional Menéndez Pelayo	uimp.es	4	4	0	0	0
Universidad Europea Miguel de Cervantes	uemc.es	Sin datos				
Universitat Abat Oliba	uao.es	2.100	2.050	0	0	0
Universidad San Jorge	usj.es	0	0	0	0	0
Universidad Católica de Ávila	ucavila.es	Sin datos				
Universidad a Distancia de Madrid	udima.es	Sin datos				
Valencian International University	viu.es	Sin datos				

Las posiciones de las universidades responden a la posición obtenida en el ranking web, únicamente en tanto que fuente de la que se ha extraído la muestra de estudio

4. Conclusiones finales

Es esencial la existencia de tecnologías y servicios asequibles que asistan a los autores web durante el proceso de marcado. Esto permitiría automatizar ciertos aspectos de este proceso y su integración a gran escala en sitios web gestionados por CMS, para su posterior integración en *datasets*, así como su reutilización en otras plataformas y servicios de forma sencilla. En consecuencia, se mejoraría sensiblemente su difusión y recuperación en la Red, sin la necesidad de diseñar y establecer procesos costosos de marcado semántico manual, o inversión en tecnologías complementarias.

La efectividad del marcado automático aún tiene que ponerse a prueba con rigor. Funciona mejor con textos en inglés que en otros idiomas y hay que hacer varios ensayos de marcado, o seleccionar varios servicios, para obtener un marcado óptimo. Nuestra experiencia sugiere que en última instancia siempre debe recurrirse a una combinación del marcado manual y automático.

Aunque el marcado semántico de textos en un blog no ofrece una ventaja evidente o directa al editor de un web, su incorporación a los CMS más extendidos ayudará a hacer crecer un “excedente semántico” en la Web que sea aprovechable a partir de cierto volumen por otros agentes para

producir valor. “Los excedentes grandes son diferentes de los pequeños” (Shirky, 2010), y para alcanzar esa masa crítica es necesario reducir el coste de entrada a la larga cola de la web semántica a través de las herramientas más usadas. Dicho contenido semántico también es susceptible de integrarse con *datasets* externos más amplios y, por tanto, en el ecosistema *linked data*.

Estas prácticas facilitarían el desarrollo de estudios ciberométricos que usarían la información semántica introducida en el mercado para obtener nuevos indicadores (como los utilizados en este trabajo para conocer el tamaño y visibilidad web de universidades), y que se pueden expandir a niveles más específicos (especialmente atributos y clases). En ese sentido, *Sindice* ha demostrado ser una fuente útil para la cibermetría, aunque se precisan más estudios para comprobar su cobertura, posibles limitaciones y funcionalidades aplicadas al estudio del crecimiento, evolución y uso de contenido marcado en la Web.

5. Notas

1. También participan en este proyecto *Microsoft* (con su buscador *Bing*), *Yahoo!* y *Yandex*.
2. *SaaS* (*software as a service*): modelo de software en el que las aplicaciones y los datos se ubican en la “nube” y

cuyo acceso y uso se realizan generalmente a través de un navegador web.

6. Bibliografía

Balmin, Andrey; Hristidis, Vagelis; Papakonstantinou, Yanis (2004). "Objectrank: authority-based keyword search in databases". En: *Procs of the 30th intl conf on very large data bases*, v. 30, pp. 564-575.

<http://www.vldb.org/conf/2004/RS15P2.PDF>

Franz, Thomas; Schultz, Antje; Sizov, Sergej; Staab, Steffen (2009). "TripleRank: ranking semantic web data by tensor decomposition". En: Bernstein, Abraham *et al.* (ed.). *The semantic web. ISWC 2009*. Springer, v. 5823, pp. 213-228.

<http://data.semanticweb.org/pdfs/iswc/2009/paper279.pdf>
http://dx.doi.org/10.1007/978-3-642-04930-9_14

García-García, Alicia (2012). *Datos abiertos enlazados linked open data (LOD) en documentación científica*. Valencia: Universidad Politécnica de Valencia.

<http://riunet.upv.es/handle/10251/18272>

Jeh, Glen; Widom, Jennifer (2002) "SimRank: a measure of structural-context similarity". En: *Procs of the 8th ACM Sigkdd intl conf on knowledge discovery and data mining*. New York, pp. 538-543.

<http://ilpubs.stanford.edu:8090/508/1/2001-41.pdf>

<http://dx.doi.org/10.1145/775107.775126>

Khalili, Ali; Auer, Sören; Hladky, Daniel (2012). "The RDFa content editor: from wysiwyg to wysiwym". En: *Computer software and applications conf (Compsac), 2012 IEEE 36th Annual*, pp. 531-540.

http://svn.aksw.org/papers/2012/COMPSAC2012_RDFaCE/public.pdf

<http://dx.doi.org/10.1109/COMPSAC.2012.72>

Khare, Rohit; Çelik, Tantek (2006). "Microformats: a pragmatic path to the semantic web". En: *Procs of the 15th intl conf on world wide web*. ACM: New York, pp. 865-866.

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.307.7135&rep=rep1&type=pdf>

<http://dx.doi.org/10.1145/1135777.1135917>

Kleinberg, John M. (1999). "Authoritative sources in a hyperlinked environment". *Journal of the ACM*, v. 46, n. 5, pp. 604-632.

<http://www.cs.cornell.edu/home/kleinber/auth.pdf>

<http://dx.doi.org/10.1145/324133.324140>

Longqing, Shi; Qingfeng, Zhao (2011). "Data sources of we-

bometrics". En: *7th Intl conf on computational intelligence and security*, pp. 1312-1315.

<http://dx.doi.org/10.1109/CIS.2011.291>

Mendes, Pablo N.; Jakob, Max; García-Silva, Andrés; Bizer, Christian (2011). "DBpedia spotlight: shedding light on the web of documents". En: *Procs of the 7th Intl conf on semantic systems*, 1-8.

<http://goo.gl/JQ2DEs>

<http://dx.doi.org/10.1145/2063518.2063519>

Meymandpour, Rouzbeh; Davis, Josep G. (2013). "Ranking universities using linked open data". En: *LDOW2013*.

<http://events.linkedata.org/ldow2013/papers/ldow2013-paper-09.pdf>

Oren, Eyal; Delbru, Renaud; Catasta, Michele; Cyganiak, Richard; Tummarello, Giovanni (2008). "Sindice.com: a document-oriented lookup index for open linked data". En: *Intl journal of metadata, semantics and ontologies*, v. 3, n. 1, pp. 37-52.

<http://dx.doi.org/10.1504/IJMSO.2008.021204>

Pastor-Sánchez, Juan-Antonio (2011). *Tecnologías de la web semántica*. Barcelona: Editorial UOC, colección El profesional de la información, 1. ISBN: 978 84 9788 474 7

Pastor-Sánchez, Juan-Antonio. (2012). "Los CMS como pieza fundamental en el despliegue de la web semántica". *Anuario ThinkEPI*, v. 6, pp. 184-189.

Saorín, Tomás; Peset, Fernanda; Ferrer-Sapena, Antonia (2013). "Factores para la adopción de linked data e implantación de la web semántica en bibliotecas, archivos y museos". *Information research*, v. 18, n. 1.

<http://InformationR.net/ir/18-1/paper570.html>

Sellés-Carot, Alicia; Orduña-Malea, Enrique; Serrano-Cobos, Jorge (2013). "Estrategias y oportunidades tecnológicas en la generación de linked data en las bibliotecas". *Mi biblioteca*, pp. 54-59.

Shirky, Clay (2010). *Cognitive surplus: creativity and generosity in a connected age*. New York: Penguin Press. (edición en castellano: *El excedente cognitivo: creatividad y generosidad en la era conectada*. Barcelona: Ediciones Deusto, 2012. ISBN: 978 8423428632). ISBN: 978 0143119586

Stuart, David (2012). "FOAF within UK academic web space: a webometric analysis of the semantic web". En: Widén, Gunilla; Holmberg, Kim (ed.). "Social information research". *Emerald Group Publishing Ltd.*, v. 5, pp. 173-191. ISBN: 978 1 78052 832 8



2^{es} JORNADES VALENCIANES de DOCUMENTACIÓ

innovació i ocupabilitat #jvdoc13

17 – 18 octubre 2013
València

<http://jornades.cobdcv.es>



LA GESTIÓN DE CONTENIDOS COMO ACTIVIDAD ESTRATÉGICA EN EMPRESAS DE RADIODIFUSIÓN. ESTUDIO DE CASOS EN LA RADIO COMERCIAL ESPAÑOLA



Manuel Fernández-Sande, Dolores Rodríguez-Barba y Miriam Rodríguez-Pallares



Manuel Fernández-Sande, doctor en ciencias de la información por la *Univ. Complutense de Madrid (UCM)*, licenciado en periodismo (*UCM*) y en documentación (*Univ. Carlos III de Madrid*), es profesor en la *Fac. de Ciencias de la Información (UCM)*. Investiga sobre el mercado radiofónico europeo, historia de la radio, dirección estratégica de empresas de comunicación, y organización y gestión del conocimiento. Es autor de *Los orígenes de la radio en España I y II* (2006) y codirector de *Panorámica de la comunicación y de los medios en Brasil y España* (2013).

<http://orcid.org/0000-0002-0740-2630>

manuel.fernandez@ucm.es



Dolores Rodríguez-Barba es doctora en ciencias de la información. Trabaja en el *Departamento de Periodismo IV* de la *Universidad Complutense de Madrid* como profesora en el Grado y Master oficial de Periodismo, y como investigadora en el ámbito de la organización y gestión de empresas de comunicación. Ha trabajado en *TVE* y en las áreas de programación y dirección técnica en varias cadenas de radio. Es miembro del grupo de investigación *MediaCom UCM* y co-fundadora de *Ocendi*.

<http://orcid.org/0000-0001-8437-5536>

lolar@ucm.es



Miriam Rodríguez-Pallares es diplomada en biblioteconomía y documentación por la *Univ. de Salamanca*, máster en periodismo por la *UCM* y doctoranda en el *Depto. de Periodismo IV* de la *UCM*. Su trabajo fin de máster lleva por título: *La documentación informativa en los medios de comunicación. Perspectiva económica y funcional*. Desde 2007 ha trabajado en el área de documentación y gestión de contenidos de *El país*, en el *Área de Gobierno de Familia y Servicios Sociales* del Ayuntamiento de Madrid y en *Everis BPO*; como periodista, en la *Unión de Asociaciones de Estanqueros de España* y como colaboradora de la *Fundación Mapfre*.

<http://orcid.org/0000-0002-5486-0298>

mrpallares@pdi.ucm.es

*Universidad Complutense de Madrid
Facultad de Ciencias de la Información, Departamento de Periodismo IV
Avda. Complutense s/n. 28040 Madrid, España*

Resumen

Se presentan los resultados parciales del análisis comparativo realizado en tres de las principales cadenas de radiodifusión comercial españolas (*Cadena Ser*, *Onda Cero* y *Cope*) para identificar y analizar la estructura de los departamentos encargados de la gestión de contenidos así como el modelo y las herramientas de gestión. El estudio se presenta aplicando la metodología de casos y desde una perspectiva funcional, es decir, en relación con la actividad principal de cada una de las empresas, que fundamenta su modelo de negocio en la comercialización de contenidos informativos, culturales y de entretenimiento. En la segunda fase del análisis, se han realizado entrevistas en profundidad a los responsables de los cuadros altos de los organigramas de las empresas radiofónicas.

Palabras clave

Sistemas de gestión de contenidos, SGC, Empresas de radiodifusión, Documentación en medios, Documentación radiofónica, *Cadena Ser*, *Onda Cero*, *Cope*, *Enciclomedia*, *Dalet Plus*, *Paloo*.

Title: Content management as a strategic activity in broadcasting companies. Case studies about private radio networks in Spain

Abstract

We present partial results of a comparative analysis in three major commercial broadcasters in Spain -*Cadena Ser*, *Onda Cero* and *Cope*- to identify and analyze the structure of the departments responsible for content management, as well as the model and management tools used. The study applies a case methodology and takes a functional perspective, that is, a business model based on the core business of each company, i.e., informational, cultural, or entertainment content. For the second phase of the analysis, we conducted in-depth interviews with the heads of the executive team of each broadcasting company.

Keywords

Content management systems, CMS, Broadcasting companies, Media documentation, Broadcasting documentation, Spain, *Cadena Ser*, *Onda Cero*, *Cope*, *Enciclomedia*, *Dalet Plus*, *Paloo*.

Fernández-Sande, Manuel; Rodríguez-Barba, Dolores; Rodríguez-Pallares, Miriam (2013). "La gestión de contenidos como actividad estratégica en empresas de radiodifusión. Estudio de casos en la radio comercial española". *El profesional de la información*, septiembre-octubre, v. 22, n. 5, pp. 392-398.

<http://dx.doi.org/10.3145/epi.2013.sep.03>

1. Introducción

Las organizaciones empresariales llevan a cabo políticas estratégicas, con objetivos de superación y competitividad, que complementan los factores productivos propios de la economía industrial -capital, trabajo y recursos humanos- y se basan en la innovación tecnológica y organizativa como factores de éxito. Esto supone una constante vigilancia crítica de la actividad empresarial aun en épocas de estabilidad: "la cultura de lo intangible está dejando paso a una nueva manera de generar riqueza" (Carrillo-Durán; Nuño-Moral, 2010).

En este contexto la gestión de contenidos en la industria de la comunicación cobra un doble protagonismo (Rodríguez-Pallares, 2012):

- desde una perspectiva funcional, enfocada a la mejora de la distribución y la revalorización de su producción y, por lo tanto de su marca en el mercado, que busca informar, formar y entretener;
- desde una perspectiva estratégica, enfocada a la obtención de beneficios a largo plazo, que pueden repercutir en el aprendizaje organizacional.

La empresa radiofónica ha experimentado en los últimos años una importante evolución en paralelo a los procesos de transformación de los mercados informativos. Podemos definirla como una "organización social que, mediante la gestión y administración de elementos humanos, técnicos, tangibles, intangibles y recursos financieros, produce, distribuye -por medio de diferentes soportes y canales- y comercializa contenidos sonoros y otros recursos multimedia que pretenden satisfacer las necesidades informativas y de entretenimiento de sus públicos objetivos, con el fin de obtener el máximo beneficio posible". (Fernández-Sande; Peinado-Miguel, 2012).

Estas empresas han incorporado la gestión de contenidos sonoros y multimedia como un activo fundamental de sus sistemas productivos y de comercialización. La gestión de contenidos es un proceso complejo que requiere políticas de gestión y un software especializado, que denominamos sistemas gestores de contenido (*content management sys-*

tems o CMS) (Cox; Yeadon, 2002). Implica un conjunto de destrezas y actuaciones profesionales que permiten un tratamiento sistemático y global de los diferentes tipos de información (García-Morales; Bustelo-Ruesta, 2001). En el concepto de gestión de contenidos confluyen por tanto componentes tecnológicos, estrategia empresarial y prácticas profesionales. El objetivo final de todo proceso de gestión de contenidos es la creación, gestión, presentación y difusión de información (Pérez-Montoro, 2005) con la mayor eficiencia y eficacia posible. Para lograrlo se requiere un único centro documental que disponga de una plataforma tecnológica capaz de ofrecer servicio a todos los editores y productores que trabajen en red en la empresa radiofónica (García-Lastra, 2012).

“ En el concepto de gestión de contenidos confluyen componentes tecnológicos, estrategia empresarial y prácticas profesionales ”

La bibliografía científica ha tratado de establecer una diferenciación conceptual entre gestión del conocimiento y gestión de contenidos. Delimitar ambos procesos resulta necesario desde un punto de vista teórico pero también es importante para el correcto funcionamiento organizacional. La interacción permanente entre la generación de conocimiento y la gestión de contenidos dificulta en la práctica esta delimitación. La primera se centra en el diseño de estrategias, procesos y estructuras que permitan a la organización incorporar y aprovechar el conocimiento de todos sus miembros (Fernández-Marcial, 2006). La mayoría de autores coinciden en que se encuentra en un plano superior a la gestión de contenidos aunque ambos procesos son totalmente interdependientes. (Nonaka; Takeuchi, 1995; Newman, 1997; Pearson, 1999; Davenport; Prusak, 2001; Bueno-Campos, 2004)

2. Objetivos y metodología de la investigación

El objetivo general es conocer los modelos de gestión de contenidos desde una perspectiva funcional de tres de las

principales cadenas de radiodifusión privadas de España: *Ser*, *Onda Cero* y *Cope*. En concreto, se pretende:

- identificar y analizar la estructura de los departamentos encargados de la gestión de contenidos;
- identificar la visibilidad de la gestión de contenidos en el seno de la organización;
- identificar el modelo y las herramientas de gestión de contenidos.

La metodología empleada es la del estudio de casos en tres fases:

- análisis exhaustivo de la bibliografía científica y profesional;
- recogida de documentación propia de cada una de las empresas;
- entrevistas en profundidad a los responsables de los cuadros directivos de cada medio, realizadas entre julio de 2010 y abril de 2013 (la relación de entrevistados se muestra al final del artículo).

3. Modelos de gestión de contenidos en *Cadena Ser*, *Onda Cero* y *Cope*

3.1. Identificación del departamento responsable

No existen dudas sobre la importancia de la gestión de contenidos en el diseño general de estrategias empresariales, especialmente en las empresas de comunicación, que deben dar prioridad a la organización de la producción y la explotación de dichos contenidos atendiendo a criterios documentales con un fin informativo y de difusión y no sólo probatorio. Sin embargo, en los últimos años, a consecuencia de los procesos de convergencia tecnológica se producen situaciones dispares en el conjunto de los medios, motivadas por estrategias no muy definidas: creación de líneas de negocio cuyos objetivos se cumplen en el primer año y que tras superar el período de viabilidad son suprimidas por falta de rentabilidad (Peinado-Miguel; Rodríguez-Barba, 2012), simples logros a exhibir, o actividades residuales.

En los casos estudiados, *Cadena Ser* es la única empresa radiofónica con un departamento de documentación corporativa que desempeña todas las funciones propias de la gestión de contenidos con fin informativo, incluyendo producción y publicación. *Cope* por su parte dispone de un departamento de documentación orientado hacia la preservación de contenidos. Por último *Onda Cero* carece en su estructura organizativa de un servicio especializado a pesar de contar con un CMS que mantiene los contenidos durante 1 año en su división analógica (tabla 1).

Tabla 1: Modelos de gestión de contenidos

Empresa	CMS	Uso	Alcance
<i>Ser</i>	<i>Enciclomedia</i>	Descripción y gestión Repositorio Plataforma de publicación	<i>Prisa Radio</i>
<i>Onda Cero</i>	<i>Dalet</i>	Repositorio temporal (1 año) Acceso a fondos de A3TV	<i>Onda Cero</i> <i>Europa FM</i>
<i>Cope</i>	<i>Knosys</i>	Descripción y gestión Repositorios	<i>Cope</i> <i>Cadena 100</i> <i>Rock FM</i>

3.2. Posicionamiento organizativo y evolución de la gestión de contenidos

El proceso de implantación del modelo de gestión de contenidos en un medio, las razones que han llevado a tal decisión y la evolución posterior proporciona las claves para conocer su funcionalidad y valoración empresarial, su peso en la organización.

Es imposible impermeabilizar los modelos de gestión de contenidos vigentes en las emisoras analizadas de la influencia que, tanto a nivel funcional como organizativo, ejercen sobre ellas sus empresas matrices. El modelo de gestión de contenidos de la *Cadena Ser* se entiende teniendo en cuenta que es la cabecera de la red internacional que conforma *Prisa Radio*; el de *Onda Cero* en el marco de *Atresmedia*. Los modelos organizativos divisionales ya sean geográficos o por productos (De-Mateo; Bergés; Sabater, 2009) determinan el diseño y el posicionamiento organizacional de la gestión de contenidos en la empresa.

“*Cadena Ser* es la única empresa con un departamento de documentación que realiza todas las funciones propias de la gestión de contenidos con fin informativo”

El *Departamento de Documentación* de la *Cadena Ser* fue creado en 1988, tras la inclusión de la emisora en el accionariado del *Grupo Prisa*, con el fin de salvaguardar la documentación sonora con la que contaba la marca radiofónica más longeva de España. Su labor de archivo no fue cuestión baladí por la novedad que éste suponía en el sector privado. El cambio de paradigma en cuanto a gestión de contenidos llegó a *Prisa Radio* (entonces *Unión Radio*) en 2004. El objetivo era pasar de un modelo local a un modelo integracionista que vinculase el total del valor informativo generado en todas las sedes de *Prisa Radio* en un mismo repositorio de datos que nutriese, a su vez, a todo el conjunto de emisoras del grupo a ambos lados del Atlántico. Se optó por un programa gestor de contenidos propio, denominado *Enciclomedia*, que empezó a desarrollar en 2001 la empresa *Atos Consulting*. Paralelamente se fue diseñando y organizando el *Departamento de Documentación* con un concepto más funcional, hasta llegar a denominarse primero *Gestión de Contenidos* y en la actualidad *Documentación Corporativa*.

En 2012 *Documentación Corporativa* amplió sus competencias para asumir procesos editoriales secundarios, es decir, para convertirse en publicador de contenidos y no sólo en descriptor y gestor. Para ello, se reforzó la vinculación del programa de gestión de contenidos con las líneas de negocio online pese a que, como en todos los casos analizados, siguen siendo líneas de actuación independientes orgánica y funcionalmente. *Enciclomedia* dispone de pasarelas automáticas que la convierten en una plataforma de lanzamiento directo hacia publicaciones webs, telefonía móvil, apps, etc. Estas funciones representan un importante salto cualitativo respecto a las bases de datos “estáticas”.

La aportación documental a la programación de la *Cadena Ser* es la más visible de los 3 casos analizados:

- por el apoyo al trabajo de las redacciones en sus labores retrospectivas y comparativas;
- por la valoración del trabajo documental en el medio, hecho que distingue este caso de todos los demás y denota un reconocimiento empresarial a esta labor que repercute directamente en el producto y en la organización, particularmente en la plantilla.

Atresmedia Radio justifica la ausencia de un departamento propio en *Onda Cero* alegando la posibilidad de acceso al repositorio *Atresmedia TV*, sin plantearse un repositorio conjunto

El caso de *Onda Cero* es diferente. No existe un departamento de gestión de contenidos que colabore con la producción informativa y en el proceso de reutilización y rentabilización de los contenidos generados por el grupo *Atresmedia Radio*. *Onda Cero* cuenta tan sólo con un archivo físico donde se almacena el continuo de su programación *on air* (en vivo, no enlatada), con códigos descriptivos temporales, lo que significa que la búsqueda y reutilización de sus contenidos propios depende de un responsable del archivo, los contenidos no están vinculados y el usuario debe conocer la franja horaria en la que ha sido emitido su objeto de búsqueda. *Atresmedia Radio* justifica la ausencia de un departamento propio a nivel funcional alegando que *Atresmedia TV* da acceso a *Onda Cero* al repositorio de datos documentales de la televisión -gestionado externamente por *Accenture-*, solventando sus necesidades informativas más inmediatas. También se apunta el ahorro de costes para el grupo mediático, aun cuando parece que la rentabilización real vendría de la mano de un modelo de gestión y archivo común para todos los contenidos.

Cope por su parte cuenta con un departamento de gestión de contenidos, *Documentación*, creado en 1989 tras reconfiguraciones empresariales internas. Se creó con el propósito de no perder horas de grabación de incalculable valor, parte de las cuales salieron por las puertas de *Radio Popular* sin más explicación que la marcha de sus locutores estrella y un antiguo modelo de archivo de fichas arcaico, poco exhaustivo y de escasa supervisión. Hasta ese momento, alentada por su naturaleza de "radio de estrellas", la gestión y custodia de los contenidos de audio en la *Cope* era competencia de cada programa. Las copias legales estaban sujetas a una normativa hoy ya caduca que exigía la conservación durante sólo un mes de la emisión en antena en soporte *cassette*, frente a los tres meses que obliga la ley actual. Desde su creación, el *Departamento de Documentación* ha sido objeto de pocos cambios y cuenta con recursos limitados tanto informáticos como humanos (actualmente dos personas).

En los casos analizados se han encontrado tres ejemplos diferentes de posicionamiento organizativo del área de gestión de contenidos:

- en la *Cadena Ser* el *Departamento de Documentación Corporativa* depende de la presidencia de *Prisa Radio*;
- en la *Cope* depende de la dirección de programas;
- en *Onda Cero* no existe unidad departamental de gestión de contenidos.

En los tres casos la ubicación del departamento de documentación en la estructura de la empresa permite deducir el nivel de reconocimiento y el potencial, tanto organizativo como funcional, del que dispone la gestión de contenidos.

3.3. Modelos y aplicaciones de gestión de contenidos

Los sistemas de gestión de contenidos pueden dividirse en WCM (*web content management*) y ECM (*enterprise content management*). Ambos abarcan tareas de creación, gestión, distribución, y publicación de la información en un modelo centralizado (Robertson, 2003). Cada organización

Tabla 2: Departamentos y herramientas de gestión de contenidos *Cadena Ser*, *Onda Cero* y *Cope*.

			
Servicio documental y actual gestor de contenidos			
Fecha de creación del departamento de gestión de contenidos	1988	No existe	1989
CMS/MAM	Sí, permanente (<i>Enciclomedia</i>)	Sí, conservación temporal y con prestaciones limitadas (<i>Dalet Plus</i>)	Sí, permanente (<i>Knosys</i>) + WCM independiente (<i>Paloo</i>)
Naturaleza del programa actual (adquisición/propia)	Propia	Adquisición	Adquisición
Organigrama departamental			
Dependencia del departamento	Dirección General de <i>Prisa Radio</i>	Fonoteca depende del Director de Emisiones y Subdirector de Programas	Dirección de Programas
Cabecera del servicio de gestión de contenidos	Sí	No	Sí
Ubicación del departamento	Sede principal de <i>Prisa Radio</i> . Gran Vía, 32. 28013 Madrid	No hay departamento	Sede principal <i>Cope</i> . Alfonso XI, 4. 28014 Madrid

adapta el CMS a sus necesidades particulares. Así, en las empresas de comunicación se usa comúnmente la nomenclatura MAM (*media asset management*) para identificar un CMS que, manteniendo las mismas bases conceptuales, no se basa en un modelo estándar, independientemente de si se trata de un gestor web o no. Lo más habitual es que se adapten productos prefabricados a las exigencias del contexto de aplicación -*Onda Cero* con el MAM de *Dalet*-, o que sea la empresa quien diseñe su propia aplicación -*Cadena Ser* con *Enciclomedia*-.

Radio Popular cuenta con un departamento de gestión de contenidos con limitaciones funcionales que pierde peso organizativo frente a la división digital

El modelo de gestión de contenidos de *Prisa Radio* está condicionado fundamentalmente por su naturaleza internacional. El grupo radiofónico cuenta con dos grandes sistemas de actividad en su modelo:

- Local, basado en la aplicación *Redacta*: cada emisora cuenta con una aplicación identificada como un entorno de edición y archivo local para los periodistas, con campos de descripción reducidos y para uso interno. Los sistemas locales de las diferentes emisoras están interconectados entre sí para facilitar el intercambio de archivos.
- En cadena, basado en la aplicación *Enciclomedia*: cuando el contenido manejado en una emisora local concreta traspasa los intereses locales para alcanzar trascendencia para toda la marca, el contenido es analizado y seleccionado para pasar a formar parte del CMS.

La internacionalidad de su modelo de gestión de contenidos exige una descripción de perfiles de usuario muy definida con el fin de evitar duplicaciones y accesos dobles a la manipulación de contenidos. Por esta razón, se ha diseñado un modelo jerarquizado cuya cabecera está en Madrid y de la cual dependen orgánicamente los demás puntos clave con permiso para nutrir el sistema en cadena: España, Colombia, Chile, Argentina y México.

Por su parte, el contexto de trabajo de *Onda Cero Radio* se divide en tres grandes áreas:

- entorno de producción: cuyo soporte informático es el gestor de contenidos *Dalet*, donde se custodian los continuos durante un período máximo de un año, con una ficha descriptiva generada por los propios periodistas; pasado este tiempo, pasan al departamento de *Fonoteca* en formato físico, que cuenta con una base de datos simple de acceso únicamente para el departamento;
- entorno de emisión: cuyo soporte informático es a día de hoy *Viva NT*;
- entorno de intercambio de contenidos entre las diferentes emisoras de *Onda Cero* por FTP.

Resulta desconcertante que *Onda Cero* haya hecho una importante inversión en la adquisición de un MAM, pero no rentabilice las prestaciones de un gestor de contenidos como fuente, sino tan sólo como salvaguarda temporal de

su emisión *on air*. Evidentemente, la custodia de sus propios fondos garantiza la posibilidad de recuperarlos pese al consiguiente retraso que implica su archivo físico, pero en ningún caso multiplica su valor ni favorece la rentabilización del trabajo realizado. Podría decirse que se mantiene un modelo de custodia permanente de copias de seguridad. La ausencia de una actividad de descripción y vinculación de contenidos exhaustiva evita la diferenciación de perfiles de usuario, incluso en la *Fonoteca*, donde los dos responsables, aun teniendo vinculación jerárquica, intercalan privilegios de actuación.

Por último, la actividad documental de *Cope* en el área no digital basa su objetivo primario en la custodia de los continuos y todo corte de audio preseleccionado por los redactores y editores de programas e informativos y extraído de su propio sistema de emisión. En ocasiones puntuales funciona como fuente y genera contenidos documentales coadyuvantes para la producción informativa. El modelo de gestión de contenidos se sustenta en el gestor documental *Knosys* y las prestaciones del paquete *Office* de *Microsoft*. Una plantilla de tan sólo dos profesionales vinculados jerárquicamente exime la necesidad de definir perfiles de usuario, compartiendo ambos los privilegios de actuación.

Las empresas radiofónicas comerciales en España carecen de un planteamiento estratégico global que les permita integrar el potencial de la gestión de contenidos en la actividad corporativa

Actualmente se está estudiando la posibilidad de unificar todo el proceso de producción radiofónica en un único CMS, desde los sistemas de producción, emisiones y archivo, al multimedia. De llevarse a cabo este proyecto, podría posicionar al *Grupo Radio Popular* en la vanguardia en gestión de contenidos en el panorama nacional. El objetivo es la vinculación total de los contenidos de la empresa en un mismo sistema integrado sustentado en una compleja arquitectura informática basada en capas semánticas inteligentes. El salto cualitativo sería revolucionario para *Radio Popular* (Luis Martín-Pérez, 9 de abril de 2013).

4. Conclusiones

Las tres empresas radiofónicas analizadas cuentan con modelos de gestión de contenidos que presentan grandes diferencias, tanto en sus aspectos organizativos como funcionales. En ningún caso se identifica un planteamiento estratégico global que permita integrar y aprovechar el potencial de la gestión de contenidos en la actividad corporativa. Se mantiene todavía muy arraigada la concepción del archivo sonoro estático y se detecta una infraexplotación de los contenidos sonoros y multimedia a pesar del importante desarrollo en lo que se refiere a herramientas tecnológicas y prácticas profesionales registradas en los últimos años.

De los tres casos analizados *Prisa Radio* representa el modelo de gestión de contenidos más evolucionado. Dispone de un departamento de documentación consolidado en su

estructura organizativa, capaz de desarrollar procesos de producción propia y difusión de información aportando de esa forma un valor añadido más visible a la oferta radiofónica. También se diferencia por haber desarrollado un CMS de creación propia adaptado a sus características y objetivos, *Enciclomedia*, que posibilita un modelo de trabajo en red internacional.

Prisa Radio ha consolidado la gestión de contenidos como parte importante de su cultura y estrategia empresarial, a pesar de la inestabilidad que provocan los cambios organizativos que se han producido y que se puedan producir en un futuro, todo indica que en los próximos años se intensificará la rentabilización de los contenidos y se reforzará el área de *Documentación Corporativa*.

Se detecta una infraexplotación de los contenidos sonoros y multimedia

Las otras dos empresas analizadas, han optado o bien por la externalización de los procesos de gestión, caso de *Atresmedia* y *Accenture*; o por la adquisición de aplicaciones informáticas externas que requieren una posterior adaptación, caso de *Cope* con *Knosys* y *Paloo*.

En *Onda Cero* no parece que la empresa matriz decida a corto plazo crear un departamento de gestión de contenidos propio con capacidad de generar valor añadido. Por su parte *Cope* se plantea mejoras cualitativas en su modelo de gestión de contenidos, pero este posible desarrollo tecnológico, funcional y estratégico parece condicionado por la evolución económica del grupo y su capacidad de generar nuevas vías de financiación.

Por último hay que destacar que los perfiles de usuario entre los gestores de contenido son más rígidos cuanto más grande y jerarquizada es la red de trabajo, y más polivalentes cuanto más pequeña y menos definida es la misma.

Entrevistas

Ángeles Afuera (abril 2012)

Jefa del *Departamento de Documentación Corporativa* de la *Cadena Ser*.

Miguel-Ángel Álbez-Sánchez (junio 2012)

Responsable de la *Fonoteca y Archivo Sonoro* de *Onda Cero*.

Francisco Castaño-Igual (junio 2012)

Jefe del *Departamento de Informática y Sistemas* de *Onda Cero*.

Ester Turu-Llimós (junio 2012)

Subdirectora de *Informativos* de *Onda Cero*.

Alicia Sánchez-Gómez (julio 2010)

Documentalista de la *Cadena Ser* Madrid.

Luis Rodríguez-Pi (agosto 2012)

Director de *Antena* de la *Cadena Ser* Madrid.

Alicia Ayala (agosto 2012)

Documentalista de la *Cadena Cope*.

Luís Martín-Pérez (abril 2013)

Jefe del área *Internet* de la *Cadena Cope*.

Maribel Sánchez (julio 2012)

Antigua responsable y creadora del *Departamento de Documentación* de la *Cadena Cope*.

Mónica Muñoz-García (noviembre 2012)

Coordinadora de *ondacero.es* y *europafm.com*

Bibliografía

Bueno-Campos, Eduardo (2004). "Fundamentos epistemológicos de dirección del conocimiento organizativo: desarrollo, medición y gestión de intangibles". *Economía industrial*, n. 357, pp. 13-26.

<http://goo.gl/U5Vwid>

Carrillo-Durán, María-Victoria; Nuño-Moral, María-Victoria (2010). "La documentación en la evaluación y gestión de la imagen corporativa". *El profesional de la información*, v. 19, n. 2, pp. 123-132.

<http://dx.doi.org/10.3145/epi.2010.mar.02>

Cox, Andrew; Yeadon, Jane (2002). "Estrategias prácticas en gestión de contenidos". *El profesional de la información*, v. 11, n. 6, pp. 466-474.

<http://www.elprofesionalde lainformacion.com/contenidos/2002/noviembre/7.pdf>

<http://dx.doi.org/10.1076/epri.11.6.466.15462>

Davenport, Thomas H.; Prusak, Laurence (2001). *Conocimiento en acción. Cómo las organizaciones manejan lo que saben*. Madrid, Pearson Educación. ISBN: 978 9879460290

De-Mateo, Rosario; Bergés, Laura; Sabater, Marta (2009). *Gestión de empresas de comunicación*. Sevilla: Comunicación Social. ISBN: 978 8496082908

Fernández-Marcial, Viviana (2006). "Gestión del conocimiento versus gestión de la información". *Investigación bibliotecológica*, v. 20, n. 41, pp. 44-62.

<http://www.ejournal.unam.mx/ibi/vol20-41/IBI002004102.pdf>

Fernández-Sande, Manuel; Peinado-Miguel, Fernando (2012). "La empresa radiofónica actual". En: Gallego-Pérez, Juan-Ignacio; García-Leiva, Trinidad. *Sintonizando el futuro: radio y producción sonora en el siglo XXI*. Madrid: Instituto RTVE. pp. 29-61. ISBN: 978 8488788856

García-Lastra, José-María (2012). "Del modelo productivo de la era analógica al de la radio multicanal". En: Gallego-Pérez, Juan-Ignacio; García-Leiva, Trinidad. *Sintonizando el futuro: radio y producción sonora en el siglo XXI*. Madrid: Instituto RTVE. pp. 167-188. ISBN: 9788488788856

García-Morales, Elisa; Bustelo-Ruesta, Carlota (2001). "Tendencias en la gestión de la información, la documentación y el conocimiento en las organizaciones". *El profesional de la información*, v. 10, n. 12, pp. 4-7.

<http://eprints.rclis.org/10040>

<http://dx.doi.org/10.1076/epri.10.12.4.6540>

Newman, Victor (1997). "Redefining knowledge management to deliver competitive advantage". *Journal of knowledge management*, v. 1, n. 2, pp. 123-128.

<http://dx.doi.org/10.1108/EUM000000004587>

Nonaka, Ikujiro; Takeuchi, Hirotaka (1995). *The knowled-*

ge creating company. Oxford University Press. ISBN: 978 0195092691

Pearson, Thomas A. (1999). "Measurement and the knowledge revolution". *Quality progress*, v. 32, n. 9, pp. 31-37.

Peinado-Miguel, Fernando; Rodríguez-Barba, Dolores (2011). "Nuevos modelos de negocio para la radio". En: Ortiz-Sobrino, Miguel-Ángel; López-Vidales, Nereida. *Radio 3.0 una nueva radio para una nueva era: la democratización de los contenidos*. Madrid: Fragua. pp. 141-162. ISBN: 978 8470744365

Pérez-Montoro, Mario (2005). "Sistemas de gestión de contenidos en la gestión del conocimiento". *BiD: textos univer-*

sitaris de biblioteconomia i documentació, n. 14. <http://www.ub.edu/bid/14monto2.htm>

Robertson, James (2003). "So, what is a content management system". *Step two designs*, June. http://www.steptwo.com.au/papers/kmc_what/index.html

Rodríguez-Pallares, Miriam (2012). "El capital intangible como clave estratégica en la empresa radiofónica española: el caso de la cadena Ser". En: Fernández-Sande, Manuel; Adami, Antonio (coord.). *Panorama da comunicação e dos meios no Brasil e Espanha / Panorámica de la comunicación y de los medios en Brasil y España*. São Paulo (Brasil): Intercom, pp. 473-495. ISBN: 9788582080153 <http://www.intercom.org.br/e-book/panorama.pdf>



Colección de libros de bolsillo El profesional de la información (Editorial UOC)

Más información:

<http://www.elprofesionalde lainformacion.com/libros.html>



GESTIÓN DE LOS FONDOS DOCUMENTALES EN RADIO NACIONAL DE ESPAÑA



Carmen Marta-Lazo y Miguel-Ángel Ortiz-Sobrino



Carmen Marta-Lazo es profesora titular de comunicación audiovisual y publicidad de la *Universidad de Zaragoza*. Doctora en ciencias de la información por la *Universidad Complutense de Madrid* y master de *Radio Nacional de España*. Es directora de *Radio.Unizar* e investigadora principal del *Grupo de Investigación en Comunicación e Información Digital (Gicid)*. Ha publicado sobre los géneros audiovisuales y su conversión hacia los cibergéneros fruto de las nuevas aplicaciones digitales.
<http://orcid.org/0000-0002-0004-1094>

Universidad de Zaragoza, Facultad de Filosofía y Letras
Pedro Cerbuna, 12. 50009 Zaragoza, España
cmarta@unizar.es



Miguel-Ángel Ortiz-Sobrino es profesor contratado doctor de la *Universidad Complutense de Madrid*. Doctor en ciencias de la información por la misma universidad. Entre 2000 y 2009 ha dirigido el *Instituto Oficial de Radio y Televisión*. También ha sido director adjunto de *Radio Televisión Madrid* y jefe de programas de *Onda Madrid*. De 1989 a 2009 ha dirigido el master de radio *Universidad Complutense/RNE*. Es profesor del área de Empresa Informativa en la *Universidad Complutense de Madrid* y delegado para *Medialab/Inforadio* de la *Facultad de Ciencias de la Información* de esa universidad.
<http://orcid.org/0000-0001-8129-3244>

Universidad Complutense de Madrid, Facultad de Ciencias de la Información
Av. Complutense, s/n. 28040 Madrid, España
maortiz@ucm.es

Resumen

En 2011, *Radio Nacional de España* implementó un nuevo sistema para la gestión de contenidos de sus fondos documentales denominado ARCA, que facilita a los responsables de los programas radiofónicos acceso directo a los contenidos documentales y sitúa a los documentalistas en el centro del proceso productivo. Se analiza la incidencia del sistema ARCA en las rutinas profesionales de la producción de programas. Se recoge el tipo de contenidos del fondo documental que pueden ser utilizados para mejorar la calidad de la programación, tanto en la emisión convencional como en las emisiones paralelas de valor añadido en la web radiofónica. Por último, se examina el nuevo sistema teniendo en cuenta la función de la radio pública en el mantenimiento de la memoria sonora y documental del Estado español y su organización territorial autonómica.

Palabras clave

Fondos documentales, Bases de datos, Documentación radiofónica, Documentación periodística, Perfiles profesionales, Documentalistas, Radio, Sistemas de gestión de contenidos, Archivos, ARCA, RNE.

Title: Management of documentary archives at *Radio Nacional de España*

Abstract

In 2011, Spain's public radio network, *Radio Nacional de España*, introduced a new content management system for its documentary archive, called ARCA. This new system allows those in charge of radio programs direct access to documentary content, and places documentalists at the center of the production process. In this article, the impact of the ARCA system on the professional routines of program production is analyzed. The article describes the typology of documentary archive content that can be used to improve programming quality, both in conventional broadcasting and in parallel broadcasting of added value on the radio web. Finally, the new system is analyzed, keeping in mind public radio's role in maintaining the documentary and sound archive of the Spanish State and its autonomic territorial organization.

Keywords

Documentary archive, Databases, Broadcasting documentation, Journalistic documentation, Professional profiles, Documents, Archive, Content management systems, Radio, ARCA, RNE.

Marta-Lazo, Carmen; Ortiz-Sobrino, Miguel-Ángel (2013). "Gestión de los fondos documentales en *Radio Nacional de España*". *El profesional de la información*, septiembre-octubre, v. 22, n. 5, pp. 399-403.

<http://dx.doi.org/10.3145/epi.2013.sep.04>

Artículo recibido el 30-04-2013
Aceptación definitiva: 26-07-2013

1. Introducción: la documentación audiovisual en la era digital

La documentación audiovisual es un área de conocimiento que en las últimas décadas ha evolucionado de forma sustancial como consecuencia de los avances tecnológicos, de la convergencia mediática y de un mayor marco conceptual. A partir del reconocimiento científico de la documentación en el ámbito de las ciencias de la información, se exploraron sus fundamentos teóricos y se apostó por su ramificación atendiendo a los diversos tipos de registros, según el medio de comunicación. Entre ellos, la radio y sus nuevos desarrollos propiciados por la aparición de las nuevas plataformas digitales (Nuño-Moral, 2007; Torregrosa, 2008; García-Jiménez, 2002).

Las herramientas tecnológicas para el tratamiento de la información aparecidas en la era digital están favoreciendo una mejor optimización de los recursos documentales, independientemente de los soportes concretos en los que están albergados (Torregrosa, 2008, p. 157; Caldera-Serrano, 2008; Caridad et al., 2011). En este sentido, la universalización de las TIC ha mejorado la organización y automatización de la documentación, permitiendo una mayor participación de los actores que intervienen en la gestión de los contenidos. Ahora, con la digitalización de los medios, el radiofonista realiza parte del trabajo documental a partir de la consulta directa de la base de datos del medio o mediante la captura de la información en internet (Micó-Sanz; Masip-Masip; García-Avilés, 2009). Este *modus operandi* da lugar a una transformación del rol del documentalista como intermediario de la información y le acerca a un terreno profesional en el que se le reclama un papel más activo para proponer temas informativos de interés y hacer más accesibles los recursos documentales adecuados para el trabajo informativo (Rubio-Lacoba, 2007). Además de una mayor accesibilidad, la digitalización de los fondos documentales está permitiendo la puesta en valor de la capacidad de recuperar contenidos y contextualizar la información en esta convergencia mediática, cada vez más compleja y multimodal (Area; Guarro, 2012).

2. Objeto y metodología de la investigación

Este artículo es el resultado de una investigación cuyo objetivo ha sido el análisis y revisión de las prestaciones del sistema ARCA (*Archivo y conservación de contenidos audiovisuales*) para la gestión de los contenidos documentales de *Radio Nacional de España (RNE)*, implantado a partir de julio de 2011. La investigación se ha realizado con tres perspectivas:

- tipología documental a la que da acceso este nuevo sistema;
- en qué medida sus prestaciones contribuyen a mantener actualizado el repositorio de la memoria sonora de España, como tradicionalmente lo había venido haciendo el Archivo Sonoro de la emisora pública (García-Jiménez, 2002);
- incidencia en la polivalencia de funciones de los productores de programas y en el desplazamiento de los documentalistas hacia labores más especializadas o de gestión tecnológica de los contenidos de los fondos documentales.

En cuanto a la metodología empleada, se ha partido de la lectura de bibliografía científica para contextualizar el objeto de estudio y, a partir de ahí, se ha procedido al análisis de contenidos y funciones del sistema ARCA. Por último, se han realizado entrevistas en profundidad a los principales responsables del Centro de Documentación de RNE, así como a los responsables de los Servicios Informativos, a la Dirección de Contenidos, a la Dirección de Programas y a la Dirección del área de Tecnología, cuyos departamentos son los principales usuarios de esta plataforma. Se ha optado por la entrevista en profundidad, porque permite contextualizar y obtener una visión más cualitativa de las respuestas (Sierra-Bravo, 2008, p. 352). El contraste de las opiniones de los entrevistados ha permitido revisar la utilización de los fondos documentales y las prestaciones de la plataforma ARCA, obteniéndose una panorámica de 360º que abarca el punto de vista de todos los usuarios de la plataforma.

Esta investigación ha sido realizada por el *Grupo de Investigación en Comunicación e Información Digital (Gicid)* de la *Universidad de Zaragoza* y el *Grupo Mediacom-UCM* de la *Universidad Complutense de Madrid*, con el apoyo de la *Asociación Española de Radios Universitarias*.

3. Análisis de resultados

Una vez elaborado el trabajo de campo, se han extraído resultados que ponen de manifiesto las prestaciones del sistema ARCA, su incidencia en las competencias de los diferentes profesionales de la radio y su valor como herramienta para la conservación de la *memoria sonora* de España.

3.1. ARCA: un sistema de gestión de contenidos para la convergencia digital

Desde julio de 2011, *Radio Nacional de España* comparte ARCA con los centros de producción de la *Corporación Radiotelevisión Española (RTVE)*. Los fondos documentales se estructuran teniendo en cuenta su tipo –texto, foto, audio y video–, y están ubicados en 13 bases de datos:

- *Agencias* incluye 3.950.000 registros con teletipos seleccionados de las agencias de noticias *EFE*, *Reuters* y *France Press*, que se cargan online a lo largo del día, de forma automática.
- *Hemeroteca* contiene 2.850.000 documentos.
- *Prensa*: 1.750.000 informaciones.
- *Fotos*: 250.000 fotografías.
- *Biblioteca* referencia 50.000 volúmenes: anuarios, enciclopedias, diccionarios, monografías, publicaciones especiales y libros electrónicos, que se encuentran en los centros de producción de RTVE: *Televisión Española (TVE)* en Barcelona, *Instituto RTVE* (Madrid), *RNE* (Prado del Rey, Madrid) y *TVE Informativos* (Torrespaña, Madrid).
- *Música clásica*: 390.000 registros.
- *Música ligera*: 1.175.000 grabaciones de varios géneros: pop, rock, jazz...
- *Música tradicional*. Manifestaciones del folclore español e internacional.
- *Palabra* es la que tiene un mayor valor histórico, tanto por su variedad, por su antigüedad y, en algunos casos, por el carácter único de algunos de los 585.000 documentos sonoros que contiene (Ariza-Chicharro, 2004). A estos do-

cumentos se suma el registro de las emisiones nacionales de los canales de la cadena pública desde 2001.

- *Informativos de Televisión Española (TVE)*: 1.575.000 grabaciones sobre la actualidad.
- *Programas* no informativos de la televisión pública estatal, con 375.000 registros.
- *Producción de programas de TVE*.
- *NO-DO*.

3.2. Dos niveles de usuarios del sistema

En *RNE* existen dos tipos de permisos para acceder al sistema *ARCA*. En el caso de los profesionales de realización y producción de programas, el permiso permite realizar cualquier tipo de búsqueda, entrar en todas sus bases de datos, visualizar o escuchar los documentos y descargarlos para su utilización en la producción de programas. Los documentalistas, además, pueden dar de alta nuevos documentos en las bases de datos, catalogar y analizar su contenido, solicitar su digitalización y modificar los metadatos. De igual manera, pueden gestionar los lenguajes documentales utilizados por el sistema, arbitrando tesauros y léxicos para validar nuevos términos.

3.3. Más polivalencia de los realizadores de programas

Prácticamente todos los profesionales de *Radio Nacional de España* que intervienen en la programación –guionistas, informadores, técnicos, realizadores y productores– tienen capacidad de consulta y selección de recursos de alguna de las bases del sistema *ARCA* para incorporarlos a sus programas. Además, en ocasiones proponen a los documentalistas criterios para la adquisición de nuevos materiales y sugieren la conservación de determinados documentos. La facilidad de consulta y descarga de los fondos potencia la polivalencia en los roles de estos profesionales encargados de la elaboración de los programas. No obstante, a lo largo de esta investigación se ha constatado que, en este escenario de convergencia mediática, los documentalistas de la radio pública estatal conservan su papel de apoyo a los usuarios del fondo en la búsqueda de recursos documentales para la producción de programas (Prieto, 2007).

3.4. Mayor protagonismo de los documentalistas

El acceso a los fondos documentales de otros perfiles profesionales que participan en la elaboración de los programas está desplazando el rol de los documentalistas hacia posiciones más especializadas que les ha permitido una mejor capacitación tecnológica, así como un mayor dominio de la metodología para la búsqueda de documentación y la valoración exhaustiva de los documentos en aras de contrastar su veracidad: una cuestión ampliamente demandada desde los profesionales del sector audiovisual (Hidalgo, 2005). Ahora su actividad tiende a centrarse en el trabajo de catalogación, administración y establecimiento de criterios de conservación y consulta especializada.

En todo caso, los profesionales de la radio pública que han sido entrevistados en el trabajo de campo consideran que en un entorno convergente como el actual caracterizado por la abundancia informativa de internet, la labor de los documentalistas será cada vez más importante porque esa información necesita ser contrastada (Fernández-Vegue¹,

2013). A esa tarea habría que sumar su importante papel de apoyo al desarrollo de competencias para la investigación y la búsqueda de recursos multimedia entre los productores de los programas (Bawden, 2002, pp. 361-408).

La facilidad de consulta y descarga de los fondos potencia la polivalencia en los roles de los profesionales encargados de la elaboración de los programas

3.5. Hacia la dimensión multimedia

La confirmación de que la radio pública estatal se está alineando con las tendencias multimedia de la radio, es otro de los resultados de este estudio (González-Aldea, 2011, p. 89). Aunque en el caso de *RNE* los fondos documentales utilizados en la programación convencional se circunscriben mayoritariamente a documentos sonoros y escritos, es cada vez más habitual el uso de fragmentos de programas televisivos y fotos en su web corporativa. Los *podcasts*, *blogs* y *webcasters* (streamings) asociados a la programación de *RNE* demandan el uso de documentos audiovisuales para estas manifestaciones de radio 2.0 que ofrecen un valor añadido a su programación convencional, como es el caso de los blogs de los programas o la última oferta de *Radio 3*, denominada “Radio3 Extra”, a la que se puede acceder a través del portal interactivo de *RTVE*.

<http://www.rtve.es>

3.6. Depositario de la memoria sonora de España

Mención especial merece la revisión del binomio *ARCA*/memoria histórica. En el caso de *Radio Nacional de España*, dice García-Jiménez (2002) que su fondo documental desempeña, también, la función de archivo sonoro histórico. Consciente de esa función, la radio pública estatal está grabando y digitalizando las 24 horas de emisión nacional de todos sus canales. Sin embargo, aunque se prevé hacerlo a corto plazo, todavía no se conservan en *ARCA* las desconexiones territoriales diarias producidas desde las diferentes cabeceras autonómicas y locales de *RNE*. No obstante, los documentos sonoros producidos desde las emisoras territoriales y locales de *RNE* que tienen interés informativo para toda España, habitualmente son recogidos por los programas de emisión nacional. En todo caso, el objetivo de la cadena pública es conseguir que todas sus emisiones de carácter territorial y local puedan estar grabadas en *ARCA*.

4. Conclusiones y discusión

Las conclusiones de la investigación apuntan a un cambio de escenario y a un nuevo protagonismo de la gestión de los contenidos documentales para la producción de programas en radio. Como afirma Rubio-Lacoba (2012), en los medios de comunicación nunca antes se había hablado tanto de etiquetas, de indización y de contextos informativos relacionados. Hoy nos encontramos ante un nuevo marco en el tratamiento informativo y documental en el que las rutinas profesionales y procesos de producción informativos están demandando un replanteamiento del papel de los profesionales de la radio que participan en la elaboración de los contenidos.

En alguna ocasión se ha afirmado que la radio pública se caracteriza, todavía hoy, por un cierto encorsetamiento en las funciones asignadas a las diferentes categorías laborales, lo que dificulta su polivalencia y, por tanto, la adaptación a los nuevos roles que demanda la convergencia de medios y la implantación de las TIC (López-Vidales; Ortiz-Sobrino, 2011). Sin embargo, en la investigación se ha comprobado que, en lo que se refiere a la utilización de documentos y recursos audiovisuales, la situación ha cambiado. Ahora, prácticamente todos los profesionales de la radio pública que participan en la elaboración de programas se implican en el proceso de búsqueda y selección de contenidos documentales para sus programas.

Prácticamente todos los profesionales de la radio pública que participan en la elaboración de programas se implican en la búsqueda y selección de contenidos documentales

A pesar de que algún investigador, como Giménez-Rayó (2013), acusa de cierto intrusismo a los periodistas y profesionales radiofónicos por el hecho de que puedan consultar y seleccionar con plena autonomía los contenidos de los fondos documentales, no se puede dar la espalda a los nuevos modelos productivos que se están implantando con la convergencia de medios. Así lo entienden los documentalistas de RNE que no sólo consideran positivo que los productores de contenidos accedan a los fondos documentales, sino que procuran simplificar los sistemas de búsqueda e identificación de los documentos para facilitar su utilización (Pérez-Carabias², 2013). Creen que en un futuro inmediato el papel del documentalista se irá desplazando hacia labores más especializadas en la gestión de los fondos documentales, aunque sigan manteniendo las funciones de consulta y selección.

Este desplazamiento de los documentalistas hacia territorios profesionales de mayor especialización ya ha sido puesto de manifiesto en alguna ocasión. Algunos autores han venido constatando que la convergencia mediática está propiciando la aparición de nuevos perfiles profesionales relacionados con las nuevas necesidades de catalogación y gestión de los contenidos para las redacciones digitalizadas de la radio y la televisión, como es el caso de los gestores de archivo (Meana et al., 2010). No obstante, entre los responsables de las diferentes áreas de la cadena productiva de la radio pública estatal se detecta que, cuando la búsqueda de fondos es relativamente compleja o especializada, los profesionales que intervienen en la programación siguen recurriendo a los documentalistas porque la premura de tiempo con la que habitualmente se trabaja hace, en muchas ocasiones, imposible conciliar la búsqueda de contenidos del fondo documental con la urgencia de la emisión. Esto significa que, aún hoy, existe una convivencia de las competencias tradicionales de los documentalistas con sus nuevas funciones derivadas del impacto de las TIC y la convergencia de medios.

Uno de los aspectos que tras el trabajo de campo ha hecho cambiar la apreciación de partida de los responsables de

esta investigación, ha sido la incompatibilidad de ARCA con el sistema informático utilizado para la producción de programas y su acceso a buscadores de la Red. A primera vista, sorprendía que el sistema ARCA –que nació con vocación de plataforma única de acceso a las fuentes documentales– no contemplara entre sus enlaces externos la conexión directa a internet o su compatibilidad con los sistemas digitales de producción de programas –*Dalet* y *MAR4*– utilizados por RNE. Sin embargo, los responsables de esta investigación han asumido las tesis de los documentalistas del fondo documental de la radio pública, partidarios de mantener la independencia de ARCA respecto a otras plataformas porque facilita la inmunidad de los fondos documentales ante posibles ataques virales provenientes de la Red o de los sistemas de producción de programas.

En relación con las prestaciones del sistema ARCA y el acceso a las diferentes bases de datos del fondo documental de RTVE, es pertinente incorporar en este epígrafe de conclusiones una observación que incide en el carácter inacabado del proyecto. Antes se ha comentado que la nueva plataforma permite al usuario de RNE el acceso a muchos de los contenidos documentales de TVE. Sin embargo, la documentación producida desde el Área de Informativos de la televisión pública estatal está gestionada desde una plataforma digital diferente, denominada *Invenio*. Aunque estos fondos documentales ya se están incorporando al sistema ARCA, hasta que el proceso de migración no finalice será necesario el envío por RDSI de algunos de los contenidos informativos emitidos por la televisión pública estatal, desde el Centro de Documentación de Informativos de TVE al centro de producción de la radio, en caso de ser solicitados por alguno de los programas.

Por último, en lo que respecta al papel de la radio pública estatal en la conservación de la memoria sonora de España, también se debe considerar a los fondos documentales gestionados desde ARCA como un proyecto dinámico e inacabado. En la investigación se ha comprobado que las emisiones diarias de carácter nacional de las diferentes cadenas de la radio pública estatal se graban y conservan en formato digital. Sin embargo, esta tarea sólo podrá considerarse completa una vez que acabe de implementarse la migración de los fondos documentales existentes en los archivos de los centros territoriales de *Radio Nacional de España* y la digitalización de las emisiones diarias de los contenidos autonómicos emitidos desde sus diecisiete Centros de Producción Territorial.

Notas

1. Juan Fernández-Vegue es director de Informativos de *Radio Nacional de España*.
2. Mónica Pérez-Carabias es directora del Archivo Sonoro de *Radio Nacional de España*.

5. Bibliografía

Area, Manuel; Guarro, Amador (2012). “La alfabetización informacional y digital: fundamentos pedagógicos para la enseñanza y el aprendizaje competente”. *Revista española de documentación científica*, n. monográfico, pp. 46-74. <http://dx.doi.org/10.3989/redc.2012.mono.977>

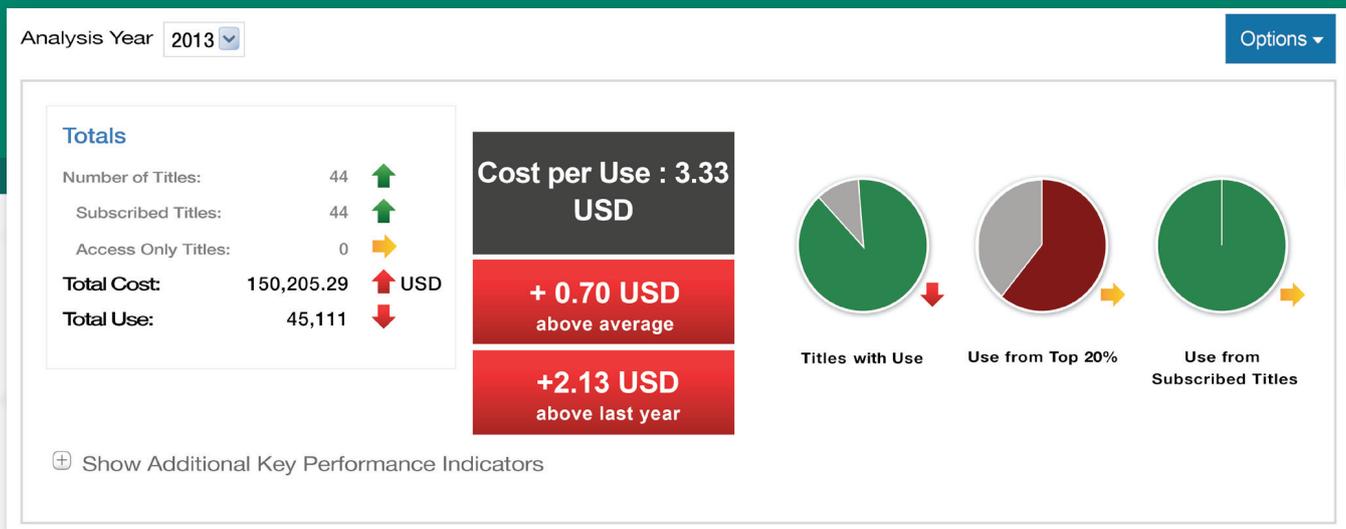
- Ariza-Chicharro, Rosa-María** (2004). "El Archivo de la Palabra de Radio Nacional de España". *Revista general de información y documentación*, v. 14, n. 2, pp. 29-58.
- Bawden, David** (2002). "Revisión de los conceptos de alfabetización informacional y alfabetización digital". *Revista anales de documentación*, n. 5, pp. 361-408.
- Caldera-Serrano, Jorge** (2008). "Changes in the management of information in audio-visual archives following digitalization: Current and future outlook". *Journal of librarianship and information science*, n. 40, pp. 13-20. <http://dx.doi.org/10.1177/0961000607086617>
- Caridad, Mercedes; Hernández, Tony; Rodríguez, David; Pérez, Belén** (2011). *Documentación audiovisual. Nuevas tendencias en el entorno digital*. Madrid: Síntesis. ISBN: 978 84 975 6746 6
- García-Jiménez, Antonio** (2002). "El Centro de Documentación Informativa de RNE". En: Galdón, Gabriel (coord.). *Teoría y práctica de la documentación informativa*. Barcelona: Ariel, pp. 273-292. ISBN: 978 84 344 1293 4
- Giménez-Rayó, Mabel** (2012). "La documentación audiovisual en televisión en el mundo 2.0: retos y oportunidades". *Trípodos*, n. 31, pp. 79-97.
- González-Aldea, Patricia** (2011). "La radio en Internet: las webs de las cadenas analógicas tradicionales". En: Ortiz-Sobrino y López-Vidales (eds.). *Radio 3.0: una nueva radio para una nueva era*. Madrid: Fragua, pp. 89-123. ISBN: 978 84 7074 436 5
- Hidalgo-Goyanes, Paloma** (2005). "La documentación audiovisual de las televisiones. La problemática actual". *Documentación de las ciencias de la información*, n. 28, pp. 159-171.
- López-Vidales, Nereida; Ortiz-Sobrino, Miguel-Ángel** (2011). "Viejas nomenclaturas, viejas competencias. Perfiles profesionales en la radio española". *Revista Telos*, n. 87. <http://goo.gl/Nio4QF>
- Meana-Alonso, Sonsoles; Muñoz-De-la-Peña, Paz; Sáez-Carreras, Susana** (2010). "El gestor de archivo, nuevo perfil profesional en la redacción única de TVE". En: *12^{es} Jornades datalanes d' Informació i Documentació*, Barcelona. <http://goo.gl/LRGwO7>
- Micó-Sanz, Josep-Lluís; Masip-Masip, Pere; García-Avilés, José-Alberto** (2009). "Periodistas que ejercen de documentalistas (¿y viceversa?). Nuevas relaciones entre la redacción y el archivo tras la digitalización de los medios". *El profesional de la información*, v. 18, n. 3, pp. 284-290. http://eprints.rclis.org/13756/1/Masip_EPI.pdf
- Nuño-Moral, María-Victoria** (2007). *Documentación en el medio radiofónico*. Madrid: Editorial Síntesis. ISBN: 978 84 9756 460 1
- Prieto, Laura** (2007). "Información y documentación radiofónica: espacios para un interés común. La experiencia de Radio Nacional de España". *El profesional de la información*, v. 16, n. 5, pp. 443-499. <http://eprints.rclis.org/12437/1/prieto.pdf>
- Rubio-Lacoba, María** (2007). *Documentación informativa en el periodismo digital*. Madrid: Editorial Síntesis. ISBN: 978 84 9756 459 5
- Rubio-Lacoba, María** (2013). "Nuevas destrezas profesionales para periodistas: el vocabulario colaborativo del diario El País". *Trípodos*, n. 31, pp. 65-78. <http://eprints.rclis.org/18511>
- Sierra-Bravo, Restituto** (2008). *Técnicas de investigación social*. Madrid: Thomson. ISBN: 978 84 2832 429 8
- Torregrosa, Juan-Francisco** (2008). *Documentación informativa en organizaciones de comunicación*. Madrid: Editorial Universitat. ISBN: 978 84 7991 230 7

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 (96,69 +21% IVA euros/año), puesto que en esta modalidad no hay que cubrir los gastos de imprenta ni de correo postal.

Optimize your collection with ease



With EBSCONET® Usage Consolidation & Analytics...

you get instant insight into critical usage data, including cost per use, in the same application where you manage your subscriptions.

- + Perform on-the-spot analysis of journal subscriptions and packages
- + View results in easy-to-read dashboards
- + Generate detailed reports to share
- + Capture and report on usage for all your online resources



ESTRATEGIAS DE DIFUSIÓN DE UNA REVISTA CIENTÍFICA. UN EXPERIMENTO CON *EL PROFESIONAL DE LA INFORMACIÓN*



Sergi Cortiñas-Rovira y Xavier Ramon-Vegas



Sergi Cortiñas-Rovira es profesor del *Depto. de Comunicación* de la *Univ. Pompeu Fabra (UPF)*. Doctor en comunicación social por la *UPF*, licenciado en ciencias químicas por la *Univ. de Barcelona (UB)* y periodismo por la *UPF*, investiga sobre divulgación de la ciencia. Dirige el proyecto competitivo I+D+i *El periodismo científico en España y las nuevas tecnologías de la información (TIC): mapa de situación y propuesta de actuación para mejorar los procesos comunicativos*, del que forma parte este artículo (*M^o de Economía y Competitividad*, Ref.: CSO2011-25969).
<http://orcid.org/0000-0002-7252-5418>

sergi.cortinas@upf.edu



Xavier Ramon-Vegas es profesor e investigador del *Departamento de Comunicación* de la *Universitat Pompeu Fabra*. Doctorando en comunicación, es máster en estudios avanzados en comunicación social y licenciado en comunicación audiovisual y periodismo por la *UPF*.
<http://orcid.org/0000-0002-4478-5626>

xavier.ramon@upf.edu

Departamento de Comunicación, Universitat Pompeu Fabra
Roc Boronat, 138. 08018 Barcelona
<http://www.upf.edu/decom/es>

Resumen

Partiendo de los fundamentos teóricos sobre el papel de los gabinetes de prensa en la transmisión de contenidos de las revistas científicas, se analiza un experimento realizado entre noviembre de 2011 y noviembre de 2012 para difundir los contenidos de *El profesional de la información*. El documento detalla la propuesta metodológica, las fuentes utilizadas, los focos de envío y la selección y edición de la información. Se ahonda además en las características de los impactos obtenidos. Finalmente se exponen propuestas futuras para potenciar la transmisión y monitorización de los contenidos científicos mediante las notas de prensa y las redes sociales.

Palabras clave

Difusión del conocimiento científico, Revistas científicas, Gabinetes de comunicación, Notas de prensa, Experimento, *El profesional de la información*.

Title: Strategies for disseminating a scientific journal. An experiment with *El profesional de la información*

Abstract

Starting from the theoretical foundations of the role of press offices in the transmission of the content of scientific journals, the article analyses an experiment carried out between November 2011 and November 2012 with *El profesional de la información*. The methodological approach, the sources, the dissemination targets and the selection and edition of the information are discussed. Moreover, an in-depth analysis of the characteristics of the achieved impacts is presented. Finally, future research is suggested to enhance the transmission and monitoring of scientific content through press releases and social networking sites.

Keywords

Dissemination of scientific knowledge, Scientific journals, Communication offices, Press releases, Experiment, *El profesional de la información*.

Cortiñas-Rovira, Sergi; Ramon-Vegas, Xavier (2013). "Estrategias de difusión de una revista científica. Un experimento con *El profesional de la información*". *El profesional de la información*, septiembre-octubre, v. 22, n. 5, pp. 405-414.

<http://dx.doi.org/10.3145/epi.2013.sep.05>

Artículo recibido el 19-06-2013
Aceptación definitiva: 26-07-2013



Figura 1. El Periòdic. Periòdic digital de la Comunitat Valenciana

1. El papel de los gabinetes de prensa en la difusión de contenidos científicos

La actividad de los gabinetes de prensa en España cuenta con una dilatada trayectoria (Almansa-Martínez, 2011, pp. 36-53). Sin embargo la difusión de los contenidos científicos por parte de los medios ha sido escasa. Por un lado “las noticias científicas tan sólo aparecen en los medios de comunicación cuando el descubrimiento está avalado por una publicación de impacto o por un premio de prestigio” (Domínguez-Romero, 2002, p. 50). Por otra parte, se percibe una baja visibilidad de las investigaciones españolas frente a los contenidos extranjeros (Nieto; Vendrell, 2001), así como la minimización de campos fuera de la medicina, agricultura o medio ambiente (Elías, 2002).

En el ecosistema comunicativo actual, las revistas científicas deben aprovechar todas las fórmulas que ofrece el entorno digital para gestionar eficientemente las relaciones con la prensa

Pese a todo, en el ecosistema comunicativo actual, marcado por la instantaneidad y la superabundancia informativa (Fernández-Asenjo; De-la-Torre-Alfaro, 2009), las revistas científicas deben aprovechar todas las fórmulas que ofrece el entorno digital para gestionar eficientemente las relaciones con la prensa (Almansa-Martínez, 2011; García-Orosa, 2009). La publicación en medios y sitios web institucionales es clave en la estrategia de promoción de los artículos científicos, junto con la difusión en repositorios y redes sociales (Baiget; Torres-Salinas, 2013, pp. 86-89). En esta línea la revista *El profesional de la información* (EPI) realiza una tarea de comunicación intensiva en su web y blog, así como en las redes sociales Facebook y Twitter. No obstante, potenciar los contenidos de EPI entre medios e instituciones era, y en cierta medida sigue siendo, un reto pendiente de superar. <http://www.elprofesionaldelainformacion.com/notas>

<https://www.facebook.com/elprofesionaldelainformacion>
http://twitter.com/revista_EPI

2. Experimento. Difusión de contenidos de EPI

2.1. Definición y propuesta metodológica

Con el objetivo de captar el interés informativo (Sissons, 2012) e intentar lograr impactos para la visibilidad de la revista, se realizó durante el período de 13 meses comprendido entre noviembre de 2011 y noviembre de 2012 un experimento de campo (Priest, 2009, p. 83). Este método se basa en “manipular la variable independiente para determinar si tiene una influencia en la variable dependiente” (Bryman, 2012, p. 50). En este caso se tomó como *variable independiente* una muestra

de artículos de EPI. Mediante el experimento los investigadores 1) diseñaron una estrategia de difusión, 2) transformaron los contenidos en notas de prensa y 3) recogieron datos sobre el impacto que éstas tuvieron sobre la *variable dependiente*, que en este experimento es el espacio informativo de los medios e instituciones (Campbell; Stanley, 1966; Priest, 2009; Webster; Sell, 2007).

2.2. Creación de un marco para la difusión de contenidos

En diciembre de 2011 se elaboró una primera base de datos de medios, instituciones y periodistas. Se tomó como punto de partida la *Agenda de comunicación 2011*, editada por la *Secretaría de Estado de Comunicación* (Ministerio de la Presidencia del Gobierno de España)¹. Se seleccionó un total de 1.120 direcciones relevantes de correo electrónico² de medios de diferentes soportes (prensa escrita, radio, televisión

Tabla 1. Destinatarios de la difusión considerados en el estudio

1	Periódicos nacionales y autonómicos de información general
2	Prensa nativa online
3	Revistas especializadas en tecnología, comunicación, cultura y arte
4	Revistas de información general
5	Revistas especializadas en ciencia
6	Agencias de noticias
7	Emisoras de radio y televisión
8	Listas de distribución científica (Agencia SINC, Associació Catalana de Comunicació Científica, Incyt, etc.)
9	Portales de comunicación (Portal de la comunicació, CSIC, Noticom, etc.).
10	Expertos y académicos
11	Departamentos de prensa y gabinetes de comunicación de universidades
12	Departamentos de comunicación de instituciones, gobiernos y empresas
13	Autores de blogs

e internet), tipo (generalistas y especializados) y ámbitos geográficos (nacionales, autonómicos y locales). Posteriormente se actualizó la base de datos a raíz de la publicación de la *Agenda de la comunicación 2012*³. Los contenidos de esta fuente se complementaron y ampliaron durante el estudio con recursos específicos (como el directorio *CRUE Gabinetes de Comunicación de las Universidades Españolas*⁴), webs institucionales y otras fuentes electrónicas, con el objetivo de hallar los contactos más pertinentes para cada ocasión. Se combinaron pues los correos genéricos con otros más específicos, dirigidos a algunos profesionales individuales, pues-

to que como señala **De-Bustos** (2012, p. 223), actualmente el “gabinete de prensa debe tener al redactor como punto de partida de una red de difusión en la que ocupa el centro indiscutible. El periodista ha desplazado al medio como referencia del gabinete de prensa”. También se difundieron las notas de prensa entre autores de blogs. En contraste con la posición mostrada por **Dans** en su artículo “Y el lugar de las notas de prensa... ¿qué?”⁵, creemos que los contenidos no sólo tienen un valor potencial para los medios, sino que las notas de prensa deben apuntar al espectro más amplio de

destinatarios posibles, siempre y cuando el envío de estas notas se realice de forma ética, responsable y adaptada a las necesidades e intereses específicos de los receptores.

2.3. Selección y redacción de las informaciones

Durante el estudio, llevado a cabo entre noviembre de 2011 y noviembre de 2012, se seleccionaron diez artículos, tomando como criterios básicos su calidad, interés, relevancia, carácter noticiable (**Baines; Egan; Jefkins, 2004; De-Se-mir; Ribas; Revuelta, 1998; Pander-Maat; De-Jong, 2012**) y



Figura 2. Universitat Pompeu Fabra

Tabla 2. Artículos de EPI utilizados en el experimento y número de impactos obtenido por cada artículo

Artículo	Impactos
Tur-Viñes, Victoria; Fonseca-Mora, María-Carmen; Gutiérrez-San-Miguel, Begoña (2012). “Ética de la publicación científica: iniciativas y recomendaciones”. <i>El profesional de la información</i> , v. 21, n. 5, pp. 491-497. http://dx.doi.org/10.3145/epi.2012.sep.07	5
Torres-Salinas, Daniel; García-Moreno-Torres, José; Robinson-García, Nicolás; Delgado-López-Cózar, Emilio; Herrera, Francisco (2011). “Rankings ISI de las universidades españolas según campos y disciplinas científicas (2ª ed. 2011)”. <i>El profesional de la información</i> , v. 20, n. 6, pp. 701-709. http://dx.doi.org/10.3145/epi.2011.nov.16	4
Hilera-González, José-Ramón; Fernández-Sanz, Luis; Suárez-Morón, Esther (2012). “Evaluación de la legibilidad de webs de universidades”. <i>El profesional de la información</i> , v. 21, n. 5, pp. 468-474. http://dx.doi.org/10.3145/epi.2012.sep.04	3
Isasi-Andrieu, Alberto; López-Carrera, Ángel; Ruiz-Ibáñez, Pilar; Angulo-Redondo, Iñaki (2012). “Gazela: asesor digital de redes sociales para adolescentes”. <i>El profesional de la información</i> , v. 21, n. 5, pp. 514-519. http://dx.doi.org/10.3145/epi.2012.sep.11	2
Benito-Blázquez, Belén; Doñate-Cifuentes, Ana; Gómez-Domínguez, David; Marquina-Arenas, Julián; Martínez-Cañadas, Evelio (2011). “Biblogtecarios: el poder de las conexiones y el crowdsourcing”. <i>El profesional de la información</i> , v. 20, n. 6, pp. 677-681. http://dx.doi.org/10.3145/epi.2011.nov.12	0
Blanco, Elena; Casaldàliga, Anna (2012). “Papel de la biblioteca en la difusión de la producción científica en la <i>Universitat Pompeu Fabra</i> ”. <i>El profesional de la información</i> , v. 21, n. 6, pp. 627-631. http://dx.doi.org/10.3145/epi.2012.nov.10	0
Izquierdo-Castillo, Jessica (2012). “Distribución online de contenidos audiovisuales: análisis de 3 modelos de negocio”. <i>El profesional de la información</i> , v. 21, n. 4, pp. 385-390. http://dx.doi.org/10.3145/epi.2012.jul.09	0
Larrondo, Ainara; Larrañaga-Zubizarreta, José; Meso, Koldo; Agirreazkuenaga, Irati (2012). “Convergencia de medios y redacciones: el caso de la <i>Radio Televisión Pública Vasca (EITB)</i> ”. <i>El profesional de la información</i> , v. 21, n. 4, pp. 347-353. http://dx.doi.org/10.3145/epi.2012.jul.03	0
Leydesdorff, Loet (2012). “Comparación y predicción de las cuotas mundiales de publicación de los EUA, EU-27 y China usando la nueva interfaz <i>Web of science</i> versus <i>Scopus</i> ”. <i>El profesional de la información</i> , v. 21, n. 1, pp. 43-50. http://www.elprofesionaldelainformacion.com/leydesdorff-espanol.pdf	0
Martínez-Sanz, Raquel (2012). “Estrategia comunicativa digital en el museo”. <i>El profesional de la información</i> , v. 21, n. 4, pp. 391-395. http://dx.doi.org/10.3145/epi.2012.jul.10	0



Figura 3. Parc de Recerca Biomèdica (PRBB)

representatividad respecto a los ejes temáticos de EPI, que se muestran en la tabla 2. Se trataron cuestiones referentes a la distribución audiovisual, bibliotecas, comunicación museística, convergencia mediática, redes sociales, producción y publicación científica o legibilidad web.

Cada pieza (la nota de prensa, en términos clásicos) se redactó siguiendo las características de la redacción periodística: estructuración en pirámide invertida, u orden lógico frente a orden cronológico; atención a las seis preguntas básicas (6 w questions); claridad, precisión, rigor, concreción y concisión en el texto y en los elementos de titulación (McCurry, 2004; Fernández-Asenjo; De-la-Torre-Alfaro, 2009; García-Orosa, 2009; Foster, 2012; Pander-Maat, 2007; Pander-Maat; De-Jong, 2012; Risquete, 2006; Sissons, 2012). También se atendió a la naturaleza hipertextual de la redacción online (Díaz-Noci, 2002; Salaverría, 2005) y se acompañó cada difusión del artículo completo publicado en EPI en pdf.

La redacción de las notas informativas se rigió por los criterios éticos fundamentales de la profesión (Fernández-Asenjo; De-la-Torre-Alfaro, 2009; Risquete, 2006). Según explica el Col·legi de Periodistes de Catalunya, “los gabinetes de comunicación prestan un servicio de información y de comunicación social, de acuerdo con las premisas deontológicas



Figura 4. Universitat de València

gicas del periodismo”⁶. Se marcaron los límites entre información y publicidad o marketing (De-Bustos, 2012; Foster, 2012), y se evitó la inclusión del lenguaje promocional descrito por Pander-Maat (2007, pp. 69-71), esto es, adjetivos y adverbios intensificadores, adjetivos valorativos o grados superlativos, entre otros. Asimismo se preservaron los principios de veracidad, libertad, justicia y responsabilidad social (Alsuis, 2010; Frost, 2011), recogidos en los códigos éticos del periodismo promovidos por la Unesco, la Federación Internacional de Periodistas, el Consejo de Europa o la FAPE⁷,

así como en los documentos de referencia en las relaciones públicas, elaborados por la Public Relations Society of America⁸ o la International Public Relations Association⁹.

El proceso de elaboración de cada nota (lectura y análisis del artículo original, adaptación y maquetación) y su distribución y envío (redacción y selección de los destinatarios relevantes para cada caso) requirió una dedicación aproximada de 3,5 horas por artículo. En el caso de las notas específicas o personalizadas, el tiempo utilizado fue superior (entre 5 y 6 horas).

3. Resultados del experimento

3.1. Difusiones realizadas

Se envió un total de 38 notas de prensa, correspondientes a 10 artículos publicados en *El profesional de la información*, aplicando varias estrategias de difusión:

- se elaboraron notas de prensa con un único enfoque, basadas en siete artículos (a saber: Blanco; Casaldàliga, 2012; Benito-Blázquez et al., 2011; Isasi-Andrieu et al., 2012; Izquierdo-Castillo, 2012; Larrondo et al., 2012; Leydesdorff, 2012; Martínez-Sanz, 2012);
- se realizaron notas personalizadas, como en el artículo “Ética de la publicación científica: iniciativas y recomendaciones” (Tur-Viñes; Fonseca-Mora; Gutiérrez-San-Miguel, 2012), del que se mandaron notas específicas a cada una de las universidades de las autoras. Esta estrategia también se refleja en otros artículos, como “Evaluación de la legibilidad de webs de universidades” (Hilera-González; Fernández-Sanz; Suárez-Morón, 2012), a partir del cual surgieron 6 notas: 1 pieza general y 5 notas específicas para los gabinetes de universidades (Alcalá de Henares, Barcelona, UNED, Pompeu Fabra y Politécnica de Madrid). A partir del artículo “Rankings ISI de las uni-

versidades españolas según campos y disciplinas científicas (2ª ed. 2011)” (Torres-Salinas *et al.*, 2012), se elaboró una nota personalizada para cada uno de los 21 centros recogidos en el estudio.

3.2. Impactos obtenidos

Se detectaron 14 impactos correspondientes a cuatro de los 10 temas mencionados anteriormente. Los seis restantes no obtuvieron impacto alguno. Según el impacto recibido:

1) el mejor resultado en términos de recepción fue para el artículo de **Tur-Viñes; Fonseca-Mora; Gutiérrez-San-Miguel** (2012). Los resultados científicos de este trabajo fueron recogidos por la *Universidad de Salamanca*¹⁰, *Diariodesalamanca.es*, *Salamancauniversitaria.com*, *Salamanca24horas.com*¹¹ y la *Agencia de Noticias para la Difusión de la Ciencia y la Tecnología (DiCYT)*¹².

“Cuatro artículos de *El profesional de la información* generaron 14 impactos en medios de comunicación y plataformas institucionales”

2) informaciones procedentes del artículo de **Torres-Salinas *et al.*** (2012), se reflejaron en las páginas de *El Periódic. Periódico digital de la Comunidad Valenciana*¹³, *Universitat Pompeu Fabra* de Barcelona, *Parc de Recerca Biomèdica (PRBB)*¹⁴ y *Universitat de València*¹⁵.

3) la nota del artículo “Evaluación de la legibilidad de webs de universidades” fue recogida por las páginas de la *Universitat de Barcelona*¹⁶ y de la *Universidad de Alcalá de Henares*¹⁷, así como en la página de *ESVI-AL (Educación Superior Virtual Inclusiva—América Latina)*¹⁸.

4) la nota “*Gazela*: asesor digital de redes sociales para adolescentes”, elaborada a partir del artículo de **Isasi-Andrieu *et al.*** (2012), fue publicada en el periódico *Deia.com*¹⁹ y en el blog *Ecommerce y Marketing*²⁰.

El número de impactos en medios de comunicación generalistas fue reducido en comparación con la repercusión en plataformas académicas. Cabe tener en cuenta la sobreabundancia de comunicados recibidos por los periodistas (**Fernández-Asenjo; De-la-Torre-Alfaro**, 2009), que provoca que se deseche “el 85% de las



Figura 5. *Universitat de Barcelona*

notas de prensa que llegan a las redacciones” (De-Bustos, 2012, p. 226). Además, cabe considerar el poco espacio dedicado por los medios a la investigación, las escasas fuentes de las noticias científicas publicadas en la prensa española (Elías, 2000) y “la inseguridad del redactor sobre lo que es noticia científica relevante” (Elías, 2008, p. 84).

Tampoco se pueden obviar las dificultades técnicas. El envío de la primera nota de prensa subrayó una carencia muy significativa de la *Agenda de la Comunicación*. Entre las 1.120 direcciones que integraron el primer listado de *mailing*, se obtuvo una tasa de correos rechazados del 10%. Se observaron como causas principales de estos rechazos los buzones de correo llenos y direcciones desactualizadas. Esta cifra



Figura 6. *Universidad de Alcalá de Henares*

ESVI-AL Educación Superior Virtual Inclusiva - América Latina

Salto al contenido Mapa Del Sitio Accesibilidad A+ A- Alto Contraste

Mejora de la Accesibilidad en la Educación Superior Virtual en América Latina

«Ir Abajo»

Publicación “Evaluación de la legibilidad de webs de Universidades”

El equipo ESVI-AL de Universidad de Alcalá publicó el artículo titulado “Evaluación de la Legibilidad de Webs de Universidades” en la revista “El profesional de la información”, la misma es una revista internacional de información indexada por ISI SSCI, Scopus y otras bases de datos con alto factor de impacto que contribuye a la visibilidad de los resultados del proyecto ESVI-AL.

Los autores del artículo titulado: “Evaluación de la Legibilidad de Webs de Universidades”, son: José Ramón Hiera González (Coordinador general del proyecto ESVI-AL), Luis Fernández Sanz y Esther Suárez Morón, profesores de la Universidad de Alcalá.

Figura 7. Educación Superior Virtual Inclusiva – América Latina (ESVI-AL)

manifiesta el grado de obsolescencia de un documento de referencia para el sector comunicativo español. En la *Agenda de la Comunicación 2012* se detectaron carencias similares, puesto que se incluían empresas que habían cesado su actividad o habían cambiado de denominación.

3.3. Características de los impactos obtenidos

Mediante la aplicación de la técnica del análisis del contenido (Bryman, 2012; Krippendorff, 2004; Ruiz-Olabuénaga; Aristegui; Melgosa, 1998), se estudiaron las piezas publicadas y se alcanzaron los siguientes resultados:

Se puede apreciar una ligera variación en cuanto al número de palabras del texto original. Si bien las notas enviadas desde *EPI* tenían una extensión media de 403 palabras, esta cifra ascendió hasta las 459 palabras en los materiales publicados. Sin embargo, no se produjeron grandes cambios a nivel de contenido editorial, pues se respetó el sentido y el fondo de la información transmitida. En determinados casos, como en el texto de la *Universidad de Salamanca*, el incremento de palabras respecto a la nota enviada (de 421 a 644) se acompañó de una mayor profundización en el artículo original y de información docente e investigadora de la autora. En otras ocasiones la traslación de contenidos fue casi exacta, lo que reforzó la idea de que en lugar de editar nuevamente la información, se replicó prácticamente el material proporcionado (Sissons, 2012).

La circulación libre y eficaz de los contenidos científicos es uno de los principales deberes de las sociedades democráticas modernas

En todos los casos se citó a *El profesional de la información* como fuente, aunque sólo *ESVI-AL* añadió más datos contextuales sobre la revista (índices y factor de impacto).

Todos los textos mencionaron a los autores de los artículos y su filiación, a excepción de la *UB*, que no referenció a una de las autoras del artículo sobre legibilidad de webs. Pese a todo, el título exacto del artículo publicado en *EPI* sólo se reflejó en nueve de catorce ocasiones y sólo en los casos de *UPF* y *PRBB* se incluyó la referencia bibliográfica completa.

Los enlaces a la revista o al resumen del artículo en la página web de la misma fueron escasos (*UPF*, *PRBB*, *UB* y *ESVI-AL*). Cabe añadir que sólo dos piezas (*Deia.com* y *ESVI-AL*) incluyeron citas entrecorrientes con contenido de los artículos científicos.

Finalmente, no se detectaron incorrecciones en la información transmitida, a excepción de un leve error (*UPF*) en que se señaló a “Thomson-ISI” como productor de la base de datos, cuando hablando con propiedad, la base de datos *ISI Web of knowledge* es editada por *Thomson-Reuters*.

4. Conclusiones y líneas futuras de investigación

Como se ha explicado, la muestra se ha circunscrito a una revista, un ámbito científico y un período temporal concreto. Por consiguiente, no es posible extrapolar los resultados hacia otros contextos. Sin embargo, aquellos aportan datos de interés que permiten apuntar a futuras investigaciones.

Se respetó el sentido y fondo de la información, pero en la mayoría de los casos simplemente se replicó el material proporcionado y se demostró una escasa profundización en el artículo original

Los resultados obtenidos hasta el momento indican la pertinencia de ahondar en la estrategia de realización de notas de prensa, con el objetivo de contribuir a una mayor visibilidad de los contenidos de la revista. Las notas de prensa deben seguir jugando un papel fundamental en la divulgación de la actividad científica. Una cuidada adaptación de los artículos se considera clave para una posterior difusión completa y que respete la información relevante y las fuentes utilizadas en cada trabajo. Para ello es recomendable que las revistas cuenten con un equipo de colaboradores que diseñe y aplique una estrategia de difusión de contenidos, que vele por la calidad de los mismos y que ahonde en la relación con los destinatarios de la información, por teléfono y encuentros personales, si es posible, ya que la interacción y la recepción de *feedback* (García-Orosa, 2009; Ros-Martín, 2008) también puede resultar beneficiosa para mejorar la calidad de los contenidos (Escribà; Cortiñas-Rovira, 2013).

Además conviene seguir potenciando el uso de las redes sociales²¹. Las revistas no pueden infravalorar los beneficios

de *Twitter*, utilizado de forma creciente como herramienta por parte de los profesionales para aproximarse y citar a las fuentes (Broersma; Graham, 2013). La capacidad de adaptarse a los nuevos canales y determinar su utilidad para la difusión científica deviene una tarea de primer orden.

Estas cuestiones señalan algunos retos futuros de investigación. En primer lugar, se debe seguir trabajando y monitorizando sobre las notas de *El profesional de la información*. En segundo lugar, sería relevante realizar el mismo experimento con diferentes revistas de impacto a nivel nacional e internacional, con el fin de analizar y comparar las estrategias de difusión y los impactos obtenidos. En nuevos experimentos, se debería potenciar la difusión de artículos en las redes sociales generalistas y plataformas académicas especializadas (*Academia.edu* y *ResearchGate.com*). Por último sería pertinente triangular (Tashakkori; Teddlie, 2009) la metodología experimental con entrevistas en profundidad (Marshall; Rossman, 2011; Rubin; Rubin, 1995) con autores, miembros de consejos editoriales de revistas y receptores de la información científica (investigadores, profesorado, periodistas, profesionales de los sectores relevantes, ciudadanos), para conocer mejor las expectativas de producción y recepción de los materiales. De este modo en futuros trabajos también se podría examinar si las acciones realizadas se traducen en un incremento de las suscripciones a dichas publicaciones.

Y es que más allá de las voluntades puramente promocionales de las revistas o de las entidades editoras, está en juego uno de los principales deberes de las sociedades democráticas modernas: la circulación libre y diligente de la investigación puntera de calidad, obtenida a través del *método científico*, uno de los mayores patrimonios que el conocimiento ha dado en todos los tiempos. Estamos todavía lejos de que el costoso camino que va de la comunidad científica especializada hacia el gran público sea firme y duradero; y sea, a su vez, transitado por los numerosos actores implicados de forma frecuente y eficaz.

Accesibilidad | Mapa web | USAL Digital | Directorio | Registro | Correo | Bibliotecas | Mi USAL | Buscar

UNIVERSIDAD DE SALAMANCA
CAMPO DE EXCELENCIA INTERNACIONAL

Estás en: Inicio > Noticias > Un artículo de una profesora de la Universidad de Salamanca ofrece ...

Noticia

RSS | Compartir

Un artículo de una profesora de la Universidad de Salamanca ofrece recomendaciones para mejorar la honestidad y calidad de las publicaciones científicas

16/11/2012 | Comunicación Universidad de Salamanca

La revista especializada *El Profesional de la Información* publica 'Ética de la publicación científica', de Begoña Gutiérrez San Miguel, en un número dedicado a información y derecho

La profesora de la Universidad de Salamanca Begoña Gutiérrez

búsqueda de noticias

Palabra clave

Ámbito

Fecha desde hasta

Figura 8. Universidad de Salamanca

Diario de Salamanca.es
Universitario

Universidad | Local | Cultura | Deportes | Sucosos | Opiniones | + Vida | + Noticias

Erasmus | Perdidos | Eventos | Recomendaciones | Entrevistas | Columnistas | Puerta Abierta | Reportajes

CONSERVATORIO SUPERIOR DE MÚSICA DE SALAMANCA
C/ Lazarillo de Tormes, 54 37005 Salamanca
www.consuperiorsal.com

Recomendaciones para mejorar la honestidad y calidad de las publicaciones científicas

16/11/2012 | Archivado en: twitter, Universidad, Usal | Autor: Redacción

La revista especializada *El*

Erasmus próximos Eventos
Ciclo Idiomas V. O. 'Mi semana con Ma (Inglés) 22 nov 12

Figura 9. Diario de Salamanca

Quiénes somos | Preguntas frecuentes | Apoyos oficiales | Aviso legal | Podcast | RSS

Principal | Ciencia | Cultura | Educación

Portada | Artículos | Eventos | Archivo | Blogs | Mediateca

Eventos: 21 noviembre de 2012 Congreso Internacional

ESPAÑA

Recomendaciones para mejorar la honestidad y calidad de las publicaciones científicas

La revista especializada *El Profesional de la Información* publica 'Ética de la publicación científica', de una profesora de la Universidad de Salamanca

JPA/DICYT La profesora de la Universidad de Salamanca Begoña Gutiérrez San Miguel, en colaboración con Victoria Tur Viñes, de la Universidad de Alicante, y María Carmen Fonseca Mora, de la Universidad de Huelva, ha publicado el artículo 'Ética de la publicación científica: iniciativas y recomendaciones' en la revista *El Profesional de la Información*. En él se aportan las claves para una publicación científica ética y responsable.

Según ha informado hoy la Universidad de Salamanca, el último número de esta revista internacional sobre información, bibliotecas y nuevas tecnologías de la información, recoge un artículo que, según sus autoras, responde a la necesidad ineludible de incrementar la calidad de las revistas científicas.

La creciente importancia de las publicaciones científicas de calidad, su impacto en la sociedad en general y en los procesos de acreditación de los autores, así como la necesidad de visibilidad internacional, pueden provocar un incremento de los conflictos de

Noticias relacionadas

La Universidad de Salamanca ha publicado más de 5.500 artículos científicos en los últimos 15 años

"Una lectura crítica de los artículos científicos es fundamental para el autoaprendizaje de los profesionales médicos"

Los científicos salmantinos se acercan al proceso de publicación de artículos científicos en las revistas análogas

Los expertos estiman en 500 las revistas científicas españolas 'de calidad'

España se sitúa entre los cinco países con más publicaciones científicas sobre Matemáticas

Una revista dirigida por el jefe de Cirugía de León ingresa en la lista de publicaciones por factor de

Figura 10. Agencia de Noticias para la Difusión de la Ciencia y la Tecnología (DiCYT).

5. Notas

1. Agenda de la comunicación 2011

<http://es.scribd.com/doc/51112123/Agenda-de-la-Comunicacion-2011-Espana-Ministerio-de-la-Presidencia-Secretaria-de-Estado-de-Comunicacion-Espana>



Figura 11. Deia.com

2. Para conducir el experimento se descartó el uso del correo postal. Se valoró la visibilidad que se puede llegar a obtener con este método, pero se consideraron sus inconvenientes, como la dificultad de llegar a cada receptor de forma directa, rápida y eficiente o la imposibilidad de aportar elementos de valor añadido, como los enlaces, materiales en otros soportes o el texto electrónico, que posteriormente puede ser copiado, adaptado y refundido.

3. *Agenda de la comunicación 2012*
<http://www.lamoncloa.gob.es/NR/rdonlyres/120271BF-7C97-422E-B6B0-3938B9BE2E5F/0/Agenda2012.pdf>

4. Listado de gabinetes de comunicación de las universidades españolas:
http://www.crue.org/prensa/documentos/Gabinetes_Universidades_Espanolas_marzo2013.pdf

5. **Dans, Enrique** (2012). "Y el lugar de las notas de prensa... ¿qué?". *El blog de Enrique Dans*, 27 de diciembre.
<http://www.enriquedans.com/2012/12/y-en-lugar-de-las-notas-de-prensa-que.html>

6. *Col·legi de Periodistes de Catalunya*. "Els gabinets de comunicació. Criteris de bones pràctiques professionals".
http://www.periodistes.org/fcic/admin/doc_contingut/MANUAL_BONES_PRACTIQUES.pdf

7. Recopilación de códigos éticos del periodismo realizada por los profesores **Kaarle Nordenstreng** y **Ari Heinonen** de la *University of Tampere* (Finlandia).
http://ethicnet.uta.fi/codes_by_country

8. *Public Relations Society of America (PRSA)* "Member code of ethics".
<http://www.prsa.org/AboutPRSA/Ethics/documents/Code%20of%20Ethics.pdf>

9. *International Public Relations Association (IPRA)*. "Code

of conduct".
<http://www.ipra.org/about/ipra-codes>

10. "Un artículo de una profesora de la *Universidad de Salamanca* ofrece recomendaciones para mejorar la honestidad y calidad de las publicaciones científicas". *Universidad de Salamanca*. 16/11/2012.
<http://www.usal.es/webusal/node/26597>

11. La dirección de la pieza publicada por *Diariodesalamanca.es* no está disponible. En las siguientes direcciones se pueden encontrar las otras notas refundidas:

<http://www.salamancauniversitaria.com/usal/906-recomendaciones-para-mejorar-la-honestidad-y-calidad-de-las-publicaciones-cientificas>

<http://www.salamanca24horas.com/local/77343-recomendaciones-para-mejorar-la-honestidad-y-calidad-de-las-publicaciones-cientificas>

12. "Recomendaciones para mejorar la honestidad y calidad de las publicaciones científicas". *DiCYT*, 16/11/2012.

<http://www.dicyt.com/noticias/recomendaciones-para-mejorar-la-honestidad-y-calidad-de-las-publicaciones-cientificas>

13. "La *Universitat de València*, una de las cinco primeras universidades españolas en producción científica". *El Periòdic. Periòdic digital de la Comunitat Valenciana*. 28/01/2012.

http://www.elperiodic.com/pvalencia/noticias/156101_universitat-valencia-cinco-primeras-universidades-espaolas-produccion-cientifica.html

14. Debido a la migración a la nueva web de la *Universitat Pompeu Fabra* llevada a cabo a principios de septiembre de 2012, estas direcciones ya no están disponibles.

15. "La *Universitat de València*, una de las cinco primeras universidades españolas en producción científica". *Universitat de València*. 26/01/2012.

<http://www.uv.es/uvweb/universidad/es/lista-noticias/universitat-valencia-cinco-primeras-universidades-espanolas-produccion-cientifica-1285846070123/Noticia.html?id=1285849964442>

16. "La *UB*, entre las mejor posicionadas en el 'ranking' de legibilidad de las webs universitarias españolas". *Universitat de Barcelona*, 19/10/2012.

http://www.ub.edu/web/ub/es/menu_eines/noticies/2012/10/091.html

17. "Dos profesores de la *UAH* publican un artículo sobre la legibilidad de las webs universitarias". *Universidad de Alcalá de Henares*, 15/10/2012.

http://www2.uah.es/diariodigital/index2.php?option=com_content&task=view&id=6449

18. "Publicación 'Evaluación de la legibilidad de webs de

Universidades". *Educación Superior Virtual Inclusiva – América Latina (ESVI-AL)*.

<http://www.esvial.org/?p=678>

19. "Un software online para proteger a adolescentes en Facebook". *Deia.com*. 24/11/2012.

<http://www.deia.com/2012/11/24/ocio-y-cultura/internet/un-software-online-para-proteger-a-adolescentes-en-facebook>

20. "Tecnalia ha desarrollado *Gazela*, una plataforma online para proteger a adolescentes en el uso de las redes sociales". *Ecommerce y marketing*, 26/11/2012.

<http://ecommerceymarketing.blogspot.com.es/2012/11/tecnalia-ha-desarrollado-gazela-una.html>

21. Durante el estudio, los investigadores no difundieron el contenido de las notas de prensa o de los artículos originales en las redes sociales.

Las revistas deben contar con una estrategia de difusión de contenidos, potenciar las redes sociales y ahondar en la relación con los destinatarios de la información

6. Bibliografía

Almansa-Martínez, Ana (2011). *Del gabinete de prensa al gabinete de comunicación. La dirección de comunicación en la actualidad*. Zamora: Comunicación Social. ISBN: 978 84 92860 90 6

Alsius, Salvador (ed.) (2010). *The ethical values of journalists. Field research among media professionals in Catalonia*. Barcelona: Generalitat de Catalunya. ISBN: 978 84 39383 46 8

Baiget, Tomàs; Torres-Salinas, Daniel (2013). *Informe APEI sobre publicación en revistas científicas*. Gijón: Asociación Profesional de Especialistas en Información. <http://www.apei.es/informes/InformeAPEI-Publicaciones-cientificas.pdf>

Baines, Paul; Egan, John; Jenkins, Frank (2004). *Public relations: contemporary issues and techniques*. Oxford: Elsevier Butterworth-Heinemann. ISBN: 0750657243

Broersma, Marcel; Graham, Todd (2013). "Twitter as a news source. How Dutch and British newspapers used tweets in their news coverage, 2007–2011". *Journalism practice*, publicado online 30 de mayo, pp. 1-19. <http://dx.doi.org/10.1080/17512786.2013.802481>

Bryman, Alan (2012). *Social research methods* (4th ed.). Oxford: Oxford University Press. ISBN: 978 0199588053

Campbell, Donald T.; Stanley, Julian C. (1966). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin Company. ISBN: 978 0395307878.

Col·legi de Periodistes de Catalunya (2009). *Els gabinets de comunicació. Criteris de bones pràctiques professionals*. Barcelona: Col·legi de Periodistes de Catalunya. http://www.periodistes.org/fcic/admin/doc_contingut/MANUAL_BONES_PRACTIQUES.pdf

De-Bustos, Rosa (2012). "El gabinete de prensa ante el nuevo ecosistema informativo digital". *adComunica. Revista científica de estrategias, tendencias e innovación en comunicación*, n. 4, pp. 223-227.

<http://repositori.uji.es/xmlui/bitstream/handle/10234/53579/78-127-1-PB.pdf?sequence=1>
<http://dx.doi.org/10.6035/2174-0992.2012.4.14>

De-Semir, Vladimir; Ribas, Cristina; Revuelta, Gemma (1998). "Press releases of science journal articles and subsequent newspaper stories on the same topic". *Journal of the American Medical Association*, v. 280, n. 3, pp. 294-295. <http://jama.jamanetwork.com/article.aspx?articleid=187769>
<http://dx.doi.org/10.1001/jama.280.3.294>

Díaz-Noci, Javier (2002). *La escritura digital: hipertexto y construcción del discurso informativo en el periodismo electrónico*. Bilbao: Servicio Editorial de la Universidad del País Vasco. ISBN: 8483733781

Domínguez-Romero, Martí (2002). "Divulgar la investigación con revistas científicas: el caso de *Mètode*". *Comunicar. Revista científica de comunicación y educación*, v. 19, pp. 49-53. <http://www.revistacomunicar.com/indice/articulo.php?numero=19-2002-09>

Elías, Carlos (2000). *Flujos de información entre científicos y prensa*. Tesis doctoral, Universidad de La Laguna. <ftp://tesis.bbtk.ull.es/ccssyhum/cs194.pdf>

Elías, Carlos (2002). "Influencia de las revistas de impacto en el periodismo científico y en la ciencia actual". *Reis. Revista española de investigaciones sociológicas*, n. 98, pp. 123-137. <http://www.redalyc.org/articulo.oa?id=99717918007>

Elías, Carlos (2008). "El periodismo científico como paradigma de la 'noticia acatamiento'. Una demostración desde las fuentes y una alerta de sus peligros". *Periodística*, n. 11, pp. 81-93. <http://www.raco.cat/index.php/Periodistica/article/view/245699>

Escribà, Eudald; Cortiñas-Rovira, Sergi (2013). "Internationalization and coauthorship in major communication journals in Spain". *Comunicar*, v. XXI, n. 41. <http://dx.doi.org/10.3916/C41-2013-03>

Fernández-Asenjo, Gerardo; De-la-Torre-Alfaro, Noa (2009). *Gabinetes de comunicación en mínimos. Todos los porqués sobre las funciones y disfunciones de los gabinetes pequeños y con pocos recursos*. Madrid: Fragua. ISBN: 978 8470743375

Foster, John (2012). *Writing skills for public relations: style and technique for mainstream and social media*. London: Kogan Page Publishers. ISBN: 978 0749465438

Frost, Chris (2011). *Journalism ethics and regulation* (3rd ed.). Londres: Longman. ISBN: 978 1408244685

García-Orosa, Berta (2009). *Gabinetes de comunicación on line. Claves para generar información corporativa en la Red*. Sevilla / Zamora: Comunicación Social. ISBN: 978 84 96082 78 6

Krippendorff, Klaus (2004). *Content analysis. An introduction to its methodology* (2nd ed.). Thousand Oaks, California: Sage, 2004. ISBN: 0761915443

Loeffler, Robert H. (1993). *A guide to preparing cost-effective press releases*. Binghamton (New York): The Haworth Press, 1993. ISBN: 1560241411

Marshall, Catherine; Rossman, Gretchen B. (2011). *Designing qualitative research* (5th ed.). Thousand Oaks, California: Sage, 2011. ISBN: 978 1412970440

McCurry, Mike (2004). *Media relations handbook: for agencies, associations, nonprofits and congress*. Alexandria, Virginia (USA): The Capitol Net Inc. ISBN: 1587330032

Nieto, Josep; Vendrell, Montserrat (2001). "Els gabinets de premsa com a mediadors entre els científics i els mitjans". *Treballs de la Societat Catalana de Biologia*, v. 51, pp. 159-165. <http://revistes.iec.cat/index.php/TSCB/article/view/6557/60292>

Pander-Matt, Henk (2007). "How promotional language in press releases is dealt with by journalists. Genre mixing or genre conflict?". *Journal of business communication*, v. 44, n. 1, pp. 59-95. <http://dx.doi.org/10.1177/0021943606295780>

Pander-Maat, Henk; De-Jong, Caro (2012). "How newspaper journalists reframe product press release information". *Journalism*, v. 14, n. 3, pp. 348-371. <http://dx.doi.org/10.1177/1464884912448914>

Priest, Susanna-Hornig (2009). *Doing media research: an introduction*. Thousand Oaks, California, Sage. ISBN: 978 1412960977

Risquete, Jaume (2006). "Gabinets de premsa i de comunicació. L'ambigua frontera entre el periodisme i les relacions

públiques". *Anàlisi*, n. 34, pp. 209-221.

Ros-Martín, Marcos (2008). "Metodología para la implementación de un blog corporativo externo". *El profesional de la información*, v. 17, n. 5, pp. 502-509. <http://www.elprofesionaldelainformacion.com/contenidos/2008/septiembre/03.pdf> <http://dx.doi.org/10.3145/epi.2008.sep.03>

Rubin, Herbert J.; Rubin, Irene S. (1995). *Qualitative interviewing. The art of hearing data*. Thousand Oaks, California: Sage. ISBN: 978 0803950962

Ruiz-Olabuénaga, José I.; Aristegui, Iratxe; Melgosa, Leire (1998). *Cómo elaborar un proyecto de investigación social*. Bilbao: Universidad de Deusto. ISBN: 978 8474855470

Salaverría, Ramón (2005). *Redacción periodística en internet*. Pamplona: Ediciones Universidad de Navarra (Eunsa). ISBN: 978 8431322595

Sissons, Helen (2012). "Journalism and public relations: a tale of two discourses". *Discourse & communication*, v. 6, n. 3, pp. 273-294. <http://dx.doi.org/10.1177/1750481312452202>

Tashakkori, Abbas; Teddlie, Charles (2009). *Foundations of mixed methods research: integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, California: Sage. ISBN: 978 0761930112

Webster, Murray; Sell, Jane (2007). *Laboratory experiments in the social sciences*. Burlington, Massachusetts: Academic Press. ISBN: 978 0123694898

iralis®

International Registry for Authors:
Links to Identify Scientists

es:

- una **guía** para los autores hispanos para que firmen sus trabajos en el formato internacional usual
- una **base de datos** que registra las variantes de firma usadas por cada autor en diferentes épocas
- un **buscador** que usa automáticamente todas las variantes registradas

¡Ni un día más sin decidir tu firma!
Inscríbete en <http://iralis.org>

**El formato de
firma internacional**

Con el apoyo de:

El profesional de la
información





GESTIÓN DE DATOS DE INVESTIGACIÓN: INFRAESTRUCTURAS PARA SU DIFUSIÓN



Luis-Millán González, Tomás Saorín, Antonia Ferrer-Sapena, Rafael Aleixandre-Benavent y Fernanda Peset



Luis-Millán González, doctor por la *Universitat de València (UV)*, es profesor desde 1998 en el *Departamento de Educación Física* de la UV donde imparte las asignaturas de metodología de la investigación y ejercicio físico para la calidad de vida. Sus intereses son: gestión de los datos de investigación, arquitectura de la información, estadística, software documental y ejercicio físico en la patología músculo-esquelética. Ha publicado en las principales revistas de medicina del deporte indexadas en los *Journal Citation Reports*.

<http://orcid.org/0000-0002-6478-4014>

Universitat de València, Facultat de Ciències de l'Activitat Física i l'Esport (FCAFE)
Gascó Oliag, 3, 46010 Valencia, España
luis.m.gonzalez@uv.es



Tomás Saorín es profesor asociado en la *Facultad de Comunicación y Documentación* de la *Universidad de Murcia* y documentalista de la *Comunidad Autónoma de Murcia* donde ha participado en la puesta en marcha de proyectos de gestión de contenidos e información institucional en las áreas de servicios sociales, trabajo, empleo y bibliotecas. Participa en el capítulo español de *Wikimedia* para el conocimiento libre y en acciones de divulgación del movimiento *GLAM-Wiki* para bibliotecas, archivos y museos. Ha investigado sobre estrategia digital y de edición electrónica, así como sistemas de gestión de contenidos y proyectos colaborativos.

<http://orcid.org/0000-0001-9448-0866>

Servicio Regional de Empleo y Formación, Unidad de Documentación
Av. Infante Juan Manuel, 14. 30011 Murcia, España
tsp@um.es



Antonia Ferrer-Sapena es licenciada en geografía e historia contemporánea por la *Universidad de Valencia*, y doctora en técnicas y métodos de información y documentación por la *Universidad Politécnica de Valencia (UPV)*. Es coordinadora de investigación en *Florida Centre de Formació* y miembro de su Comité de Innovación. Entre los proyectos que dirige destaca el portal de economía social *EcSocial.com*. Es profesora de la UPV, y miembro del foro *Innovem Junts*, del *Instituto de la Pequeña y Mediana Industria (Impiva)* de la *Generalitat Valenciana*.

<http://www.ecsocial.com>

<http://orcid.org/0000-0001-6432-917X>

Universidad Politécnica de Valencia
Camino de Vera, s/n. 46022 Valencia, España
anfersa@upv.es



Rafael Aleixandre-Benavent es científico titular del *Consejo Superior de Investigaciones Científicas (CSIC)* y catedrático acreditado de biblioteconomía y documentación. Es doctor en medicina, especialista en documentación médica por la *Universitat de València (UV)* y codirector de la *Unidad de Información e Investigación Social y Sanitaria (Uisys)* de la UV. Sus principales líneas de trabajo son la evaluación de la investigación y de las publicaciones científicas y los estudios sobre el acceso abierto a los datos científicos

<http://orcid.org/0000-0002-6678-8844>

Instituto de Historia de la Medicina y de la Ciencia López Piñero
Plaza Cisneros, 4. 46003 Valencia, España
rafael.aleixandre@uv.es



Fernanda Peset es profesora titular de la *Universidad Politécnica de Valencia, Departamento de Comunicación Audiovisual, Documentación e Historia del Arte* y coordinadora del programa de doctorado. Es doctor por la *Universidad de Murcia* en 2002. Su docencia y publicaciones se orientan a comunicación científica, acceso abierto, implantación del protocolo *OAI-PMH*, normalización de la información, descripción de documentos, sistemas de documentación de museos, y datos abiertos y enlazados. Participa en proyectos como *IraLIS, E-LIS, Grupo CIEPI, ODiSEA* y dirige el proyecto *I+D Datasea*.

<http://orcid.org/0000-0003-3706-6532>

Universitat Politècnica de València
Camino de Vera, s/n. 46022 Valencia, España
mpesetm@upv.es

Artículo recibido el 04-02-2013

Aceptación definitiva: 19-06-2013

Resumen

El intercambio de datos de investigación es un tema candente debido a la expansión de los métodos de trabajo colaborativos con un uso intensivo de tecnología (e-ciencia). El volumen y el intercambio de datos electrónicos están aumentando con rapidez, y si no se tratan correctamente sus beneficios no llegarán a sus potenciales interesados (investigadores, gobiernos y organismos de financiación públicos, directores de investigación y bibliotecas, sector editorial...). Este trabajo revisa su aparición y los movimientos de acceso abierto, describe un modelo de ciclo de vida para la preservación de datos, y revisa las iniciativas que tratan sobre metodologías y tecnologías. De forma paralela se presentan los avances internacionales y las necesidades españolas.

Palabras clave

Datos de investigación, Intercambio de datos, Acceso abierto, E-ciencia, Bancos de datos, Preservación, Política científica.

Title: Managing and disseminating research data

Abstract

The sharing of research data is a hot topic at the international level, especially due to the development of e-science, based on collaborative working methods that use technology in an intensive manner. The amount of digital data and of its sharing is growing fast, and a lack of proper management could limit the potential benefits to stakeholders (researchers, governments and public funding agencies, research managers and libraries, and finally the publishing sector). The aim of this review is first to place the sharing of electronic research data in the context of its appearance and of open environments; secondly, it describes the life cycle model of data preservation; finally, it describes several initiatives dealing with methodologies and technologies. In parallel, international developments and highlights needs that exist in Spain are described.

Keywords

Research data, Data-sharing, Open access, E-science, Databank, Preservation, Scientific policy.

González, Luis-Millán; Saorín, Tomás; Ferrer-Sapena, Antonia; Aleixandre-Benavent, Rafael; Peset, Fernanda (2013). "Gestión de datos de investigación: infraestructuras para su difusión". *El profesional de la información*, septiembre-octubre, v. 22, n. 5, pp. 415-423.

<http://dx.doi.org/10.3145/epi.2013.sep.06>

Introducción¹

La preservación y el intercambio de datos científicos se han convertido en temas de interés a escala internacional para quienes financian, gestionan y generan la investigación, algo que ya recogían desde 2003 las pautas sobre acceso abierto a la información científica de los *National Institutes of Health (NIH)*. **Borrego** (2012), basándose en el informe *Opportunities for data exchange (ODE, 2011)*, cifra en un 60% el número de investigadores que desearían utilizar datos primarios ajenos, mientras que aproximadamente un 40% muestra problemas para compartir los suyos.

La investigación científica genera cada vez un volumen mayor de datos digitales. Algunos se usan en publicaciones pero buena parte permanecen ocultos (*Nesta, 2010*). De la encuesta *Parse.insight* (2009) se desprende que ha aumentado el volumen de datos que corren peligro de perderse cuando ha finalizado la financiación, ya que se almacenan a nivel departamental y si se comparten es por canales no formalizados como correo o webs.

Preservar los datos digitales procedentes de la investigación con una infraestructura adecuada es prioritario para la comunidad científica y para la sociedad. Muestra de ello es la financiación del *National Science Foundation's DataNet Program* de EUA a grandes proyectos de investigación en esta área.

Existen multitud de tipos de datos digitales utilizados en los procesos de investigación: publicaciones, *raw data* (datos

brutos, sin tratar), estadísticas, gráficos, vídeos... El *W3Consortium* aconseja la utilización de formatos estándar (csv, xls, xml) para poder compartirlos con mayor facilidad. Su almacenamiento necesita instrucciones exactas, ya que no son igual de reutilizables datos que ya han sido analizados, que los directamente salidos de los instrumentos, por ejemplo.

El intercambio de resultados publicados y de datos de las investigaciones (*data sharing*) estimula descubrimientos adicionales.

La gestión de los datos de investigación es insuficiente tanto en el ámbito internacional como, especialmente, en el

Datos de investigación

La definición que tiene más consenso, al haber sido adoptada por entidades como los *National Institutes of Health (NIH)* de EUA y la *OECD* (2007), es "todo el material registrado durante el proceso investigador, reconocido por la comunidad científica y que sirve para certificar los resultados de la investigación que se realiza". Deben provenir de una fuente única y ser difíciles o imposibles de obtener de nuevo (**Torres-Salinas; Robinson-García; Cabezas-Clavijo, 2012**).

nacional. El volumen de datos electrónicos y su intercambio por métodos informales aumentan con rapidez y se corre el riesgo de perderlos al acabar la financiación de los proyectos. En España no se ha investigado sistemáticamente de qué situación partimos. La *Ley 14/2011* de 1 de junio de la ciencia, la tecnología y la innovación, que obliga a depositar las publicaciones en abierto, no incluye mención sobre los datos científicos. Este tema apenas está tratado por universidades y organismos públicos de investigación (OPI), que son quienes producen más datos.

Ante esta situación uno de los mayores retos es la investigación de nuevos métodos, tecnologías y modelos de gestión que sean capaces de identificar, clasificar, preservar y distribuir los conjuntos de datos (*datasets*), así como los flujos y ciclos de vida desde que son generados hasta que son reutilizados.

Este trabajo es una revisión de algunas tendencias internacionales que pueden ser de referencia para España. Existen cuestiones que deben ser dirimidas y armonizadas a escala internacional, como modelos, estándares y tendencias, mientras que otras están determinadas por el contexto en que se producen, como mandatos, agentes involucrados, estados de opinión y conductas, o valorización de los datos. Por esta razón se revisa el contexto internacional y se proponen aspectos a indagar en España.

Desde 2003 las pautas sobre acceso abierto de los *National Institutes of Health* contemplan los datos científicos además de la publicación de los resultados

En primer lugar se sitúa el objeto de estudio en el contexto de su aparición, los movimientos de acceso abierto; en segundo lugar se describe un modelo de ciclo de vida para su preservación; y por último se revisan métodos y técnicas aplicables. Asimismo se da cuenta de los avances internacionales y de las necesidades españolas.

Movimientos de acceso abierto

La investigación científica ha evolucionado en los últimos años a causa de varios fenómenos vinculados a la e-ciencia y los movimientos *open*. Se observa que las formas de investigación han cambiado (*The Royal Society*, 2012). El trabajo colaborativo -que ya existía por ejemplo en genómica- avanza hacia la e-ciencia, hacia la investigación facilitada por las infraestructuras electrónicas/TIC, en las que el intercambio de datos es esencial (*Shneiderman*, 2008). Un buen ejemplo es el desarrollo de vacunas contra la gripe porcina (H1N1),

http://grants.nih.gov/grants/policy/data_sharing

que se produjo en tan sólo tres meses, frente al año que había necesitado la vacuna contra el *SARS coronavirus* diez años antes.

Existe una brecha entre el deseo de utilizar los datos ajenos (60%) y el de compartir los propios (40%)

La *Agenda digital europea 2003* abogaba ya por una mayor reutilización de los datos gubernamentales, pero también del patrimonio cultural y de investigación, como complemento del acceso abierto a las publicaciones (*Comisión Europea*, 2011b; 2012; *Riding the wave*, 2010).

Preservación de datos

Preservación digital es el “proceso de establecer y mantener a largo plazo repositorios digitales de referencia capaces de salvaguardar los datos actuales y futuros” (*Higgins*, 2008). Otros la definen como “la administración de datos científicos digitales a los que se proporciona un valor añadido dándoles un contexto y una vinculación a su origen de manera que se facilite su reutilización, integración y promoción” (*Rusbridge et al.*, 2005). Para la preservación y la reutilización de datos no sólo hay que resolver aspectos técnicos, sino también de organización y de procedimiento, económicos, financieros y de personal, administración de su propiedad, obligaciones legales, requisitos de auditoría, restricciones sociales de uso, etc. Este conjunto de acciones configura lo que se denomina *data curation*, actividad que recibió un gran impulso con la creación en marzo de 2004 del *Digital Curation Centre (DCC)* en Gran Bretaña, con financiación del *Joint Information Systems Committee (JISC)*. Actualmente es un centro de referencia mundial, siendo sus propósitos crear estándares y ayudar en la generación de planes de gestión (*Borgman*, 2012).

Las bibliotecas pueden asumir una importante responsabilidad en la preservación, pues están especializadas en conocer las fuentes de información digital y los métodos y estándares para su almacenamiento y conservación (*Ogburn*,

Tabla 1. Iniciativas de carácter general

País Institución	Proyecto
Reino Unido JISC	-DCC (<i>Data Curation Centre</i>), 2004. -Ukrds (<i>United Kingdom Research Data Service</i>), 2007-2010. Estableció un modelo sostenible de costes. -DAF (<i>Data Asset Framework</i>), 2009. Creó una metodología para asesorar a quienes han de gestionar este tipo de información.
Unión Europea (Reino Unido, Holanda, Alemania, Francia, además de Suiza). FP7	<i>Parse.Insight (Permanent access to the records of science in Europe)</i> , 2008-2010. Se ocupó de la preservación de la ciencia (datos primarios y publicaciones). Llevó a cabo estudios en profundidad para diseñar una hoja de ruta que orientase la estrategia de la <i>Comisión Europea</i> sobre infraestructura técnica. Resultados de la encuesta: -Los investigadores consideran importante un reanálisis de los datos ajenos. -Los gestores ven necesario disponer en abierto los datos obtenidos con financiación pública -Los editores creen que poner los datos en abierto hará posible el avance científico
Unión Europea. FP7	Los informes de ODE (<i>Kotarski et al., 2012; Dallmeier-Tiessen et al., 2012</i>) recogen buenas prácticas y opiniones de legisladores, financiadores, operadores de infraestructuras, centros de datos, proveedores y usuarios de datos, bibliotecas y editoriales.
Unión Europea. FP7	<i>OpenAirePlus (Open access infrastructure for research Europe)</i> . Toma como base las publicaciones enriquecidas: <i>Enhanced publications</i>
Unión Europea. FP7	<i>DataCite</i> es una organización británica sin ánimo de lucro que desde 2009 asesora para que se reconozcan como contribuciones científicas los datasets compartidos. Ha implementado un esquema de metadatos (<i>DataCite</i> , 2009) basado en DOI.
Estados Unidos. <i>National Science Foundation</i>	Las agencias de financiación tienen un papel relevante en la preservación y sostenibilidad de los datos (<i>Keefer</i> , 2011). Ello repercute en las bibliotecas académicas estadounidenses, que ayudan a elaborar los planes que exige la NSF en su apartado <i>Sharing of findings, data, and other research products</i> y formularios como el <i>Data conservancy</i> .

2010). Fuera de España algunas bibliotecas de investigación ya están desarrollando estos servicios de apoyo (**Choudhury**, 2009; **Gold**, 2010).

En los últimos años la atención se ha desplazado hacia la gestión del material según su ciclo de vida: creación, intercambio y preservación. Algunos objetos tienen una historia larga y complicada, pero tener en cuenta la “continuidad del ciclo de vida” mejora su longevidad y sus perspectivas de reutilización (**Higgins**, 2008). Para ello se registra la información relevante a lo largo del tiempo y los departamentos o secciones responsables de ese objeto en particular. Por ejemplo, en una institución de educación superior los profesores crean objetos digitales, pero es posible que ciertos detalles sobre los autores residan en otra unidad de gestión, que los usuarios del objeto estén registrados previamente en otra diferente o que se envíe una copia del objeto a la biblioteca, archivo o *digital curation centre* para su preservación.

El ciclo de vida de *digital curation* sigue estos pasos:

- conceptualización del plan para la fase de creación de datos;
- creación de objetos digitales con los metadatos adecuados;
- garantía de fácil acceso por los usuarios;
- selección de objetos a conservar, respetando sus dere-

chos y asegurando las copias;

- plan de expurgo y destrucción segura del resto de objetos;
- ingesta de contenidos en el repositorio;
- preservación;
- reevaluación de los objetos que no cumplen los procedimientos;
- almacenamiento: mantenimiento de los datos de manera segura;
- garantía del acceso y reutilización de los datos públicos;
- transformación en nuevos objetos digitales a partir de los originales.

Estas actividades han estado tradicionalmente vinculadas a las bibliotecas —recolección, producción, gestión y conservación— por lo que sus profesionales podrían evolucionar hacia los roles de *data manager* o *data curator*. Sin embargo el camino no es fácil, pues sus conocimientos han de completarse con aspectos muy específicos sobre la creación científica y la transformación de datos.

En cualquier caso el elemento clave para facilitar el acceso y movimiento de datos son los metadatos, en particular los de preservación, pues informan, describen y registran este tipo de actividades. Actualmente existen varias posibilidades para realizar estas funciones (**Martínez-Urbe**; **Macdonald**,

Tabla 2. Proyectos por disciplinas

Tema	Proyecto
Científico y cultural	<i>Caspar (Cultural, artistic and scientific knowledge preservation, for access and retrieval)</i> . Destaca la dificultad de encontrar un marco común para todo el ciclo de vida de una información que es muy heterogénea: científica, artística y cultural; sin embargo, el modelo de preservación <i>Oais</i> lo hizo posible.
Ciencias físicas, biomédicas, sociales y humanidades	<i>KE (Knowledge exchange)</i> recoge las buenas prácticas de los gobiernos británico, alemán, holandés y danés. Se focaliza en la necesidad de fomentar el intercambio de datos, facilitar programas de formación y estudiar la infraestructura necesaria.

2008; Educause, 2013): DDI 3.0 u *Open archival information system (OAIS) (ISO 14721:2003)*. La opción más utilizada es *OAIS*, un marco de referencia que define las responsabilidades obligatorias de un archivo digital y expone las actividades y relaciones básicas que deben llevarse a cabo para que el sistema funcione.

Para la puesta en valor de los datos se necesitan sin embargo otro tipo de ontologías destinadas a describir el ámbito disciplinar, cómo se han tomado los datos o el formato en que se encuentran.

Infraestructuras para difusión de datos

En relación con las cuestiones revisadas en los apartados precedentes se pueden identificar diversos tipos de personas, iniciativas e infraestructuras. Hay tres grupos de personas:

- científicos, que son productores y consumidores de los datos de investigación;
- gobiernos y agencias públicas, que financian el entramado de la ciencia, si bien en este grupo se incluyen en numerosas ocasiones las bibliotecas y servicios de investigación de los OPIs;
- sector editorial científico que hace pública la ciencia.

En las iniciativas se puede distinguir:

- portales orientados a estudiar métodos y a asesorar;
- tecnología para compartir datos y directorios que recogen los bancos de datos.

Metodologías

En este campo las iniciativas son numerosas y generalmente previas a establecer las infraestructuras técnicas. Publican metodologías y protocolos para un campo científico o se constituyen en plataformas internacionales con servicios de asesoramiento (tabla 1).

Por otra parte, existen proyectos que se limitan a disciplinas (tabla 2).

También las sociedades científicas adquieren protagonismo en tanto que productoras de datos. Por ejemplo la *International Spinal Cord Society (ISCoS)* recomienda el modelo *Common data elements* del *National Institute of Neurological Disorders and Stroke (Ninds)*, de los *NIH*, su agencia natural de financiación. Por otra parte, la *American Psychological Association (APA)* enlaza multitud de políticas y repositorios autorizados para distribuir los datos.

En cuanto a las revistas, el análisis de las políticas de las editoriales sobre datos muestra una correlación entre índices de impacto altos y el hecho de que las revistas acepten y hagan disponibles *datasets*. Se ha estudiado para el área de salud (Piwowar; Chapman, 2008; Piwowar, 2011) y para todas las disciplinas, escogiendo las revistas más citadas (Alsheikh-Ali et al., 2011; García-García et al., 2012a). Qui-

<http://www.dcc.ac.uk>

zá sería necesario realizar un estudio sistemático de estas políticas, teniendo en cuenta que el papel de las editoriales se considera crucial en el intercambio de datos (Dallmeier-Tiessen et al., 2012).

Como vemos, existen multitud de actores para un escenario que está en plena ebullición. En España el grupo de trabajo sobre datos de *Recolecta* ha publicado un informe preliminar (*Grupo de trabajo*, 2012) que ofrece un panorama de la gestión de datos a escala internacional y española. Simultáneamente el *Plan Nacional de I+D+i* ha financiado en su convocatoria de 2012 el proyecto *Opendatascience, centro de recursos para la preservación y gestión de datos abiertos de investigación* (CSO2012-39632). También las fundaciones privadas están interesadas en investigar en este campo y, por ejemplo, la *Fundación Mapfre* subvenciona el proyecto *Estrategias para la gestión y la promoción del uso compartido de datos científicos sobre la salud*. Ambos proyectos están siendo desarrollados por las tres instituciones del *VLC/Campus-Valencia Campus de Excelencia Internacional: Universidad de Valencia, Universidad Politécnica de Valencia y CSIC*.

“ Las políticas sobre los datos científicos no tienen únicamente un carácter técnico ”

Tecnologías

Existen múltiples formas de compartir datos, más o menos formalizadas, pero las predominantes son las no registradas o dispersas, como se deduce de los escasos y heterogéneos casos de éxito que recogen **Torres-Salinas, Robinson-García y Cabezas-Clavijo** (2012): *ProteinDataBank* y *GenBank*.

En cuanto a la difusión formalizada, pueden adoptarse varias vías que casi coinciden con las conocidas ruta verde y dorada para fomentar el libre acceso a las publicaciones, aunque no usan el protocolo *OAI-PMH* para cosechar los

Tabla 3. Iniciativas relacionadas con el depósito de los datos

Clasificación	Nombre	Descripción
Servicios nacionales	<i>ANDS (Australian National Data Service)</i>	Objetivo: publicar, descubrir, acceder y reutilizar datos de investigación. <i>Research Data Australia</i> es un inventario y <i>My Data</i> un servicio de citación basado en DOI.
	<i>SND (The Swedish National Data Service)</i>	Coordina las bases de datos y guía a los científicos durante su investigación y en el acceso a datos externos
Infraestructuras temáticas	<i>Seadatanet (2007). Unión Europea</i>	Sistema interoperable para la gestión de datos e información marina
	<i>IDIP-ES (Science data infrastructure for preservation - Earth science), 2011. Unión Europea</i>	Infraestructura de datos sobre ciencias de la tierra
Servicios generalistas centrados en los contenidos	<i>Figshare o Dryad</i>	Permiten publicar los datos de investigación en abierto, facilitando su intercambio, cita y recuperación
	<i>Dataverse. Harvard University (EUA)</i>	Aplicación de código abierto para publicar, compartir, citar, extraer y analizar datos de investigación. Garantiza el uso de protocolos estandarizados y el reconocimiento de la autoría de los datos. Se organiza por colecciones creadas por cada institución que usa la plataforma.

metadatos. **Borrego (2012)**, siguiendo *ODE*, menciona repositorios y editoriales como las vías de almacenamiento preferidas por los investigadores. La ruta verde sería el depósito de *datasets* en repositorios o bancos específicos por disciplinas; y la dorada, almacenarlos en las plataformas editoriales junto a la publicación.

El concepto de continuidad del ciclo de vida mejora la longevidad y la reutilización de los datos

Torres-Salinas, Robinson-García y Cabezas-Clavijo (2012) siguiendo a **Kowalczyk (2011)** clasifican los repositorios y bancos de datos en centralizados por institución, descentralizados de varias instituciones, federados con una plataforma común de consulta y ciberestructuras. Nuestra propia indagación nos conduce a una clasificación más práctica:

- servicios nacionales de datos, con el caso australiano a la cabeza;
- plataformas por disciplinas; y
- plataformas de interés general, orientadas al investigador, preocupado especialmente por no perder el reconocimiento de su trabajo (**Starr; Gastl, 2011**).

Se prevén aún más líneas de actuación de cara al lanzamiento de *Horizon 2020* en la UE, el nuevo programa marco de investigación e innovación que integrará todas las líneas de investigación de la *Comisión Europea*, incluido el programa marco vigente todavía.

La tabla 3 muestra algunos ejemplos de lo que se puede llamar vía verde.

En cuanto a la vía que hemos comparado con la ruta dorada -el almacenamiento de datos en las editoriales- obviamos describir la tecnología utilizada, ya que se trata de las mismas plataformas de las publicaciones. Suele ser tecnología

propietaria en el caso de editoriales convencionales como *Sciverse* y *Wiley-Blackwell (Navarro-Molina, 2012)* o sistemas *OAI-PMH* compatibles como *Open Journal System (OJS)* o *Biomed Central* para el caso de las revistas *open access*.

Por último *Thomson Reuters* ha entrado en escena pisando fuerte con el lanzamiento del *Data citation index* en la *Web of knowledge*. Gracias a este producto, los resultados de las búsquedas mostrarán los *associated records* si la publicación tiene archivados datos en uno de los repositorios indexados (69 en enero de 2013).

Las infraestructuras que recopilan *datasets* pueden ser muy variadas en muchos sentidos -modelo, disciplina, tipos de datos, grado de reutilización, tecnología-. En la tabla 4 se listan algunos directorios.

Se buscaron repositorios de datos españoles en *DataBib* y *ODiSEA (International registry on research data)*. En el primero no aparece ninguno, y en el segundo dos (entre los 183 registrados a noviembre de 2012): *Ceacs*, de la *Fundación Juan March*, y *Digital.CSIC*; la *Universitat Pompeu Fabra* también han dado sus primeros pasos.

Según *OpenDoar*, sólo 82 repositorios ofrecen *datasets* en todo el mundo (3,6%). De los 97 repositorios españoles, que suponen el 4% del total, sólo uno acepta *datasets (Digital.CSIC)* (1%).

Por otra parte, suponiendo que los investigadores españoles estén depositando sus datos en bancos internacionales, se ejecutaron búsquedas en *ODiSEA* por los términos *Spain* y *Spanish*³: unos 20 bancos (11%) ofrecían conjuntos de datos españoles.

Conclusiones

Debido a la irrupción de los movimientos de acceso abierto en todos los ámbitos de la sociedad, el sector de la gestión de la información está en un momento de cambio, al que se suma las normativas que fomentan la reutilización de los datos para darles nuevo valor social y económico. El acceso

Tabla 4. Directorios de conjuntos de datos

Directorio	Detalles
<i>Research Data Australia</i> http://researchdata.andis.org.au	Carácter nacional
<i>CKAN/DataHUB</i> http://datahub.io	Múltiples recursos también relacionados con <i>opendata</i> y <i>linkedopendata</i>
<i>Dataverse</i> http://thedata.org	Plataforma de colecciones
<i>ODiSEA</i> http://odisea.ciepi.org	Incluye desde 2011 únicamente las bases de datos sobre datos de investigación (García-García <i>et al.</i> , 2012b y 2012c). Ante la falta de un inventario mundial de datos, similar a los de repositorios de publicaciones DOAR o ROAR, se seleccionaron aproximadamente 200 bancos.
<i>DataBib</i> http://databib.org	Inventario mundial colaborativo que incluye casi 600 items, entre los cuales aparecen también webs

a datos digitales, con los beneficios futuros que se deriven de ello, depende de las estrategias y procedimientos de preservación digital que se adopten ahora y del respaldo que reciban de las autoridades.

No obstante, entendemos que los organismos públicos de investigación y la universidad tienen características y funciones muy diferentes a las del resto de la administración pública, pues los datos que generan son la materia prima para su trabajo.

Es necesario conocer aspectos como:

- quiénes son los agentes involucrados en el flujo de los datos de investigación;
- cuáles son sus prácticas, necesidades y límites a la hora de compartir sus datos;
- qué datos se están generando, registrándolos y calificando especialmente su grado de apertura;

El trabajo de promoción y asesoría, al modo de bibliotecas como las del MIT, *Purdue University* o *California University*, es una oportunidad para nuestro sector. Unirlo a las estrategias que se están llevando a cabo para los repositorios institucionales valorizará esta información, hasta ahora escondida en los discos de los investigadores.

La complejidad de un escenario con tantos actores obliga a incluir a “infomediadores” que se encarguen de valorizar la información de cara al usuario final. El trabajo de promoción o *advocacy*, tan necesario, ha de ser liderado desde varios frentes. Minoritariamente incluirá al sector bibliotecario, junto a otros agentes relacionados con la investigación, como las oficinas de transferencia de resultados de la investigación (OTRI). Esos frentes habrán de constituirse en una red de influencia con el objetivo de contribuir a modificar los hábitos de la producción, tratamiento y difusión de los datos de investigación.

Notas

1. Este trabajo recibe financiación del Plan Nacional I+D+i del *Ministerio de Economía y Competitividad (Mineco)*: *Opendatascience*, Centro de recursos para la preservación y gestión de datos abiertos de investigación (CSO2012-39632), y de la *Fundación Mapfre*, convocatoria 2012 de ayudas a la investigación

http://www.fundacionmapfre.org/fundacion/es_es/default.jsp

2. Para escribir este artículo se ha consultado *Sciencea.com*, que ofrece servicios de valor añadido sobre los datos de *Cordis*.
<http://www.sciencea.com>
http://cordis.europa.eu/home_es.html

3. Los autores agradecen la ayuda de **Florencia Dieci**, colaboradora de *ODiSEA*.

Bibliografía

Alsheikh-Ali, Alawi A.; Qureshi, Waqas; Al-Mallah, Mouaz H.; Ioannidis, John P. A. (2011). “Public availability of published research data in high-impact journals”. *PLoS one*, v. 6, n. 9, pp. e24357.
<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0024357>

ANDS. *Australian National Data Service*.
<http://www.andis.org.au>

APA. *Links to data sets and repositories*.
<http://www.apa.org/research/responsible/data-links.aspx>

Arano, Silvia; Martínez, Gemma; Losada, Marina; Villegas, Marta; Casaldàliga, Anna; Bel, Núria (2011). “La comunidad ‘Recursos y datos primarios’ de la *Universitat Pompeu Fabra*: los repositorios institucionales como infraestructuras científicas. Estudio de caso”. *Revista española de documentación científica*, v. 34, n. 3, pp. 385-407.
<http://goo.gl/sfpnN>
<http://dx.doi.org/10.3989/redc.2011.3.834>

Borgman, Christine L. (2012). “The conundrum of sharing research data”. *Journal of the American Society for Information Science and Technology*, v. 63, n. 3, pp. 1059-1078.
<http://dx.doi.org/10.1002/asi.22634>

Borrego, Àngel (2012). “Los retos de la gestión de datos de investigación”. *Blok de bid*, 6 nov.
<http://goo.gl/10h9S>

Caspar. *Cultural artistic and scientific knowledge preservation for access and retrieval*.
<http://www.casparpreserves.eu>

Choudhury, Sayeed (2009). *Rethinking scholarly communi-*

cation: building data curation infrastructure.

<http://www.it.utah.edu/leadership/research/ciday/2009/notes/choudhury.pdf>

Comisión Europea (2003). *Agenda digital europea.*

http://europa.eu/rapid/press-release_IP-11-1524_es.htm

Comisión Europea (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data.* Final report of the High Level Expert Group on Scientific Data.

<http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>

Comisión Europea (2011a). *Datos abiertos: Un motor para la innovación, el crecimiento y la gobernanza transparente.*

<http://goo.gl/4n5Wpw>

Comisión Europea (2011b). *Propuesta de directiva del Parlamento Europeo y del Consejo por la que se modifica la Directiva 2003/98CE relativa a la reutilización de la información del sector público.*

<http://goo.gl/9iVHtH>

Comisión Europea (2012). "Dictamen del Comité de las Regiones — Revisión de la Directiva relativa a la reutilización de la información del sector público y los datos abiertos (2012/C 391/12)". *Diario oficial de la Unión Europea*, 18.12.2012.

<http://goo.gl/ef3nE>

Comisión Europea (2012). *Online survey on scientific information in the digital age.* Luxembourg: Publications Office of the European Union.

<http://goo.gl/jVX0k>

<http://dx.doi.org/10.2777/7549>

Dallmeier-Tiessen, Sunje; Darby, Robert; Gitmans, Kathrin; Lambert, Simon; Suhonen, Jari; Wilson, Michael (2012). *Compilation of results on drivers and barriers and new opportunities.*

<http://goo.gl/ChTK6>

Data asset framework (2009). *Implementation guide.* Londres: JISC.

<http://www.data-audit.eu>

Data citation index.

http://wokinfo.com/products_tools/multidisciplinary/dci

DataCite (2009). *Metadata schema for the publication and citation of research data. Version 2.2.*

<http://goo.gl/AD4sgs>

<http://dx.doi.org/10.5438/0005>

DataConservancy. *Leading the movement to build data management tools and services across institutions and disciplines.*

<http://dataconservancy.org>

Dataverse network project.

<http://thedata.org>

DCC digital curation lifecycle.

<http://www.dcc.ac.uk/digital-curation/what-digital-curation>

DDI 3.0

<http://www.ddialliance.org>

Educause (2013). *Developing an institutional research data management plan service.*

<http://net.educause.edu/ir/library/pdf/ACT11301.pdf>

Figshare.

<http://figshare.com>

García-García, Alicia; García-Massó, Xavier; Ferrer-Sapena, Antonia; González, Luis-Millán; Peset, Fernanda; Aleixandre-Benavent, Rafael (2012a). "Mejores prácticas en reuso de conjuntos de datos publicados online como material adicional a los artículos". En: *2ª Conferencia sobre calidad de revistas de ciencias sociales y humanidades (CRECS 2012)*, Valencia.

<http://www.thinkepi.net/crecs2012>

García-García, Alicia; García-Massó, Xavier; Ferrer-Sapena, Antonia; González, Luis-Millán; Peset, Fernanda; Rodríguez-Gairín, Josep-Manuel; Saorín, Tomás (2012b).

"ODiSEA: International registry on research data". En: *5ª Jornadas OS-repositorios "La motricidad de los repositorios de acceso abierto"*, 23-25 mayo 2012.

<http://e-spacio.uned.es/ocs/index.php/osrepositoriosBilbao/motricidad/paper/view/15>

García-García, Alicia; Rodríguez-Gairín, Josep-Manuel; Saorín, Tomás; González, Luis-Millán; García-Massó, Xavier; Ferrer-Sapena, Antonia; Peset, Fernanda (2012c) "ODiSEA: International registry on research data". *BiD: textos universitaris de biblioteconomia i documentació*, n. 29.

<http://www.ub.edu/bid/29/garcia2.htm>

Gold, Anna (2010). *Data curation and libraries: short-term developments, long-term prospects.* April, 4.

http://digitalcommons.calpoly.edu/lib_dean/27

Grupo de Trabajo *Depósito y Gestión de datos en Acceso Abierto* del proyecto *Recolecta* (2012). *Informe preliminar para la conservación y reutilización de los datos científicos en España*. Madrid: Fundación Española para la Ciencia y la Tecnología (Fecyt).

<http://www.fecyt.es>

Higgins, Sarah (2008). "The DCC curation lifecycle model". *The international journal of digital curation*, v. 3, n. 1, pp 134-140.

<http://goo.gl/WTlgt>

ISCOs. International SCI data sets.

<http://www.iscos.org.uk/international-sci-data-sets>

Keefer, Alice (2011). "La preservación de los datos de investigación y las agencias de financiación de la I+D". *Blok de bid*, 11 de octubre.

<http://www.ub.edu/blokdebid/es/node/130>

Knowledge Exchange-KE

<http://www.knowledge-exchange.info>

Kotarski, Rachael; Reilly, Susan; Schrimpf, Sabine; Smit, Eefke; Walshe, Karen (2012). *Report on best practices for citability of data and on evolving roles in scholarly communication. Executive summary.*

<http://goo.gl/6qktq>

Kowalczyk, Stacy; Shankar, Kalpana (2011). "Data sharing in sciences". *Annual review of information science and technology*, n. 45, pp. 247-294.

- Martínez-Urbe, Luis; Macdonald, Stuart** (2008). "Un nuevo cometido para los bibliotecarios académicos: data curation". *El profesional de la información*, v. 17, n. 3, pp. 273-280.
<http://eprints.rclis.org/19304>
<http://dx.doi.org/10.3145/epi.2008.may.03>
- Martínez-Urbe, Luis; Macdonald, Stuart** (2009). "User engagement in research data curation". En: *ECDL'09 Proceedings of the 13th European conference on research and advanced technology for digital libraries*.
<http://www.era.lib.ed.ac.uk/handle/1842/3206>
- Navarro-Molina, Carolina** (2012). *Propuesta de criterios de calidad formal para revistas científicas electrónicas basadas en aspectos de interacción persona-ordenador*. Tesis doctoral. Universidad de Valencia.
<http://roderic.uv.es/handle/10550/23755>
- Nesta (National Endowment for Science Technology and the Arts) (2009 - 2010). *Open science case studies*.
<http://www.rin.ac.uk/our-work/data-management-and-curation/open-science-case-studies>
- NIH (2003a). *Data sharing faqs*.
http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm
- NIH (2003b). *Final NIH statement on sharing research data*.
<http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>
- NIH. Office of Extramural Research. *NIH Data sharing policy*.
http://grants1.nih.gov/grants/policy/data_sharing
- Ninds *Common data elements CDE*.
<http://www.commondataelements.ninds.nih.gov>
- Ninds (2012). *Maximizing what we learn from clinical trials: Ninds common data elements*.
http://www.ninds.nih.gov/about_ninds/message/message-CDE.htm
- NSF (2009). *Grant general conditions (GC-1)*.
<http://www.nsf.gov/pubs/gc1/jan09.pdf>
- NSF. Directorate for Social, Behavioral and Economic Sciences. *Data archiving policy*.
<http://www.nsf.gov/sbe/ses/common/archive.jsp>
- ODE (2010). *Opportunities for data exchange*.
<http://www.alliancepermanentaccess.org/index.php/community/current-projects/ode>
- OECD (2007). *Principles and guidelines for access to research data from public funding*.
<http://www.oecd.org/science/scienceandtechnologypolicy/38500813.pdf>
- Ogburn, Joyce L.** (2010). "The imperative for data curation". *Portal-libraries and the academy*, v. 10, n. 2, pp. 241-246.
<http://dx.doi.org/10.1353/pla.0.0100>
- Parse.insight (2009). *Insight into digital preservation of research output in Europe. Survey report*.
http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf
- Piwowar, Heather A.** (2011). "Who shares? Who doesn't? Factors associated with openly archiving raw research data". *PLoS one*, v. 6, n. 7, p. e18657.
<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0018657>
- Piwowar, Heather A.; Chapman, Wendy W.** (2008). "A review of journal policies for sharing research data". *AMIA Annual symposium proceedings archive*.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655927/?tool=pubmed>
- Rusbridge, Chris; Burnhill, Peter; Ross, Seamus et al.** (2005). "The digital curation centre: a vision for digital curation". En: *Proceedings from local to global: data interoperability - challenges and technologies*, 2005. Forte Village Resort, Sardinia, Italy, pp. 1-11.
<http://eprints.erpanet.org/82>
- Scidip-es. *Science data infrastructure for preservation - Earth science*.
<http://www.scidip-es.eu>
- The Royal Society (2012). *Science as an open enterprise*. The Royal Society Science Policy Centre report. ISBN: 978 0 85403 962 3
<http://royalsociety.org/policy/projects/science-public-enterprise/report>
- Seadatanet.
<http://www.seadatanet.org>
- Shneiderman, Ben** (2008). "Science 2.0". *Science*, v. 319, n. 5868, pp. 1349-1350.
<http://dx.doi.org/10.1126/science.1153539>
- Starr, Joan; Gastl, Angela** (2011). "isCitedBy: a metadata scheme for DataCite". *D-lib magazine*, v. 17, n. 1-2
<http://dlib.org/dlib/january11/starr/01starr.html>
<http://dx.doi.org/10.1045/january2011-starr>
- SND. Swedish National Data Service.
<http://snd.gu.se/en>
- Torres-Salinas, Daniel** (2010a). "Compartir datos (data sharing) en ciencia: el contexto de una oportunidad". *Anuario ThinkEPI*, v. 4, pp. 258-261.
- Torres-Salinas, Daniel** (2010b). "Primeros pasos hacia la gestión de datos de investigación en las universidades: la iniciativa DAF". *Anuario ThinkEPI*, v. 4, pp. 262-265.
- Torres-Salinas, Daniel; Robinson-García, Nicolás; Cabezas-Clavijo, Álvaro** (2012). "Compartir los datos de investigación: introducción al data sharing". *El profesional de la información*, v. 21, n. 2, pp. 173-184.
<http://ec3.ugr.es/publicaciones/datasharing.pdf>
<http://dx.doi.org/10.3145/epi.2012.mar.08>

CCD

Centro de Competencia Documental

Ayudamos a nuestros clientes a generar oportunidades en sus negocios para mejorar los ámbitos

- documental y de la información
- de los activos digitales

Contribuimos a

- reducir costes, y a mejorar la eficiencia y la calidad
- mejorar los procesos, y a la reutilización y explotación eficiente de los activos digitales

Proporcionamos soporte y asesoramiento operativo, funcional y estratégico en todos los ámbitos

- de la documentación, información y del conocimiento
- de la gestión de los activos digitales

CCS

Centro de Competencia de Servicios

Nuestro personal experto, nuestra metodología y los medios tecnológicos, permiten gestionar todos los procesos documentales y de información

En nuestro equipo, disponemos de Perfiles expertos en

- procesos de gestión de la información y de los contenidos
- procesos de soporte documental
- procesos de gestión de contenidos digitales
- procesos de gestión documental

doc 6

consultores en recursos de información

www.doc6.es

Gestión
de la
información
y del
conocimiento

CCT

Centro de Competencia Tecnológica

Soluciones tecnológicas para la gestión de

- centros de documentación
- unidades de marketing y centros de comunicación
- community managers
- bibliotecas, archivos y museos
- fondos documentales de valor, ...

Soluciones basadas en

- productos distribuidos en exclusiva por doc6: Inmagic, OCLC, Lexis-Nexis
- productos y servicios desarrollados a medida por nuestro equipo técnico, según las necesidades de los clientes





GESTIÓN DE CONTENIDOS CON *DRUPAL*: REVISIÓN DE MÓDULOS ESPECÍFICOS PARA BIBLIOTECAS, ARCHIVOS Y MUSEOS



Jesús Tramullas



Jesús Tramullas es profesor titular en el *Departamento de Ciencias de la Documentación* de la *Universidad de Zaragoza*. Socio senior de *ACM e IEEE*; *ISKO*, *Anabad* y *Asociación Española de Recuperación de Información (SERI)*. Miembro de *Drupal Association*. Es investigador principal de proyectos de investigación financiados sobre web semántica, web social y servicios de información digital, y miembro del grupo de investigación *Intelligent Networks and Information Technologies (iNiT)*, reconocido por el Gobierno de Aragón. Sus líneas de investigación se centran en el diseño e implementación de servicios de información digital, gestión de contenidos, comportamiento informacional, bibliotecas digitales y herramientas de software libre para la gestión de información. <http://orcid.org/0000-0002-5374-9993>

Universidad de Zaragoza, Departamento de Ciencias de la Documentación
Pedro Cerbuna, 12. 50009 Zaragoza, España
<http://tramullas.com>
tramullas@unizar.es

Resumen

Drupal es una de las plataformas más avanzadas de gestión de contenidos. Para la elaboración de productos especializados, la arquitectura de *Drupal* utiliza módulos que añaden funciones específicas a su núcleo. Existe más de una veintena de módulos orientados a las necesidades de bibliotecas, archivos y museos. Se revisan las características de esos módulos, incluyendo evolución temporal, número de versiones, desarrolladores implicados, instalaciones y otros datos.

Palabras clave

Drupal, Módulos, Gestión de contenidos, Bibliotecas, Repositorios, Opacs.

Title: Content management with *Drupal*: a review of modules specific to libraries, archives and museums

Abstract

Drupal is one of the most advanced platforms for content management. For product development, *Drupal* architecture uses modules that add specific functionality to its core. More than 20 modules are geared to the needs of libraries, archives and museums. This paper reviews the characteristics and development of these modules, including changes over time, number of versions, developers involved, active implementations, and other data.

Keywords

Drupal, Modules, Content management, Repositories, Libraries, Opacs.

Tramullas, Jesús (2013). "Gestión de contenidos con *Drupal*: revisión de módulos específicos para bibliotecas, archivos y museos". *El profesional de la información*, septiembre-octubre, v. 22, n. 5, pp. 425-431.

<http://dx.doi.org/10.3145/epi.2013.sep.07>

Planteamiento, objetivo y método

Drupal se ha convertido en los últimos años en el sistema de gestión de contenidos (*content management system*, CMS) que mayor expansión ha alcanzado entre la comunidad de diseñadores de productos y servicios del sector de bibliotecas. Su arquitectura modular y su capacidad para implementar proyectos y funciones ex-profeso, así como el creciente

número de programadores especializados con que cuenta, han hecho posible la creación de todo tipo de productos y servicios basados en *Drupal*. El valor añadido que ofrece no ha pasado desapercibido en el ámbito de las unidades de información (Farkas, 2008; Coombs, 2009; Becker, 2010). <http://drupal.org>



El semanario británico *The economist* se pasó a *Drupal* en octubre de 2010. Tiene 3 millones de usuarios registrados y mensualmente contabiliza 25 millones de páginas vistas.
<https://drupal.org/node/915102>

Sin embargo en numerosas ocasiones se acude a *Drupal* esperando encontrar una solución directa que resuelva los problemas y necesidades de un servicio o producto determinado. Adoptar este enfoque es un error. La incorporación de funciones específicas se realiza mediante la instalación de módulos que complementan y amplían las prestaciones que ofrece su *core* o núcleo (Vacek et al., 2010). *Drupal* se ha transformado en una plataforma sobre la cual es necesario implementar elementos específicos que ofrezcan una solución a los requerimientos planteados en la fase de diseño del producto o servicio. Este cambio, que se inició en la versión 6, se ha consolidado en la versión 7, y la futura versión 8 completa esta migración hacia una plataforma de desarrollo. De ser un CMS que ofrece funciones avanzadas, se está transformando en un entorno basado en un framework PHP, como es *Symfony*. *Drupal* se ha ido separando progresivamente del enfoque adoptado por productos como *WordPress* o *Joomla*, orientados a la solución directa y casi inmediata de necesidades.

<http://symfony.com>

Este trabajo tiene como objetivo revisar el estado de evolución de los módulos de *Drupal* especialmente diseñados para implementar productos y servicios de información específicos de bibliotecas, archivos y museos, de forma que se pueda transformar en un opac o portal bibliotecario (Rosenthal; Bernardo, 2010), un gestor de referencias, un servidor OAI, una biblioteca digital (Moore, 2008) o un *front-end* para colecciones digitales. Para ello se ha establecido el siguiente método de trabajo:

- definición de las categorías de análisis, atendiendo a las funciones, evolución y uso de los módulos;
- identificación de los módulos especialmente diseñados para bibliotecas, archivos y museos;
- revisión de los objetivos, características, versiones, modificaciones y uso de los mismos;
- discusión de resultados y elaboración de recomendaciones.

Los módulos que tiene en cuenta este trabajo han sido seleccionados tomando como referencia las enumeraciones disponibles en *Library Modules*, en los archivos de la lista de correo electrónico *Drupal4Lib*, y en **Tramullas y Garrido-Picazo** (2010). La recopilación se ha completado mediante búsquedas específicas por descriptor/título/módulo relacionado en *drupal.org*. Los datos de referencia de cada uno de ellos han sido tomados de los servidores <http://drupal.org> y <http://drupalmodules.com>, disponibles a 31 de mayo de 2013. Dado que entre los objetivos del trabajo se encuentra ofrecer un indicador sobre el desarrollo de cada módulo, se han tomado en consideración las estadísticas básicas, el número de versiones liberadas, y las aportaciones o *commits*. El indicador de instalaciones activas se basa en la recopilación automática de datos que hace *Drupal* en sus versiones 6 y 7 merced al módulo *Update status*.

<http://groups.drupal.org/libraries/modules>

<http://listserv.uic.edu/archives/drupal4lib.html>

<https://drupal.org/node/329620>

No se han cubierto los módulos orientados a la gestión de documentos, dado que su campo de aplicación excede el marco del trabajo planteado, y porque las características difieren entre los módulos que podrían tomarse en consideración. Por el momento no existe una solución a la gestión de documentos o *records management* sobre *Drupal*, basada en la implementación de un módulo o conjunto de módulos especializados. Esta limitación no quiere decir que no sea posible implementar una gestión de documentos o una biblioteca digital sobre *Drupal*, a través del uso de módulos no especialmente identificados y orientados a bibliotecas, archivos y museos.

Drupal en bibliotecas, archivos y museos

- 2006 Se crea una web de referencia, *Drupalib*, actualmente con poca actividad
- 2006 Se crea el grupo *Libraries* dentro de los grupos de *Drupal*, que continúa activo:
<http://groups.drupal.org/libraries>
- 2007 Se inicia la lista de correo *Drupal4Lib*, que es el principal medio de intercambio de información entre desarrolladores y usuarios de *Drupal* en bibliotecas:
<http://listserv.uic.edu/archives/drupal4lib.html>
- 2008 Austin y Harris publican *Drupal in libraries*, dentro de la colección *Library technology reports* de la ALA, un manual introductorio que cubría *Drupal* en su versión 6
- 2010 *American libraries*, la revista web de la *American Library Association*, adopta *Drupal* como su plataforma web, titulado la noticia "Drupal: the change we need" (Fitzpatrick, 2010)
- 2010 Se publica un volumen en el que se recogen experiencias de implementación sobre *Drupal 6* en el contexto español (Tramullas, 2010). Coombs y Hollister (2010) dedican un capítulo a *Drupal* dentro de su manual sobre software libre para bibliotecas
- 2012 ALA TechSource y Neal-Schuman publican un nuevo manual de introducción, también titulado *Drupal in libraries* (Varnum, 2012), que se centra en la versión 7

La importancia de *Drupal* en el campo de las bibliotecas es creciente, como se muestra en la siguiente secuencia cronológica:

Tramullas y Garrido-Picazo (2010) realizaron una recopilación de módulos para bibliotecas y archivos, disponibles para *Drupal 6*: gestión bibliográfica, formato MARC, Z39.50, OAI, e integración en *Drupal* de opacs, incluyendo *Millennium integration* y *Sopac2*. A finales de 2011, el *Multimedia Information Technology Group* de *Cilip* realizó una breve enumeración de módulos de *Drupal* de interés para bibliotecas, bajo el título "Battle for the CMSs: *Drupal* for libraries" (*MMIT*, 2010). En él se incluían algunos módulos para *Drupal 6* que no han tenido mantenimiento posterior, y señalaba sólo a *Sopac* y *eXtensible Catalog* como integradores para opacs.

The screenshot shows the Drupal.org homepage. At the top, there's a navigation menu with links: Get Started, Community, Documentation, Support, Download & Extend, Marketplace, and About. Below the navigation is the Drupal logo and a search bar for drupal.org. A section titled 'Come for the software, stay for the community' describes Drupal as an open source content management platform. There are buttons for 'Drupal Homepage' and 'Log in / Register'. The main content area is divided into three columns: 'Why Choose Drupal?' with a 'Get Started with Drupal' button, 'Sites Made with Drupal' featuring a preview of 'Jewish Exponent', and 'Develop with Drupal' which includes a call to action to help build Drupal 8 and a table of statistics.

23,722 Modules		This week	
1,838 Themes	2,905 Code commits	686 Distributions	5,039 Issue comments
29,112 Developers			

En la parte inferior derecha pueden verse las estadísticas de uso y actividad, que son buscables. <https://drupal.org>

La incorporación de funciones específicas se realiza mediante la instalación de módulos que complementan y amplían las prestaciones que ofrece el núcleo de *Drupal*

Frente a la gran actividad que despliegan los usuarios de bibliotecas, su penetración en archivos y museos parece limitada. Las particularidades de las colecciones digitalizadas de ambos tipos de unidades, tanto en esquemas de metadatos como en formatos, así como la existencia de herramientas especializadas para su gestión, muchas de ellas libres o gratuitas, han limitado el uso de *Drupal* en estos sectores. A ello cabe añadir que en los últimos años el proceso de publicación de colecciones en internet se ha beneficiado de herramientas específicas, como *Collective access* u *Omeka*. Para los usuarios de museos existe el grupo *Museums*, con poca actividad. Esto contrasta con el interés que genera en la comunidad museística: en las últimas ediciones de la conferencia anual *Museums and the Web* siempre ha habido sesiones, talleres o presentaciones dedicadas a usos de *Drupal* en los museos. En lo que concierne a los archivos, no existe un grupo específico de usuarios de *Drupal*, pero al igual que en el caso de los museos, es posible rastrear en la bibliografía y en internet e identificar proyectos de archivos que lo utilizan como plataforma.

<http://collectiveaccess.org>

<http://www.omeka.org>

<https://groups.drupal.org/museums>

<http://www.museumsandtheweb.com>

Módulos para bibliotecas, archivos y museos

Se han identificado y revisado 30 módulos (recogidos en la tabla 1). Los datos básicos han sido tomados de los servidores de referencia indicados, y han sido completados y elaborados para identificar otros hitos significativos (fechas de primeras

versiones, *commits*, resolución de bugs, etc.). La información obtenida ha sido elaborada y organizada en fichas de detalle para cada uno de los módulos, que a su vez se han integrado en una única tabla para facilitar la visualización y la comparación.

Para su estudio se han agrupado en las siguientes categorías:

- integración con repositorio (3);
- integración con gestores de colecciones (1);
- integración de opac (6);
- gestión de ejemplares y circulación (1);
- integración de metadatos bibliográficos (6);
- autenticación de usuarios (1);
- recuperación de información/metabuscador (4);
- gestión de bibliografías (1);
- proveedor/recolector OAI (3);
- esquemas de clasificación (2);
- integración de colecciones de museos (2).

Un resumen de los datos básicos obtenidos se refleja en la tabla 2.

Frente a la gran actividad en bibliotecas, la penetración de *Drupal* en archivos y museos parece limitada

En lo que concierne a la publicación de código, sólo *Fedora commons API*, a pesar de haber sido registrado, no ha producido ni liberado ningún código. De los 29 restantes, 8 no están elaborando nuevas versiones, y de éstos sólo 2 señalan disponer de mantenimiento. Esto indica que los módulos realmente implementables en producción son 21.

Si se atiende a la disponibilidad para versiones (5, 6 y 7), *Drupal 6* es la versión con mayor disponibilidad de módulos (26), de los cuales 15 han incorporado ya código adaptado a *Drupal 7*. A éstos hay que sumar otros 5, que han sido publicados directamente para la versión 7, sin código para

Tabla 1. Módulos identificados y revisados

Nombre	Tipo	Descripción breve del módulo
<i>Fedora commons API</i>	Integración con repositorio	API para acceder y manipular contenido de repositorios <i>Fedora commons</i> desde <i>Drupal</i>
<i>Islandora</i>	Integración con repositorio	Paquete completo que integra el contenido de repositorios <i>Fedora commons</i> en <i>Drupal</i>
<i>DSpace</i>	Integración con repositorio	API para incorporar contenido de repositorios <i>DSpace</i>
<i>Contentdm integration</i>	Integración con gestores de colecciones	Conjunto de módulos que permiten el acceso a colecciones depositadas en <i>Contentdm</i>
<i>Millenium opac integration</i>	Integración de opac	Integra el contenido del opac de sigb <i>Millenium</i> con <i>Drupal</i>
<i>Opac</i>	Integración de opac	Integra el contenido del opac de diferentes sigbs, para lo que requiere la programación de conectores
<i>PMB Drupal</i>	Integración de opac	Proyecto experimental para integrar contenido del opac de <i>PMB</i> en <i>Drupal</i> .
<i>PMB connector</i>	Integración de opac	Integra el contenido del opac de <i>PMB</i>
<i>Sopac2 the social opac</i>	Integración de opac	Integra el contenido del opac de diferentes sigbs, para lo que requiere conectores
<i>eXtensible catalog Drupal toolkit</i>	Integración de opac	Solución completa para integrar un opac o un conjunto de ellos, en <i>Drupal</i> . Usa recolección OAI-PMH, estandariza los datos desde MARC 21 y <i>Dublin core</i> , y puede transformarlos en FRBR y RDA. Puede incluir información en tiempo real sobre préstamo
<i>MARC</i>	Integración de metadatos bibliográficos	Importa registros en formato MARC en nodos de <i>Drupal</i>
<i>Library</i>	Gestión de ejemplares y circulación	Sirve para crear una biblioteca básica de elementos y asignar acciones sobre ellos
<i>LT4L</i>	Integración de metadatos bibliográficos	Muestra datos de libros de un opac, que se asocian a un nodo de <i>Drupal</i> , usando su etiquetado
<i>Google books</i>	Integración de metadatos bibliográficos	Inserta en nodos los datos de libros existentes en <i>Google books</i>
<i>Open library API</i>	Integración de metadatos bibliográficos	Recupera datos de libros existentes en <i>Open library</i>
<i>Book post</i>	Integración de metadatos bibliográficos	Captura datos de diferentes servidores de libros (<i>Open library</i> , <i>WorldCat</i> , <i>LibraryThing</i> y <i>Google books</i>), y los muestra en nodos
<i>ISBN2node</i>	Integración de metadatos bibliográficos	Crea nodos con datos de libros, usando como referencia el ISBN
<i>EZProxy</i>	Autenticación de usuarios.	Implementa autenticación para usuarios de <i>Drupal</i> contra servidores <i>EZProxy</i> , para poder acceder a recursos externos bajo licencia
<i>Z39.50/SRU client</i>	Recuperación de información/metabuscador	Cliente <i>Z39.50</i>
<i>Article discovery</i>	Recuperación de información/metabuscador	Interfaz de consulta en <i>Drupal</i> contra servidores <i>Summon</i>
<i>Pazpar2 metasearch integration</i>	Recuperación de información/metabuscador	Integra en <i>Drupal</i> el metabuscador <i>Pazpar2</i> de <i>Index Data</i>
<i>Bibliography module</i>	Gestión de bibliografías	Permite implantar un sistema web para gestión de referencias bibliográficas
<i>Views OAI-PMH</i>	Proveedor/recolector OAI	Implementa un servidor OAI
<i>OAI2 for CCK</i>	Proveedor/recolector OAI	Implementa un servidor OAI
<i>Feeds OAI-PMH Fetcher/Parser</i>	Proveedor/recolector OAI	Recolecta datos de servidores OAI
<i>Hilcc taxonomy tag</i>	Esquema de clasificación	Asigna términos de clasificación a los nodos atendiendo a los <i>LCC call numbers</i>
<i>MeSH vocabulary</i>	Esquema de clasificación	Integra los términos del <i>MeSH</i> como taxonomía
<i>Apache Solr search integration</i>	Recuperación de información/metabuscador	Integra la indización y búsqueda con <i>Solr</i> en implementaciones de <i>Drupal</i>
<i>Collective access</i>	Integración de colecciones de museos	Integra datos de colecciones gestionadas con <i>Collective access</i>
<i>Adlibapi</i>	Integración de colecciones de museos	Integra datos de colecciones gestionadas con <i>Adlib</i>

las previas, lo que eleva la disponibilidad hasta los 20. Diez módulos han detenido su desarrollo en la versión 6, y no parece que vayan a liberar nuevo código adaptado a la 7 o al futuro *Drupal 8*. De todo el conjunto, sólo 8 módulos estaban disponibles para *Drupal 5*.

Discusión de resultados

No se han valorado las funciones y prestaciones que cada uno de los módulos puede ofrecer en un contexto dado, ya

que no se pretende analizar la calidad del software. La información recopilada se utiliza para delinear el panorama del desarrollo y disponibilidad de los módulos para la gestión de información en unidades y servicios de información.

La primera cuestión que cabe señalar es la ausencia de módulos orientados a la integración con sistemas de gestión de archivos. Esto se debe precisamente a la capacidad de llevar a cabo cualquier tipo de contenido y la estructura de datos co-

respondiente con módulos de uso general, y al potencial de importación de casi cualquier contenido que ofrece el módulo *Feeds*, ya señalado. Con *Drupal* es posible implementar procesos de gestión de documentos (*records management*) usando la gran variedad de módulos disponibles. Además el desarrollo de módulos que usan el estándar CMIS hace posible importar y compartir información con herramientas como *Alfresco*, *Nuxeo*, *KnowledgeTree* o *SharePoint*.

El panorama de los módulos para museos es similar. Sólo se encuentran módulos de integración de contenidos para dos sistemas de gestión de colecciones de museos (*Collective access* y *Adlib*), ya que se está usando la misma aproximación que en archivos para la generación de productos basados en *Drupal*.

Los repositorios disponen de mayor número de opciones. Hay dos grupos de módulos:

- los que son interfaces directas a software específico, como los módulos para *DSpace* o *Fedora (Islandora)*;
- los que implementan un servidor o un recolector OAI, con el estándar *Dublin core*, y son capaces de ofrecer o de importar la información propia a terceros, o de recolectar registros externos e importar la información en tipos de nodos, y se apoyan en la funcionalidad del módulo de uso general *Feeds*.

Un tercio de los módulos revisados no ha publicado versión para *Drupal 7*, lo que podría plantear problemas para el mantenimiento de los servicios a corto plazo

Las bibliotecas disponen del mayor número de posibilidades. En primer lugar, cabe destacar los módulos que proveen de una capa de interfaz contra el opac de los sigbs tradicionales: *Millenium* (Garza, 2009), *PMB*, *opac*, *eXtensible catalog...* Las instalaciones no son muy numerosas, lo que puede deberse a los requerimientos y restricciones que impone a las bibliotecas el tener que acometer una implementación nueva, cuando ya se dispone de un opac en funcionamiento. Las ventajas que ofrece la integración de una capa social y de funciones de valor añadido en el opac parecen pesar menos que el potencial coste asociado. A ello cabe añadir las mejoras que han tenido los opacs en las últimas versiones de los productos comerciales.

Un segundo grupo lo formarían las herramientas para metabuscadores, como *Article discovery* y *Pazpar2 metasearch integration (mkdru)*, que proveen de acceso a recursos Z39.50 y similares.

En un tercer grupo entrarían los módulos capaces de capturar información bibliográfica, y de integrarla y mostrarla en

The screenshot shows the website **liiscience.org** with a navigation menu: LII INFO, LII COMPARISONS, MEETINGS, FORUMS, PUBLICATIONS, LOGIN. The main content area displays the title "Soot mass concentration measurements in diesel engine exhaust using laser-induced incandescence" and a table with the following data:

Title	Soot mass concentration measurements in diesel engine exhaust using laser-induced incandescence
Publication Type	Journal Article
Year of Publication	1996
Authors	Case, ME, Hofeldt, DL
Journal Title	Aerosol Science and Technology
Volume	25
Pages	46-60
Keywords	521 Fuel Combustion and Flame Research; 612.1.1 Internal Combustion Engine Components; 612.2 Diesel Engines; 744.8 Laser Beam Interactions; 804 Chemical Products Generally; Diesel engines; Exhaust systems (engine); Laser beam effects; Scattering; soot
Abstract	Simultaneous measurements of laser-induced incandescence (LII) and elastic scattering from soot particles in diesel engine exhaust have been made. The LII signal scaled linearly with the mass concentration of the non-volatile particulate mass fraction over the entire range of engine operating loads. Over this range of conditions, the volume mean diameter of the soot particles varied from 0.07 to 0.11 μm, but the size change did not appear to affect the signal response. The scattering response did not scale linearly with the mass concentration of soot. Mass concentrations of 0.2 mg/m ³ were easily detectable, with even lower values possible. Additional techniques for determining the volatile fraction of particulate mass are described. (Author abstract). EIP/LUS (c) 1996 Engineering Information Inc.

At the bottom right of the page, there are links for Google Scholar, BibTex, RTF, Tagged, and RIS.

Módulo *bibliography* en su configuración de ejemplo en *liiscience.org*

los nodos de *Drupal*. Los datos tomados de *LibraryThing*, *The open library* o *Google books* pueden incorporarse a cualquier tipo de producto o servicio, que no tiene porque ser estrictamente bibliotecario. Contrasta el uso de los módulos de datos bibliográficos con la carencia de módulos orientados a importar o gestionar esquemas de clasificación, lo que puede deberse precisamente a la importación directa de los esquemas desde los catálogos, o a la utilización del módulo de uso genérico *Taxonomy import/export via XML*, que es capaz de integrar en las taxonomías de *Drupal* todo tipo de esquemas de clasificación, y que muestra un elevado uso (2.371 instalaciones activas).

Una cuestión clave para la implementación de los módulos es la disponibilidad de un soporte adecuado para los problemas que puedan surgir. Dadas las características de las herramientas y de la comunidad, es necesario que los autores de código documenten adecuadamente no sólo éste, sino los procesos de instalación, configuración y ejecución del mismo. En el conjunto revisado sólo catorce módulos ofrecen documentación de referencia, lo que puede suponer un problema en determinados contextos.

La revisión de la disponibilidad y del desarrollo de módulos específicos para bibliotecas, archivos y museos en *Drupal* refleja la pequeña comunidad de usuarios implicados activamente en este campo

En contraste con los módulos anteriores deben destacarse, por sus características de desarrollo y utilización, *Bibliography* y *Apache Solr search integration*. El módulo *Biblio* es un clásico de *Drupal*, que lleva en producción desde la versión 4.7 en 2006, e indica tener más de 4.500 instalaciones activas. A su alrededor se ha generado un buen conjunto de

Tabla 2. Datos básicos de cada módulo

Módulo	Núm. versión	Primera	Última	D5	D6	D7	Mant. activo	Estado desarr.	Documentación
<i>Fedora commons API</i>	0			No	No	No	No	Inact.	No
<i>Islandora</i>	8	02/12/08	18/03/13	No	Sí	Sí	Sí	Activo	Sí
<i>DSpace</i>	1	01/01/12	13/09/12	No	No	Sí	Sí	Activo	No
<i>Contentdm integration</i>	1	22/07/11	26/06/12	No	Sí	No	Sí	Activo	No
<i>Millenium opac integration</i>	17	15/01/08	12/02/13	Sí	Sí	Sí	Sí	Activo	Sí
<i>opac</i>	13	25/07/12	31/05/13	No	No	Sí	Sí	Activo	Sí
<i>PMB Drupal</i>	1	16/11/11	16/11/11	No	Sí	No	Sí	Activo	Sí
<i>PMB connector</i>	8	12/11/11	25/05/13	No	Sí	Sí	Sí	Activo	Sí
<i>Sopac2 the social opac</i>	4	17/03/09	08/03/10	No	Sí	No	No	Inact.	Sí
<i>eXtensible catalog Drupal toolkit</i>	20	31/03/11	22/05/13	No	Sí	Sí	Sí	Activo	Sí
<i>MARC</i>	7	19/12/07	16/06/10	Sí	Sí	No	No	Inact.	No
<i>Library</i>	3	09/08/08	02/03/13	No	Sí	Sí	No	Inact.	No
<i>LT4L</i>	4	14/10/10	13/06/11	No	Sí	Sí	Sí	Activo	No
<i>Google books</i>	2	06/03/12	24/02/13	No	No	Sí	Sí	Activo	Sí
<i>Open library API</i>	1	11/08/09	25/05/13	No	Sí	No	No	Inact.	No
<i>Book post</i>	1	07/10/08	07/10/08	NO	Sí	No	No	Inact.	No
<i>ISBN2node</i>	3	07/06/11	09/06/11	No	No	Sí	Sí	Activo	No
<i>EZProxy</i>	9	28/04/09	05/01/11	Sí	Sí	Sí	Sí	Activo	Sí
<i>Z39.50/SRU client</i>	4	01/05/06	25/07/09	Sí	Sí	No	No	Inact.	No
<i>Article discovery</i>	1	07/04/11	07/04/11	No	Sí	Sí	No	Activo	No
<i>Pazpar2 metasearch integration</i>	5	02/08/11	02/04/13	No	Sí	Sí	Sí	Activo	Sí
<i>Bibliography module</i>	89	23/06/06	31/05/13	Sí	Sí	Sí	Sí	Activo	Sí
<i>Views OAI-PMH</i>	8	30/05/11	21/05/13	No	Sí	Sí	Sí	Activo	No
<i>OAI2 for CCK</i>	5	06/08/08	23/11/11	Sí	Sí	No	Sí	Inact.	No
<i>Feeds OAI-PMH Fetcher/Parser</i>	3	17/10/10	06/09/11	No	Sí	Sí	Sí	Activo	Sí
<i>HILCC taxonomy tag</i>	2	12/03/09	25/02/11	Sí	Sí	No	Sí	Activo	No
<i>MeSH vocabulary</i>	4	11/01/11	11/01/11	No	Sí	No	Sí	Inact.	No
<i>Apache Solr search integration</i>	86	25/12/07	26/05/13	Sí	Sí	Sí	Sí	Activo	Sí
<i>Collective access</i>	8	14/08/11	02/02/12	No	Sí	Sí	Sí	Activo	Sí
<i>Adlibapi</i>	2	09/02/13	09/02/13	No	No	Sí	Sí	Activo	No

submódulos que amplían sus funciones, hasta configurar un completo sistema de gestión bibliográfica y de referencias. En su desarrollo se han generado casi 2.800 *commits* (aportaciones), por parte de 17 desarrolladores. La integración de *Solr* como motor de búsqueda para propósitos múltiples en *Drupal* es exitosa; su capacidad para trabajar con información estructurada en xml le da especial interés para las unidades de información. Varios de los módulos analizados incluyen una implementación de *Solr*. Al igual que *Biblio*, muestra un patrón de comunidad masivo: en cinco años, 2.100 *commits* y 53 desarrolladores.

Los datos de comunidad y desarrollo de *Biblio* y *Solr* sirven como contrapunto para el resto de módulos. La mayoría de los analizados indican uno o dos desarrolladores, con un número de *commits* bajo. Sólo *Millenium opac integration*, *opac* y *eXtensible catalog* muestran una pauta activa de *commits*, aunque *Millenium* sólo tiene 1 desarrollador. Esta pauta indica que el desarrollo de la mayoría de los módulos tiende a ser unipersonal, con poca o nula comunidad impli-

cada, y que depende de encargos específicos o de interés personal. Esta tendencia es un riesgo que puede afectar la evolución del producto, y amenazar la aparición de nuevas versiones o actualizaciones. Una segunda pauta es la que marcan empresas de software libre, que aplican el módulo como uno más de los productos que ofrecen a sus clientes, como *opac*, *Pazpar2 metasearch integration (mkdru)* y *PMB connector*. Relacionado con lo anterior está el bajo número de versiones del módulo que se suele publicar (en el recuento se han considerado tanto las versiones estables como las de desarrollo), excepción hecha de los cinco módulos señalados en este párrafo. También cabe señalar que el módulo puede dejar de actualizarse mientras cumpla adecuadamente con la finalidad para la cual fue diseñado, lo que también parece suceder en varios casos de los analizados.

La evolución de las versiones de *Drupal* también tiene su incidencia en los módulos. Un tercio de los revisados no ha publicado versión para *Drupal 7*, lo que podría plantear problemas para el mantenimiento de los servicios a corto plazo, cuando

se publique la versión 8, y se deje de dar soporte a la versión 6. La mayor parte de estos módulos publicaron su última versión hace dos o más años, y, aunque señalan encontrarse mantenidos y en producción, cabe dudar de la fiabilidad de alguno de estos datos. Sin embargo, varios tienen sus funciones cubiertas por nuevos módulos para la versión 7, o por actualizaciones de otros módulos existentes, que han ampliado sus funciones.

Conclusiones

La revisión de la disponibilidad de módulos específicos para bibliotecas, archivos y museos en *Drupal* refleja la pequeña comunidad de usuarios implicados activamente en este campo. En los próximos años, el diseño e implementación de productos y servicios más avanzados va a hacer necesario que bibliotecas, archivos y museos se planteen la necesidad de dedicar medios a la implementación de software que responda específicamente a sus necesidades. Esto deberá ser llevado cabo en colaboración, mediante la asunción o dotación compartida de recursos.

La implementación de estas funciones resulta en ocasiones compleja para un usuario de tipo medio, como suele ser el profesional del sector. En ello inciden factores como la evaluación de funcionamiento de los módulos, la compatibilidad entre versiones, las opciones de instalación y/o configuración, etc. Están apareciendo distribuciones o perfiles de instalación, que simplifican los procesos necesarios para disponer de una instalación plenamente operativa. Por ejemplo, se encuentran disponibles *eXtensible Catalog (XC) Installation*, *OAIConnect installation profile*, o *AgriDrupal* (Subirats et al., 2010).

https://drupal.org/project/xc_installation

<https://drupal.org/project/oaiconnect>

Drupal es una de las plataformas de gestión de contenidos más extendida y con mayor funcionalidad, publicada bajo licencia de software libre. En el mercado puede encontrarse un elevado número de profesionales y empresas que ofrecen soporte y desarrollo. El modelo de negocio aplicado, basado en el ofrecimiento de valor añadido, funciona correctamente, como demuestra también el ejemplo de *WordPress*. *Drupal* es por tanto una base sólida para la implementación de diferentes tipos de productos o servicios de información digital.

Bibliografía

Austin, Andy; Harris, Christopher (2008). *Drupal in libraries. Library technology reports*, v. 44, n. 4, Chicago: ALA. <http://dx.doi.org/10.5860/ltr.44n4>

Becker, Danielle A. (2010). "Adventures in *Drupal*. Designing a library website using a CMS". *Online*, v. 34, n. 5, pp. 19-21.

Coombs, Karen (2009). "Drupal done right". *Library journal*,

v. 134, n. 19, pp. 30-32.

Coombs, Karen; Hollister, Amanda (2010). *Open source web applications for libraries*. Information today. ISBN: 978 15 738 7400 7

Farkas, Meredith (2008). "CMS for next-gen websites: using *Drupal* to manage library web content". *American libraries*, v. 39, n. 10.

Fitzpatrick, Sean (2010). "Drupal: the change we need". *American libraries*. <http://www.americanlibrariesmagazine.org/blog/drupal-change-we-need>

Garza, Alejandro (2009). "From OPAC to CMS: *Drupal* as an extensible library platform". *Library hi tech*, v. 27, n. 2, pp. 252-267. <http://dx.doi.org/10.1108/07378830910968209>

Moore, Tony A. (2008). "Assessing and adapting *Drupal* content management system as a digital library application". *Procs of the American Society for Information Science and Technology*, v. 45, n. 1, pp. 1-4. <http://dx.doi.org/10.1002/meet.2008.14504503131>

Rosenthal, Danielle; Bernardo, Mario (2010). "Creating a library database search using *Drupal*". *Code4Lib journal*, v. 10. <http://journal.code4lib.org/articles/2920>

Subirats, Imma; Malapela, Thembani; Dister, Sarah; Zeng, Marcia; Goovaerts, Marc; Pesce, Valeria; Jaques, Yves; Anibaldi, Stefano; Keizer, Johannes (2012). "Reorienting open repositories to the challenges of the semantic web: experiences from FAO's contribution to the resource processing and discovery cycle in repositories in the agricultural domain". *6th Research conf on metadata and semantics research, MTSR 2012, Communications in computer and information science*, v. 343, pp. 158-167. <http://eprints.rclis.org/17733>

Tramullas, Jesús (ed.) (2010). *Drupal para bibliotecas y archivos*. Zaragoza: Fund. Zaragoza Ciudad del Conocimiento. ISBN: 978 44 613 9611 5 <http://eprints.rclis.org/14400>

Tramullas, Jesús; Garrido-Picazo, Piedad (2010). "Drupal: módulos para bibliotecas, archivos y servicios de información y documentación." En: Tramullas, Jesús (ed.) *Drupal para bibliotecas y archivos*. Zaragoza: Fund. Zaragoza Ciudad del Conocimiento, pp. 21-28. ISBN: 978 44 613 9611 5

Vacek, Rachel; Watkins, Sean; Morris, Christina M.; Keller, Derek (2010). "Improving the *Drupal* user experience". *Code4Lib journal*, v. 12. <http://journal.code4lib.org/articles/4578>

Varnum, Ken J. (2012). *Drupal in libraries*. Chicago: ALA TechSource y Neal-Schuman. ISBN: 978 15 557 0778 1



EVOLUCIÓN DE LOS REPOSITORIOS DOCUMENTALES. EL CASO SOCIALNET



Pablo Lara-Navarra, Enric Serradell-López y David Maniega-Legarda



Pablo Lara-Navarra, es licenciado en documentación por la *Universidad de Granada*, doctor por la *Universitat Pompeu Fabra*, y master de *Sociedad de la información y conocimiento* por la *Universitat Oberta de Catalunya (UOC)*. Su carrera se centra en la docencia, innovación en productos y consultoría a empresas. Es especialista en gestión de innovación, gestión de conocimiento y e-learning, temas sobre los que ha publicado numerosos libros y artículos. Ha sido director académico del *Master de tecnologías accesibles UOC-Techosite* y director de *Innovación* de la *UOC*. Es profesor de los *Estudios de Ciencias de la Información y de la Comunicación* de la *UOC* y asesor de empresas tecnológicas.

<http://orcid.org/0000-0003-0595-3161>

Universitat Oberta de Catalunya, Estudios de Información y Comunicación
Rambla de Poblenou, 156. 08018 Barcelona, España
plara@uoc.edu



Enric Serradell-López es doctor en ciencias económicas y empresariales por la *Universidad de Barcelona*, profesor de la *Universitat Oberta de Catalunya* y director de programas MBA y directivos. Su docencia e investigación está vinculada al e-learning y al análisis de los intangibles como ventaja competitiva de las empresas. Es miembro de la *Asociación Científica de Economía y Dirección de Empresas (Acede)* y forma parte de la comisión de intangibles de la *Asociación Catalana de Contabilidad y Dirección de Empresas (Accid)*.

<http://orcid.org/0000-0003-2719-1380>

Universitat Oberta de Catalunya, Estudios de Economía y Empresa
Av. Tibidabo, 39-32. 08035 Barcelona, España
eserradell@uoc.edu



David Maniega-Legarda, historiador del arte y documentalista técnico, es fundador y chief technology officer (CTO) de *ICALIA*. Ha sido director de la *Oficina Abierta de Innovación* de la *Universitat Oberta de Catalunya (UOC)* y responsable del equipo web del *Servicio de Cooperación Bibliotecaria* del *Depto. de Cultura y Medios de Comunicación* de la *Generalitat de Catalunya*. Actúa como consultor de usabilidad y accesibilidad web, estrategia tecnológica y procesos de innovación en las organizaciones. Es experto en la creación y desarrollo de proyectos de innovación centrados en e-learning. Es profesor en el *Máster de documentación digital* de la *Universitat Pompeu Fabra* donde además es coordinador de proyectos web.

<http://orcid.org/0000-0002-9878-9446>

ICALIA Solutions SL
Almogàvers, 119-123. 08018 Barcelona, España
david.maniega@icalia.es

Resumen

Se describen algunos conceptos y tendencias tecnológicas tales como identidad digital (*digital identity*), grandes conjuntos de datos (*big data*), datos en abierto (*open data*) y visualización de datos (*data visualization*) que pueden aplicarse a los repositorios documentales. Se presenta el modelo de repositorio *SocialNet*, como ejemplo de proyecto innovador diseñado en consonancia con tales tendencias.

Palabras clave

Repositorios, Datos en abierto, Grandes datos, Gestión del conocimiento, Gestión de contenidos, Archivos abiertos, *SocialNet*.

Title: Evolution of documents repositories. The *SocialNet* case

Abstract

Some concepts and technology trends such as digital identity, large sets of data (big data) open data and data visualization that can be applied to document repositories are briefly described. In this article it is presented *SocialNet* repository model as an example of innovative project designed in line with such trends.

Artículo recibido el 18-07-2013
Aceptación definitiva: 11-08-2013

Keywords

Repositories, Open data, Big data, Knowledge management, Content management, Open archives, *SocialNet*.

Lara-Navarra, Pablo; Serradell-López, Enric; Maniega-Legarda, David. "Evolución de los repositorios documentales. El caso *SocialNet*". *El profesional de la información*, septiembre-octubre, v. 22, n. 5, pp. 432-439.

<http://dx.doi.org/10.3145/epi.2013.sep.08>

1. Principales tendencias y retos en la gestión de información

Para los profesionales de la información es de capital importancia monitorizar la evolución del mercado, analizando y conceptualizando las tendencias tecnológicas para tratar de anticipar el futuro. Tienen que detectar los cambios de comportamientos del consumidor, frente a los nuevos productos y servicios de información.

El mercado tecnológico de la información es analizado por grandes consultoras como *Deloitte*, *Accenture*, *Gartner*, *IDC*..., las cuales detectan problemas, necesidades y oportunidades en la gestión de información y realizan o encargan nuevos productos/soluciones para resolverlos. Concretamente *Deloitte* (2012) vaticina:

- cambio radical de entender los productos y servicios de información debido al auge de los *smartphones* y *tablets*, de los que la consultora estima que antes de 2016 se habrán vendido unos 1.600 millones de unidades;
- crecimiento vertiginoso en el uso de redes sociales, con unos 1.500 millones de usuarios;
- computación en la nube (*cloud computing*);
- tecnologías M2M (*machine to machine*).

Existe un consenso generalizado en cuanto a las innovaciones que pueden cambiar los modelos de información actuales. Entre ellas:

- ludificación (*gamification*);
- computación en la nube (*cloud computing*);
- identidad digital (*digital identity*);
- grandes datos (*big data*);
- *davi* (*data visualization*),
- redes de colaboración/relación profesional (*social enterprise*), entre otras.

En este trabajo nos centraremos en identidad digital, grandes datos y visualización de datos, que consideramos determinarán la evolución de los repositorios documentales.

2. Los repositorios como generadores de identidad digital

Los repositorios documentales han ido evolucionando desde su uso básico de sacar a la luz trabajos no publicados (*preprints*), a constituirse en una verdadera ventana para construir la identidad digital de autores e instituciones. Actualmente van más allá de las actividades de organizar, archivar, preservar documentos, y tienen un papel determinante como difusores de la producción intelectual de conocimiento.

Valga como ejemplo la descripción del repositorio *Digital.CSIC*, que indica que es una "estrategia de difusión de las pu-

blicaciones científicas, ofrece la posibilidad de incrementar la visibilidad de las mismas y de llegar a un número mayor de lectores, lo que potencialmente se traducirá además en un número más elevado de citas a los trabajos".

<http://digital.csic.es/faqs/#faq1>

Pero ¿qué es la identidad digital? Haciendo un símil, al igual que medimos el impacto medioambiental de nuestra sociedad sobre los recursos naturales con la "huella de carbono", en el mundo virtual empieza a definirse el concepto de "huella digital" (*Deloitte*, 2012) o activos digitales (*Cantera*, 2013) como el impacto de nuestra presencia en la Red, el rastro de información y conocimiento que generamos a lo largo de nuestra vida digital.

Las personas tenemos varias identidades digitales correspondientes a nuestras diferentes facetas (personal, profesional, consumidor, socio-cultural...). *Deloitte* (2012) afirma "que el número de credenciales que definen a un individuo ha crecido exponencialmente y la mayoría de usuarios cuentan con alguna forma de huella digital".

Nuestro entorno cambia y se expande a través de las diferentes redes a las que estamos conectados, por lo que gestionar las identidades resulta cada vez más complejo. Por ello se busca poder hacerlo mediante una única tecnología.

En el caso concreto de la documentación, *Giones-Valls* y *Serrat-Brustenga* (2010) indican que "la gestión de la identidad digital personal se define como la habilidad de gestionar con éxito la propia visibilidad, reputación y privacidad en la red como un componente inseparable y fundamental del conjunto de habilidades informacionales y digitales, las cuales se han convertido en fundamentales para vivir en la sociedad informacional".

Es de vital importancia el uso que damos a nuestros activos digitales (cuentas, perfiles, webs, blogs, listas, documentos en repositorios...) para tener una notoriedad correcta y una buena reputación en la Red. Debe controlarse con eficacia la identidad digital, ya que ésta ha superado las fronteras de nuestro PC y debemos ser conscientes de que la información que define quién somos, qué hacemos, dónde estamos, cuándo decidimos, por qué actuamos y cómo pensamos se encuentra disgregada y distribuida por la nube de internet.

Gartner (2012) asegura que "la nube reemplazará gradualmente al PC como el lugar donde las personas mantienen su contenido personal, tienen descritas sus preferencias, el acceso a sus servicios, convirtiéndose en el centro de su vida digital".

Aunque existen servicios que ayudan a la identificación y estandarización de los nombres de los autores (como *Author Resolver*, *Inspire*, *IraLIS*, *Scopus Author ID*, *ResearcherID* y



Figura 1. Tableau
<http://www.tableausoftware.com/es-es>

recientemente *Orcid* –que parece por fin el sistema definitivo-) carecen de una visión global de la identidad digital y aportan soluciones parciales, aunque todos tienden a vincular los nombres de los investigadores a sus instituciones y publicaciones, favoreciendo el proceso de descubrimiento científico.

3. Repositorios, big data y open data

Según *IDC (Gantz et al., 2008)* la producción digital total del mundo fue de 161 exabytes en 2006 y se preveía crecer hasta 988 exabytes en 2010, es decir, multiplicarse por 6 en cuatro años (1 exabyte = 1.000 millones de gigabytes). Otra estimación de IDC-EMC (2011) prevé que el volumen de datos alcance los 35.000 exabytes en 2020.

Un parte de estos datos corresponde a la información que manejamos usualmente los bibliotecarios y documentalistas, mucha de la cual se ubica en repositorios. Es difícil saber el número exacto de repositorios existentes, y mucho más conocer el número de documentos (textos, presentaciones, audios, vídeos...) que albergan. Algunos servicios pueden ofrecernos una idea:

- *Ranking web de repositorios*, del Laboratorio de Cibermetría del CSIC, tiene censados 1.650:
<http://repositories.webometrics.info/es>
- *Directory of Open Access Repositories (OpenDOAR)* de la University of Nottingham lista 2.340:
<http://www.opendoar.org>
- *Registry of open access repositories (ROAR)* de la University of Southampton tiene registrados 3.480:
<http://roar.eprints.org>

Los sistemas con robots recolectores o cosechadores (*harvesters*) pueden también darnos una indicación del volumen de documentos accesibles:

- *Recolecta*, de la Fundación Española para la Ciencia y la Tecnología (Fecyt), indexa los registros bibliográficos de 286 colecciones y repositorios.
<http://www.recolecta.net>



Figura 2. Piktochart
<http://piktochart.com>

- *Hispana*, del Ministerio de Educación, Cultura y Deporte, indexa 4,6 millones de documentos y otros objetos digitales procedentes de 202 repositorios.
<http://hispana.mcu.es>
- *Universia* indexa casi 18 millones de documentos, tesis, etc., de 226 repositorios.
<http://biblioteca.universia.net>

Por último citaremos *OAster*, de OCLC, con 23 millones de registros recolectados de 1.100 repositorios.
<http://www.oclc.org/oaister.en.html>

Como en otros tipos de acumulaciones de datos e información, en los repositorios –todos los citados son libremente accesibles- es posible utilizar aplicaciones de inteligencia (*business intelligence*) con las cuales pueden extraerse informes sobre autores e instituciones: qué, cuánto, cómo y con quién investigan, impacto obtenido, etc. Entre las aplicaciones de inteligencia cabe citar desde las hojas de cálculo, hasta softwares como *reporting and querying*, *online analytical processing (olap)*, cuadros de mando integrales (*dashboards*), minería de datos (*data mining*), etc. Todos estos sistemas han culminado en una de las grandes tendencias tecnológicas actuales: *big data*, que es el análisis masivo de ingentes cantidades de datos (logs de webs, datos meteorológicos, estelares, biológicos...) para extraer tendencias y conclusiones que sirvan para tomar decisiones. Las técnicas *big data* permiten tratar cantidades de información sin precedentes, para las cuales hasta ahora no existían ni software ni máquinas suficientemente rápidas.

Big data

Kusnetzky (2010) define las técnicas *big data* (grandes cantidades de datos) como las herramientas, procesos y procedimientos que permiten crear, manipular y gestionar conjuntos muy grandes de datos, así como sus instalaciones de almacenamiento. Según **Dans** (2011) *big data* es el “tratamiento y análisis de enormes repositorios de datos, tan desproporcionadamente grandes que resulta imposible tratarlos con las herramientas de bases de datos y analíticas convencionales”.



Figura 3. *Many Eyes*
<http://www-958.ibm.com/software/data/cognos/manyeyes>

Martínez-López (2012) afirma que tratar con volúmenes de información fuera de lo habitual se convierte en un problema de varios órdenes: de escalabilidad (¿podemos seguir creciendo?, ¿hasta cuánto?); de tratamiento (¿podremos tratar esos datos? ¿y cuándo tengamos el doble?); y de disponibilidad (¿es viable mantener online todos estos datos?).

La información contenida en la mayoría de repositorios documentales individuales se pueden analizar con las técnicas y herramientas clásicas, pero si se pudieran agregar todos ellos configurando un conjunto de datos (*data set*) distribuido, entonces se necesitarían técnicas *big data*.

Es pronto para saber cuál será la trascendencia y alcance futuro de los grandes conjuntos de datos, tanto en el ámbito profesional como para la ciudadanía en general, pero **Tim Berners-Lee** (TEDtalks, 2009) afirmó: “si la gente coloca sus datos en la Web, junto a datos gubernamentales, científicos, y los propios de la comunidad, podrán ser usados por otras personas en formas que nunca podrían haberse imaginado”.

Open data

Son *open data* (datos abiertos o en abierto) los que se pueden capturar, tratar, analizar, modificar, reutilizar o difundir libremente para proporcionarles un nuevo valor informativo.

Mazzo-Iturriaga (2010) dice que “*open data* se asemeja al movimiento *open source*, pero a diferencia de éste no apunta a desarrollar software sino a posibilitar el acceso y reutilización de datos”. **Peset, Ferrer-Sapena y Subirats-Coll** (2011) enfocan el concepto destacan la importancia de los estándares técnicos, de forma que para que los datos sean realmente abiertos y reutilizables deben ofrecerse en formatos normalizados. Estas autoras también se refieren a la exposición de los datos en la Web descritos en RDF (*resource description framework*), con indicación de la relación que puede existir entre esos datos y otros.

Si los repositorios de datos en abierto aplicaran estándares –y por tanto, gracias a esto, los datos fueran reutilizables–, permitirían generar clasificaciones, ontologías y formas de



Figura 4. *Creately*
<http://creately.com>

explotación nuevas mediante las técnicas de minería de datos, *big data* y visualización.

4. Visualización de datos en repositorios

La mayoría de repositorios documentales se han centrado en mejorar el almacenamiento, organización, clasificación, interoperabilidad, accesibilidad, usabilidad..., pero se han preocupado poco de aplicar técnicas de visualización de datos que harían la información más clara, atractiva y fácil de comprender. La presentación de la información de forma gráfica es una forma de análisis que permite identificar tendencias y patrones de relaciones que de otro modo serían muy difíciles de encontrar o ver.

Las técnicas de visualización de datos se encuentran muy ligadas a *big data*, pues son muy adecuadas para presentar los resultados resumidos del análisis de grandes cantidades de información. Algunos ejemplos de sistemas que permiten crear infografías online se citan en el blog *BVA innovación* (2013):

Tableau (figura 1)

<http://www.tableausoftware.com/es-es>

Piktochart (figura 2)

<http://piktochart.com>

Many Eyes (figura 3)

<http://www-958.ibm.com/software/data/cognos/manyeyes>

Creately (figura 4)

<http://creately.com>

Visual.ly

<http://visual.ly>

Los sistemas de realización de infografías y de visualización de datos en general constituyen una parte de la tendencia conocida como narrativa digital (*digital storytelling*). A los repositorios, sean documentales o de datos, se les pueden incorporar sistemas de visualización que permitan consultarlos por diferentes caminos siguiendo las necesidades o deseos del usuario, en tiempo real (**Shirky**, 2012). **Ferrer-**

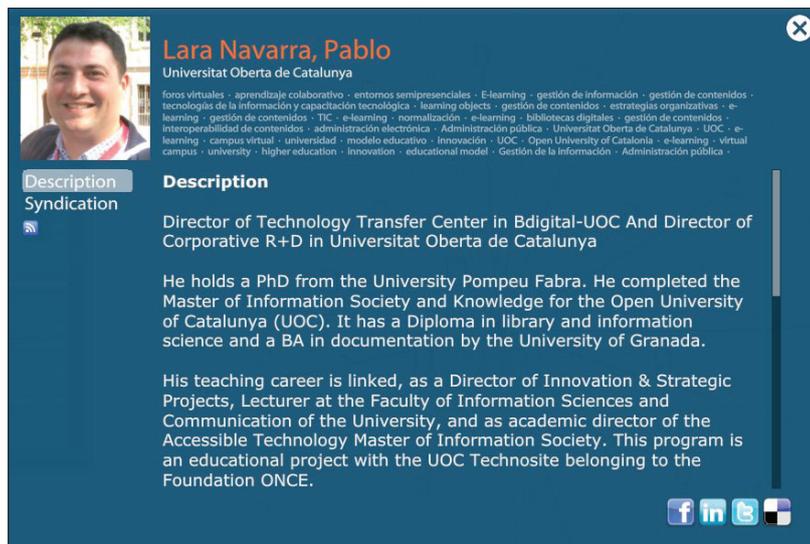


Figura 5. Modelo de ficha de autor y descripción

Sapena y Sánchez-Pérez (2013) apuntan que para contar buenas historias se necesitan capacidades para el análisis de los datos y buenas visualizaciones.

Según Díaz-Delgado (2009) la narrativa digital es “enfrentarse a un nuevo paradigma discursivo: el hipertexto, un sistema de escritura electrónica que organiza información de modo no lineal, con base en estructuras red, esto es, estructuras constituidas por nodos y enlaces”. Y, desde un punto de vista documental, Herrera (2011) propone estrategias para desarrollar una taxonomía en narrativa digital, en la que prevalecen la capacidad de contar de una forma dinámica, con datos en tiempo real y una distribución masiva.

En el contexto de este artículo, consideramos importante la capacidad que deben tener los repositorios documentales de narrar visualmente estructuras de relaciones, como mi-

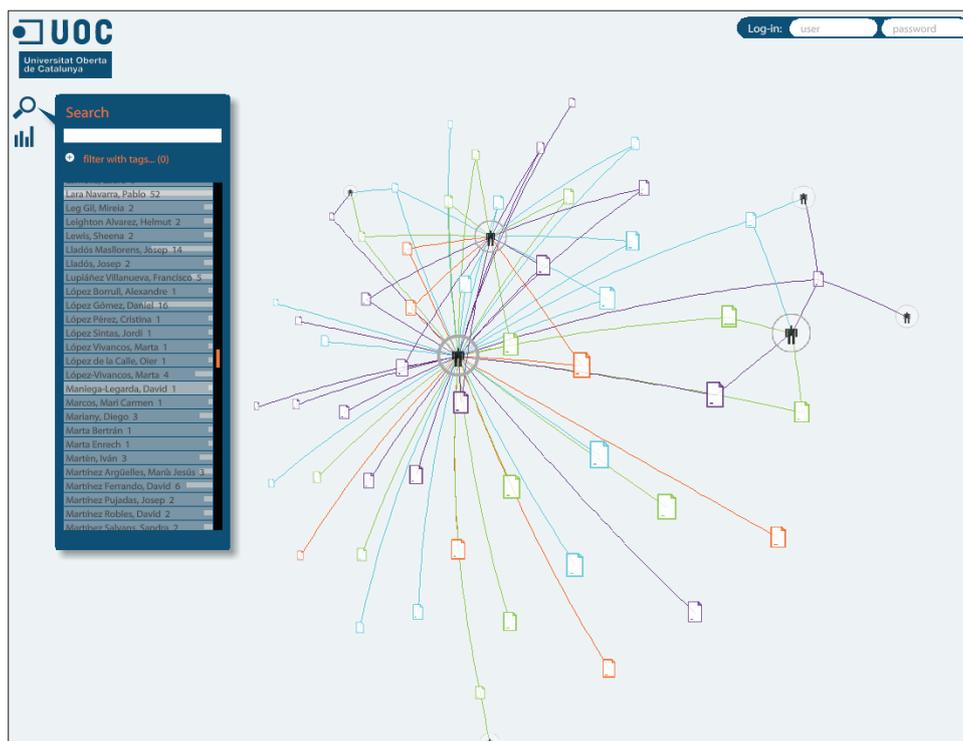


Figura 6. Relaciones autor-documentos y autor-autor a través de documentos coautorados

nimo entre autores, documentos e instituciones.

5. SocialNet: gestión de datos de repositorios documentales

El proyecto SocialNet se concibió para incorporar las tendencias tecnológicas descritas en las secciones anteriores: identidad digital, big data, datos abiertos y visualización a los repositorios documentales. Aclaramos que en este momento SocialNet funciona sobre un repositorio de prueba con un número limitado de nodos por lo que no se trata propiamente de un ejemplo de big data, pero podría serlo aplicado a un conjunto mayor.

<http://www.innovauoc.org/socialnet>

Identidad digital

SocialNet presenta toda la información que genera un autor, tanto interna al repositorio como en internet, una vez deposita un documento. Se trabaja el concepto de huella documental, que aglutina la producción del mismo, su red de colaboradores, y sus redes personales y profesionales, para de esta forma obtener una visión global del mismo.

Para ello se genera una ficha que recoge todo su activo digital disponible, y se le asignan automáticamente una o más instituciones, y un conjunto de palabras clave extraídas de los documentos producidos tanto por el autor como por sus colaboradores. Estas etiquetas representan el conocimiento que tiene un autor, bien por él mismo o bien por sus coautores a quienes se supone puede preguntar. Es decir, se le otorga una nueva dimensión en su identidad digital, ya que,

se le atribuye tanto el conocimiento real propio, como el potencial que puede obtener de su red de colaboradores. En la actualidad es tan importante identificar en la misma red “lo que se sabe” como “quién sabe lo que uno no sabe”.

En la ficha de cada autor (figura 5) se creó la sección “syndication” con el listado de documentos con los que está relacionado. Este apartado permite acceder a cada documento haciendo clic sobre su título. Para acabar de definir esta identidad digital las fichas tienen una pequeña biografía y la vinculación con las redes sociales externas, lo que permite obtener una visión completa del autor y de sus conocimientos.

Big data y open data

El problema a resolver no era tanto la capacidad del tratamiento masivo de información, sino las dificultades en la extracción de los datos en abierto y su homogeneidad. Como indica **Abril-Martorell** (2013), consejero delegado de *Prisa*, “lo más interesante del *big data* no es la cantidad de información a la que podremos acceder, sino en qué medida seremos capaces de utilizarla correctamente para aportar valor”.

Para resolver esta cuestión, se analizaron y compararon diferentes estándares como *Learning object metadata (LOM)*, *Sharable content object reference mode (Scorm)* y *Dublin core cualificado (DC)*. Sea decidió trabajar con *DC cualificado*, y se creó un sistema de notación y etiquetado en XML *DC* para marcar los campos de cada registro. Para ello se utilizó un cosechador (*harvester*) que confirmó la falta de estandarización de los datos que se encuentran de libre acceso en los repositorios documentales.

Una vez extraídos los datos, el trabajo se centró en poder tratarlos y visualizarlos en tiempo real.

Con el fin de poder visualizar conjuntos de datos complejos (por ej., cuando hay muchos coautores o muchos documentos) se podrían solapar líneas y nodos, se diseñaron funciones que permiten ajustar los parámetros básicos del sistema de visualización.

Visualización de información

El sistema permite generar visualizaciones complejas de datos de repositorios documentales. Los usuarios pueden *narrarse su propia historia* navegando por los autores y documentos de forma interactiva.

Al introducir un nombre de autor la cosechadora rastrea en los repositorios agregados, y una vez hallados los trabajos se vuelcan en la base de datos. Seguidamente se extraen los coautores de los que a su vez se buscan los trabajos. De esta manera, la red del sistema se va retroalimentando y crece paulatinamente a medida que se introducen nuevos documentos.

El componente principal del sistema es un visor de grafos, un espacio hiperbólico que se autoordena y dimensiona para que los elementos representados no se solapen, en el cual aparecen dos tipos de nodos: autor y documento, así como las relaciones entre ellos (figura 6).

Los nodos de autor tienen un círculo con un grosor de línea proporcional a la cantidad de documentos que ha escrito. Los nodos de documento se representan por un símbolo con un grosor de borde proporcional a la cantidad de autores con que están relacionados. Los colores indican la procedencia del documento. Al abrir un nodo se despliega el conjunto de conexiones con los que está relacionado y se dibujan automáticamente las relaciones con otros nodos en la red.

Al pasar el ratón sobre un nodo aparecen iconos adicionales que permiten esconder el nodo al visor, o en el caso de los autores, acceder a la ficha de información detallada. También aparece una ventana flotante con una información mínima sobre cada nodo, que ayuda a reconocerlo.

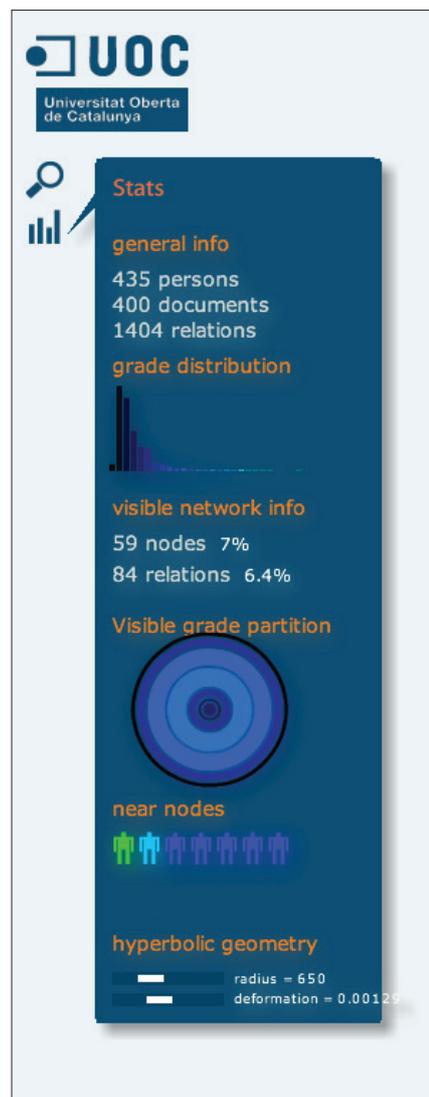


Figura 7. Panel de indicadores de red

Estadísticas y detalles de la subred de autores visualizada en la figura 6. El panel (que aparece haciendo clic en el pequeño icono en forma de diagrama de barras) contiene las siguientes secciones:

- General info: Número de autores, documentos y relaciones de la totalidad de la red tratada.
- Grade distribution: diagrama de barras representando la distribución de autores según el grado de relaciones (primer coautor, coautor del coautor, etc.).
- Visible network info: número y proporción de nodos, y relaciones visibles en pantalla.
- Visible grade partition: Gráfica radial con el nº de nodos (autores) según el grado o distancia al autor.
- Near nodes: listado de nodos más próximos.
- Hyperbolic geometry: controles del radio de la representación visual y de deformación de la geometría para centrar la parte que se desea visualizar.

Disponibilidad del producto

La investigación y el desarrollo realizados en torno al proyecto *SocialNet* para el tratamiento de datos de repositorios de información y visualización de relaciones autores-documentos, nos permiten abordar nuevos modelos de explotación de los repositorios, tanto para generadores como para distribuidores de información. Así, por ejemplo, *SocialNet* ayudó a la empresa *Bestiario* a diseñar usos de su producto

Quadrigram, una herramienta de gestión de grandes datos y datos abiertos, y diseño de narraciones visuales.

SocialNet se diseñó como un modelo ad hoc para estudiar y analizar la generación de redes sociales en unos repositorios concretos, pero podría aplicarse a otros como *E-LIS* o *Dialnet*, por ejemplo, y de esta manera ofrecer nuevos modelos de visualización de las redes de información subyacentes en esos repositorios.

Sin embargo, hacer una adaptación o un nuevo diseño a partir de la estructura básica de *SocialNet* saldría muy caro, pues el modelo descrito en este artículo estaba dirigido a una explotación muy concreta. Sería recomendable acudir a la citada empresa *Bestiario*, desarrolladora de *Quadrigram*, producto que puede ser utilizado por cualquier repositorio que iniciase un nuevo proyecto sobre sus datos. *Quadrigram* se ofrece con unas tarifas bastantes asequibles, como se describe en su web, y también cuenta con asesoramiento especializado de sus desarrolladores para orientar las posibilidades y alcance de un proyecto.

4. Conclusiones

Se vislumbra un gran recorrido para los profesionales de la información tanto en el campo académico como en el profesional y empresarial, con diferentes especialidades y campos de conocimiento emergentes, pero para ello hay que estar atento a la evolución de las tecnologías como por ejemplo visualización de datos, identidad digital, datos abiertos, grandes conjuntos de datos, y otras que a buen seguro aparecerán.

Los bibliotecarios-documentalistas actuales están familiarizados con las bases de datos y los repositorios documentales, pero deberían pensar en dotarlas de valor añadido mediante sistemas de *narrativa digital*: infografías y sistemas de visualización de datos para que las consultas por parte de los usuarios finales resulten más eficaces, y sea más fácil sacar conclusiones. De esta forma a partir de los conjuntos de datos se extrae información, que a su vez deriva en conocimiento colectivo.

SocialNet es un ejemplo de cómo se pueden interconectar los datos para dotarlos de un nuevo valor informativo. Destaca su capacidad de organizar visualmente una red de coautoría, de ofrecer información estadística tanto individual como colectiva, ayuda a verificar el conocimiento que tiene un autor, sus campos de conocimiento y los de sus pares.

5. Bibliografía

Abril-Martorell, Fernando (2013). Grandes datos, grandes historias, julio.
<http://toyoutome.es/blog/grandes-datos-grandes-historias/22218>

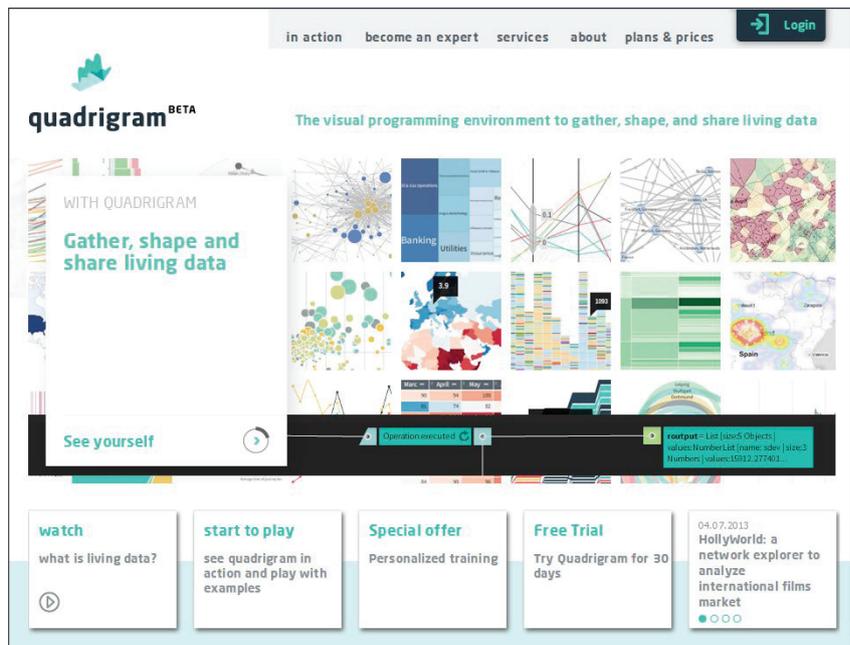


Figura 8. *Quadrigram*
<http://www.quadrigram.com>

Accenture (2012). Accenture technology vision.
<http://www.accenture.com/es-es/Pages/insight-accenture-technology-vision-2012.aspx>

BBVA Innovación (2013). Innovadata: Periodismo de datos e innovación, mayo.
<https://www.centrodeinnovacionbbva.com/blogs/planta29/posts/22047-innovadata-periodismo-de-datos-e-innovacion>

Berners-Lee, Tim (2009). "The next Web of open, linked data". *TED*, February.
https://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html
<http://www.w3.org/DesignIssues/LinkedData.html>

Cantera, José (2013). Descodificación del ADN de la experiencia de cliente del ecommerce en España.
<http://www.kpmgdigital.es/?p=701>

Dans, Enrique (2011). "Big Data: una pequeña introducción". *El blog de Enrique Dans*, 19 octubre.
<http://www.enriquedans.com/2011/10/big-data-una-pequena-introduccion.html>

Deloitte (2012). *Predicciones de Deloitte para el sector de tecnología, medios de comunicación y telecomunicaciones*
[http://www.deloitte.com/assets/Dcom-Mexico/Local%20Assets/Documents/mx\(es-mx\)TMT2012_Esp.pdf](http://www.deloitte.com/assets/Dcom-Mexico/Local%20Assets/Documents/mx(es-mx)TMT2012_Esp.pdf)

Díaz-Delgado, Natalia (2009). *Narrativa digital*.
<http://narrativa--digital.blogspot.com.es>

Ferrer-Sapena, Antonia; Sánchez-Pérez, Enrique A. (2013). "Open data, big data: ¿hacia dónde nos dirigimos?". *Anuario ThinkEPI*, v. 7, pp. 150-157.

Gantz, John F. et al. (2008). *The diverse and exploding digital universe. An updated forecast of worldwide information growth through 2011*. An IDC white paper, sponsored by EMC, 16 pp., March 2008.
<http://www.emc.com/collateral/analyst-reports/diverse->

exploding-digital-universe.pdf

Gartner (2013). *Gartner identifies the top 10 strategic technology trends for 2013*.

<http://www.gartner.com/newsroom/id/2209615>

Giones-Valls, Aina; Serrat-Brustenga, Marta (2010). "La gestión de la identidad digital: una nueva habilidad informacional y digital". *BID: textos universitarios de biblioteconomía i documentació*, juny, n. 24.

<http://bid.ub.edu/24/giones2.htm>

Herrera, Raquel (2011). "Propuesta de estrategias para desarrollar una taxonomía en narrativa digital". *Hipertext.net*, n. 9.

<http://www.upf.edu/hipertextnet/numero-9/taxonomia-narrativa-digital.html>

IDC-EMC (2011). *Digital universe study. Extracting value from chaos*.

<http://www.emc.com/leadership/programs/digital-universe.htm>

KPMG (2000). *Knowledge management research report*.

http://www.providersedge.com/docs/km_articles/kpmg_km_research_report_2000.pdf

Kusnetzky, Dan (2010). "What is "Big Data"?. *ZDNet*, 16 febrero.

<http://www.zdnet.com/blog/virtualization/what-is-big-data/1708>

Martínez-López, Héctor (2012). "The next big thing". *El blog d'UPCnet*, 27 enero.

<http://blog.upcnet.es/the-next-big-thing>

Mazzo-Iturriaga, Rodrigo (2010). *Linked open data: qué es y ejemplos en el mundo*

<http://www.bcn.cl/de-que-se-habla/open-data-link-data>

Peset, Fernanda; Ferrer-Sapena, Antonia; Subirats-Coll, Imma (2011). "Open data y linked open data: su impacto en el área de bibliotecas y documentación". *El profesional de la información*, marzo-abril, v. 20, n. 2, pp. 165-173.

<http://dx.doi.org/10.3145/epi.2011.mar.06>

Shirky, Clay (2012). "Cooperation and corruption". *Crooked Timber*.

<http://crookedtimber.org/2012/07/02/cooperation-and-corruption>

Directorio EXIT (EXpertos en Tratamiento de la Información)

<http://directorioexit.info>

The screenshot shows the profile page for Julián Marquina Arenas on the Directorio EXIT website. The page includes a header with the 'exit' logo and '2875 expertos'. Below the header is a search bar and a navigation menu. The profile section features a photo of Julián Marquina Arenas, his name, and a list of contact and professional details. At the bottom, there are links to search for him on various platforms and a footer with copyright information and site navigation.

exit 2875 expertos
Directorio de expertos en el tratamiento de la información

Julián Marquina Arenas

Exit ID: 1064
IraLIS: Marquina-Arenas, Julián ¿Qué es?
Institución: Baratz, Servicios de Teledocumentación SA
Dirección: Raimundo Fernández Villaverde, 28
Código postal: 28003
Ciudad: Madrid
Provincia: Madrid
País: España
Teléfono: +34-914 560 360
Móvil: +34-629 960 337
Correo-e: rechib@gmail.com
Correo-e personal: julian_marquina@yahoo.es
Web institucional: <http://www.baratz.es> Dirección válida
Web personal: <http://www.julianmarquina.es/> Dirección válida
Descriptor personal: Blogs and wikis, digital libraries, e-resources, open source
Especialidades: Comunicación; Documentación en general; Gestión de contenidos; Recuperación de información y búsquedas; Web 2.0 y redes sociales

Buscar en Google vCard
Buscar en Google Scholar XML
Buscar en Bing RDF
Buscar en Yahoo! similares

© Tomás Baiget & Josep-Manuel Rodríguez-Gairín, 2005-2013
Diseño de interacción: Jorge Franganillo
Español · English

Sobre Exit
Misión
Comité evaluador
Confidencialidad

Participa
Inscripción
Cooperaciones
Enlaza a Exit

Una ventana a las oportunidades profesionales

La nueva versión 2013 localiza la dirección postal en *Google maps* y permite buscar perfiles de expertos similares a uno dado.

PERCEPCIÓN MUSEOGRÁFICA DE LA BIBLIOTECA HISTÓRICA O PATRIMONIAL: PERSPECTIVAS Y REFLEXIONES EN TORNO A LOS FONDOS Y LIBROS ANTIGUOS

Manuel-José Pedraza-Gracia



Manuel-José Pedraza-Gracia, doctor en historia por la *Universidad de Zaragoza (UZ)*, fue bibliotecario de dicha universidad entre 1979 y 1989. Es profesor titular del *Área de Ciencias de la Documentación (UZ)*. Ha publicado diversos trabajos científicos en materias relacionadas con la historia del libro y de las instituciones documentales, libro antiguo y tasación y valoración del libro antiguo. Ha dirigido los cursos de verano de la *Universidad de Zaragoza* sobre tasación, valoración y comercio del libro antiguo.

<http://orcid.org/0000-0001-9686-9031>

Facultad de Filosofía y Letras, Universidad de Zaragoza
Depto. de Ciencias de la Documentación e Historia de la Ciencia
Pedro Cerbuna, 12. 50009 Zaragoza, España
pedraza@unizar.es

Resumen

Se analizan las relaciones existentes entre bibliotecas con fondos patrimoniales y museos. Se estudian especialmente las características similares del patrimonio bibliográfico y los objetos patrimoniales conservados en los museos, de las condiciones de uso y de las características de los usuarios, de la gestión de estos fondos por parte de los profesionales. Finalmente se analizan las bibliotecas que únicamente poseen este tipo de fondos.

Palabras clave

Biblioteca patrimonial, Biblioteca histórica, Museos, Patrimonio bibliográfico, Libro antiguo.

Title: Museographic view of history or cultural heritage libraries: perspectives and reflections about collections and old books

Abstract

The paper analyses the similarities between cultural heritage libraries and museums. First, books are analyzed as heritage objects and museum artifacts. Secondly, researcher's needs and interests concerning rare books are studied, as well as the knowledge and skills a librarian must possess. Finally, main functions and special features of heritage libraries are outlined.

Keywords

Heritage libraries, Museums, Bibliographical heritage, Rare books, Old books.

Pedraza-Gracia, Manuel-José (2013). "Percepción museográfica de la biblioteca histórica o patrimonial: perspectivas y reflexiones en torno a los fondos y libros antiguos". *El profesional de la información*, septiembre-octubre, v. 22, n. 5, pp. 440-447.

<http://dx.doi.org/10.3145/epi.2013.sep.09>

Introducción

Los objetos y bienes que constituyen los patrimonios de los pueblos se caracterizan por poseer un conjunto de valores de tan singular significación que llegan a conformar la identidad de esos pueblos (Arjona, 2003; Abranches, 1988; Pedraza-Gracia, 2008; Racedo et al., 2004; Unesco, 2000; 2009). Este principio -que se ha fijado claramente en la legislación del Estado Español, en alguna de las normativas de la Comunidades Autónomas como Navarra, La Rioja, País Vasco o Cataluña y en las de Iberoamérica como Perú o Venezuela, entre otras- afecta a objetos y bienes de muy diverso tipo,

algunos de los cuales se custodian en los diferentes centros de información: museos, archivos y bibliotecas, e incide de diversas maneras en ellos. Las bibliotecas como custodios y como constituyentes de ese conjunto de objetos y bienes también se ven afectadas por este principio: desde el objeto -el libro o la colección-, desde el usuario -el investigador o el ciudadano interesado-, desde el bibliotecario y sus formas de actuación y gestión, y desde la propia biblioteca. Si se analizan las soluciones que se adoptan en las bibliotecas con fondos patrimoniales se aprecia que muchas de ellas son muy próximas, si no son coincidentes, con las que afec-

tan a la pieza museística, al usuario, al conservador y al museo.

Sin embargo, esta proximidad no suele ser advertida y no existen trabajos científicos desde la biblioteconomía y la documentación que pongan en relación las bibliotecas que poseen fondos patrimoniales con los centros museísticos, lo que no quiere decir que no se hayan estudiado desde otras perspectivas.

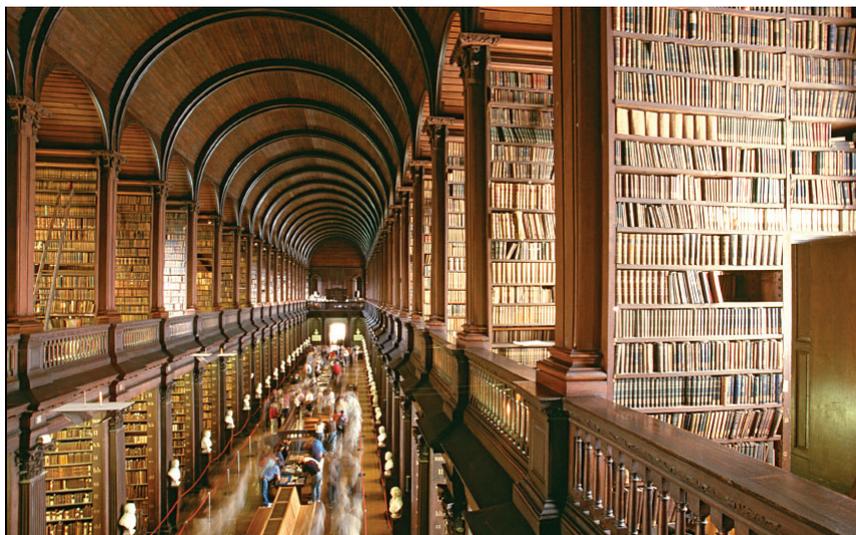
El interés de este trabajo se encuentra en el análisis, desde una posición exclusivamente teórica, de esta posible relación entre ambos tipos centros de información: biblioteca patrimonial, histórica o de reserva, con fondos históricos, conventuales, de reserva, preciosos curiosos o raros... y museo; especialmente por lo que hace referencia a su promoción y difusión.

El libro antiguo

Cuando la biblioteca conserva entre sus colecciones y fondos conjuntos o unidades documentales de cierta antigüedad –lo que puede derivar en la posesión de valores patrimoniales contrastables-, se suele producir una disfunción entre este conjunto bibliográfico y el resto de los materiales conservados en la misma. Estas disfunciones tienen como fundamento las diferencias existentes entre el valor y uso del documento bibliográfico contemporáneo y el valor y uso de los documentos bibliográficos antiguos.

El documento editado fabricado manualmente puede ofrecer diferencias entre los distintos ejemplares de la edición. La existencia de conjuntos de ejemplares de la misma edición denominados, de forma específica, emisiones y estados tiene su origen en una casuística compleja. Si estas diferencias han sido planificadas intencionadamente, la variante recibe el nombre de emisión y si no ha sido planificada intencionalmente, el de estado (**Moll**, 1979; **Clemente**, 2003, pp. 251-253; **Martín-Abad**, 2004a, pp. 53-59, pp. 81-85). La identificación de la edición es aún más compleja si a la existencia de variantes se añade que la diferenciación entre las sucesivas ediciones –reediciones- no es sencilla debido a que su mención no suele ser relevante por lo que respecta a los contemporáneos del documento y no suele figurar de forma evidente en él. Todavía hay que añadir la existencia de ediciones ilegales, contrahechas, piratas, falsificadas y subrepticias (**Clemente**, 2003, p. 254). En consecuencia, en el proceso de producción del libro en el período de la imprenta manual se crean diferentes ediciones, no siempre sencillas de identificar, que poseen variantes, en ocasiones, complejas en su apreciación. Pero no por ello se debe obviar el trance imprescindible de la identificación (**Martín-Abad**, 2004b, pp. 6-12).

Independientemente de las vicisitudes de su fabricación, con el transcurso del tiempo, el libro puede sufrir las acciones del medio ambiente y las intervenciones del ser humano que introducen nuevos elementos diferenciadores: ataques



Long room, Old Library, Trinity College, Dublín

Foto: <http://www.thezoom.com/2012/08/5921>

biológicos, una encuadernación especialmente cuidada o rica, una indicación de posesión, la pérdida de integridad (motivada o no), la inclusión en un facticio, una anotación aclaratoria, una marca de fuego, el uso descuidado..., entre una casuística variadísima de actuaciones (superada la más drástica: la propia desaparición del documento). Así se obtienen unidades bibliográficas perfectamente diferenciables que pueden acabar denominándose por el lugar en el que son conservados o por el nombre del poseedor. La procedencia del ejemplar resulta uno de los elementos diferenciadores más esenciales y a la par uno de los principios de estudio de los fondos y bibliotecas (**Pearson**, 1998; **Martín-Abad**, 2004a; **Santos-Aramburo**; **Torres-Santo-Domingo**, 2004). Algunos de estos factores inciden también y de la misma manera en los bienes patrimoniales que se custodian en museos.

Existen diferencias entre el valor y uso del documento bibliográfico contemporáneo y el valor y uso de los documentos bibliográficos antiguos

Por todo lo mencionado, de aquí se puede extraer una nueva consecuencia: el libro antiguo no puede tener el mismo interés de búsqueda y uso que el libro moderno, aunque se encuentren en el mismo centro. El documento moderno como documento de investigación o consulta se busca y se usa por su contenido (científico, literario, artístico, de ocio...). El documento antiguo como documento de investigación o consulta presenta características específicas:

a) En muy escasas ocasiones el libro antiguo es buscado para su lectura, para el ocio o para el estudio del contenido como ocurre con los documentos modernos. Se busca en este caso, por tanto, la obra. Este tipo de actuación responde a obras poco conocidas, aún no editadas por autores modernos o incluidas entre los millones de ejemplares digitalizados disponibles en la Red. Quizás esto es lo único que puede justificar esta actuación, ya que en la mayoría de las ocasiones no resulta pertinente –ni recomendable- usar con

el único fin de su lectura una edición antigua de una obra cuyas ediciones modernas presentan complementos informativos que ayudan y mejoran su comprensión, aunque es cierto que en ese proceder existe un componente fetichista muy propio de la bibliofilia o, mejor, de la bibliomanía. No es comprensible usar una primera edición de las *Comedias* de Lope para su mera lectura o intentar estudiar en determinados niveles de aproximación a las ideas de Newton sus primeras ediciones. Sólo los expertos en determinadas materias se aproximan a estas ediciones príncipe con objeto de realizar análisis de los textos minuciosos tendentes a discernir mediante el estudio del medio los mensajes en toda su intención.

“ Más que una herencia de nuestros antepasados, los objetos patrimoniales son un préstamo de nuestros sucesores ”
(Tugores; Planas, 2006)

b) En otras ocasiones el ejemplar se busca como unidad que pertenece o perteneció a un conjunto (colección, edición, variante, obra de un encuadernador concreto, producto bibliográfico de una imprenta, libros censurados...) persiguiendo, por ejemplo: la identificación de una edición, el análisis minucioso del texto de una variante, la autenticación de una anotación, una procedencia específica, la anotación de un censor, la reconstrucción de una colección disgregada, la lectura de la parte tachada por el censor, la identificación de un ex libris... (todo ello mediante la comparación con otros ejemplares del mismo u otro conjunto, de manera principal). Una de estas actuaciones es la de adscribir un ejemplar a una edición específica, muy importante cuando se trata de libro antiguo, lo que se ha denominado la identificación del ejemplar ideal, definido perfectamente por **Moll** (1979) como “el que ofrece el estado más perfecto de la obra, según la intención del autor, editor e impresor”, que ha de permitir, por ejemplo, la correcta catalogación del documento. La identificación resulta aún más complicada en los impresos *sine notis*. De nuevo aquí se cuenta con la ayuda de la Red donde ya existe un número significativo de ejemplares digitalizados.

c) En otras se quiere estudiar un ejemplar determinado de una colección dotado de valores y de información estructural concreta que lo independiza y diferencia de todos los demás (sin que, como en el caso anterior, sea imprescindible que lo que busca el investigador sea el propio producto editorial). El ejemplar es estudiado por aquello que solamente él posee: una anotación marginal, una dedicatoria, un dibujo, un grabado adherido a las guardas, un cuaderno de anotaciones cosido, una miniatura... En este caso, aunque estuviese digitalizado hay aspectos que únicamente se pueden percibir desde el original. De la misma manera que hay aspectos de una obra pictórica, por ejemplo, que sólo se pueden apreciar ante ella y jamás ante una reproducción porque la materialidad de la obra trasmite también información y sentimiento; lo que no impide que se hayan creado herramientas gráficas e informáticas que facilitan el análisis pormenorizado de la

obra pictórica, aunque en su mayoría son herramientas de investigación presencial y no de uso. **Jan Bos** (2012, p. 18) ha puesto de manifiesto la existencia de una ausencia efectiva de identidad ante la investigación entre el documento original y la copia digital, lo que se percibe de manera iconográfica en “Ceci n'est pas une pipe” de René Magritt¹. Por esta causa, aquí no se pretende tratar las posibilidades que ofrecen los documentos digitalizados ni las bibliotecas digitales, que quedan para análisis posteriores.

Uso, conocimiento y transmisión

El patrimonio tiene como imperativo legal la transmisión. La identidad de un pueblo se sostiene en la existencia de un aglutinante, la cultura, que a su vez se sustenta en la existencia de bienes culturales, no siempre tangibles. Si desaparecen estos bienes desaparece ese aglutinante y con ello el pueblo como tal, como ya se argumentó en la introducción. Es ésta la raíz del genocidio cultural, ese crimen contra toda la humanidad que se produce cuando los pueblos que se apoderan de otro destruyen su patrimonio (arrasando las bibliotecas y archivos, penalizando, despreciando o prohibiendo el uso de una lengua o privándolo de sus obras artísticas, por ejemplo). Por esta misma causa, la legislación antepone la preservación al uso. Los investigadores en patrimonio aducen siempre una premisa: más que pensar que los objetos patrimoniales son una herencia de nuestros antepasados se debería tener presente que son un préstamo de nuestros sucesores (**Tugores; Planas, 2006**). El bibliotecario que trabaja con fondos antiguos no debería olvidarlo. La biblioteca con respecto al patrimonio bibliográfico tiene la obligación de activar ese patrimonio para su conocimiento y uso por parte del ciudadano. Se trata de dotar de su función social al patrimonio o, si se quiere, su uso social (**Prats, 1997, p. 27**).

“ Debido a las vicisitudes de su elaboración y en el transcurso del tiempo, el libro antiguo deviene en ejemplar único ”

La actividad de la biblioteca varía con respecto a los fondos patrimoniales. Por lo que respecta al libro moderno la biblioteca debe servir para cubrir las necesidades informativas, de formación, de ocio... de los usuarios; resulta evidente que la utilización del documento prevalece sobre la conservación, la rentabilidad de la inversión en la colección se mide por el grado de uso y satisfacción del usuario. En lo referente a los fondos y unidades documentales antiguos, la responsabilidad y el punto de mira varían de forma notable, la biblioteca es responsable del mantenimiento y conservación del bien patrimonial, de tal suerte que el uso está supeditado a la preservación. El foco cambia de manera radical.

La racionalización del uso de estos materiales en las bibliotecas se convierte entonces en un problema que se ha de resolver casi de forma cotidiana y que se presenta con dos vertientes: qué se usa y quién lo usa.

Es evidente que cuando un documento pueda resultar dañado por el uso deberá preservarse su utilización hasta ha-

ber encontrado la manera de superar los problemas que ocasionan los desperfectos y éstos mismos. También resulta evidente que, aunque sea de forma imperceptible, cualquier grado de uso supone una degradación del documento, por lo que se puede reservar la utilización del mismo a aquellos usuarios que puedan aportar o mejorar los valores del documento o de la colección (Serrai, 1992). El problema radica, por lo tanto, en resolver la paradoja presentada anteriormente: ¿cómo accede un ciudadano a un bien que por ser patrimonial le pertenece?

Los responsables deben potenciar el acceso al conocimiento y al uso de los bienes patrimoniales mediante exposiciones permanentes y temporales, préstamo para exposiciones, charlas y conferencias, jornadas de puertas abiertas, publicaciones, ediciones facsimilares y otros recursos similares. Y es aquí donde se encuentra otra similitud entre el museo y la biblioteca con fondo patrimonial. Existe un uso reservado a la investigación de los documentos, ya que la investigación genera nuevo conocimiento sobre el documento estudiado (en lo que respecta a sus contenidos o a sus aspectos formales o materiales) que se supedita a unas reglas de uso estrictas encaminadas a la preservación del documento y que el investigador conoce. Pero se debe potenciar de forma decidida el conocimiento del patrimonio propio de los legítimos propietarios, que sepan de los valores patrimoniales del mismo, independientemente de su nivel de preparación y conocimiento del documento, facilitando su interpretación, lo que también es un mandato legislativo. De aquí se debe deducir que no solamente se atenta contra el patrimonio impidiendo su investigación, sino que también es un atentado el impedir su conocimiento y disfrute mediante el acceso a las actividades de difusión que informen sobre el mismo, que pongan de manifiesto su valor cultural.

En este sentido las exposiciones se muestran como uno de los mejores métodos: acercan el documento al ciudadano, le proporcionan información sobre sus valores, su trascendencia, su autor, sus características..., en suma, proporcionan un conocimiento que sólo puede redundar en aumento de interés por el documento y por su investigación.

El investigador es quien se aproxima al libro o a la biblioteca con un conocimiento del mismo de alto nivel (de la misma manera que el investigador se aproxima a una pieza en el museo), conoce su procedencia y su génesis, sabe cómo se trata el objeto y reconoce su importancia. Es un usuario profesionalizado, cuya actividad posee trascendencia para el centro, ya que sus aportaciones sobre las piezas generan nuevo conocimiento. Su producción (científica) aporta siempre nuevo(s) valor(es) añadido(s) a los documentos o piezas que posee el centro, lo que conduce a que esa producción científica sea imprescindible para el centro de investigación. De esta manera se produce un efecto dinamizador del centro

y de la investigación, ya que estos nuevos conocimientos y valores añadidos producen un efecto llamado sobre otros investigadores que, a su vez, aportarán nuevo conocimiento y valores al documento y al centro, y así sucesivamente.

El libro antiguo requiere un proceso de identificación imprescindible para poder llevar a cabo un tratamiento correcto

Junto a este usuario investigador hay también un ciudadano interesado, un usuario que ansía conocer, no profesionalizado, que no suele tomar un contacto directo (físico) con el documento bibliográfico o la pieza pero que es el objeto del centro, es el destinatario de la función social del mismo, asiste a conferencias y a las exposiciones, a las jornadas de puertas abiertas, se interesa por los catálogos de exposiciones... y, con ello, se forma e informa. Es este usuario el que proporciona sentido al centro.

Especial mención merecen los jóvenes y los niños que han de ser especialmente cuidados, deben organizarse actividades propias para esas edades y visitas especialmente preparadas, con zonas lúdicas relacionadas con el centro, el libro antiguo y con recursos específicos; también en las bibliotecas históricas o patrimoniales. No obstante, se debe diferenciar estructuralmente entre los espacios y funciones relacionados con la investigación, con los dedicados a la difusión y el conocimiento, y con los dedicados a los procesos.

No puede existir un tipo de usuario sin el otro. La ley determina como una de las finalidades principales del centro que custodia bienes patrimoniales que el ciudadano los (re) conozca y disfrute como propios. Pero también es esencial que un centro de información propicie la investigación sobre sus fondos patrimoniales como método de valoración de los mismos y de difusión de los nuevos conocimientos que se obtienen como fruto de esas investigaciones.



Biblioteca General Histórica, Universidad de Salamanca



Caja de conservación (Tacón-Claváin, 2010a)

El bibliotecario

La actividad del profesional ante los libros antiguos también posee características específicas. Son tantas y tan trascendentes las diferencias que existen entre los documentos impresos antiguos y los contemporáneos que forzosamente requieren que el profesional que se encarga de ellos haya de poseer una especialización (Richard, 2010). Por esta causa, sería necesaria una reflexión desde las instituciones que se dedican a la formación en información y documentación sobre si ofrecen las competencias y aptitudes que se precisan en estos centros. La vertiente patrimonial es demandada y tiene mucha aceptación entre los estudiantes de los centros en los que se ha cuidado y ofertado.

Las labores especializadas de selección y adquisición no son iguales a las de los libros contemporáneos, puesto que requieren el conocimiento de un mercado complejo en el que se adquieren ejemplares específicos de una edición, que presentan particularidades que hacen preferible a uno sobre los otros.

Aspectos elementales como el sellado, colocación de tejuelos o ubicación del documento en la estantería, en aras de la preservación, ofrecen algunas dificultades que no presentan los libros modernos, hasta incluso la incorporación del libro en cajas especialmente preparadas para su protección. Por todo ello son precisos conocimientos sobre preservación.

Tampoco el tratamiento es el mismo, el libro antiguo requiere una identificación previa e imprescindible: es preciso conocer qué es lo que se tiene delante para poder efectuar un tratamiento correcto. Los elementos que ofrecen ese conocimiento en los documentos antiguos pueden estar muy escondidos en el propio documento o requerir herramientas bibliográficas específicas. Sin esta actividad, que apenas afecta al libro contemporáneo, resulta imposible llevar a efecto cualquier actuación sobre un libro antiguo (restaurar, catalogar, tasar...). Se trata de una actividad múltiple y con diversas facetas que afecta a la obra, a la edición, a la variante y a aspectos formales como la encuadernación o las marcas de propiedad, entre otros. La catalogación exige, además de las normas específicas, un conocimiento profundo del producto bibliográfico antiguo y de los múltiples aspectos formales ya mencionados.

La difusión también es diferente, como se ha visto anteriormente. Hay que dar a conocer estos documentos al ciudadano y a los investigadores mediante muy diversos métodos: la reproducción del documento para investigación, preservación y difusión, la publicación de trabajos de investigación sobre los fondos del centro y, especialmente, las exposiciones, tanto en el mismo centro como el préstamo para exposiciones externas, son actuaciones que se han mostrado muy eficientes. Todas estas actividades exigen conocimientos relativos a seguros, sobre aspectos técnicos de la edición, de la reproducción digital... (Mouren, 2012).

De la misma manera que hay que prever que el conjunto de actuaciones llevadas a cabo en la biblioteca para disponer un libro en un estante a la espera de su uso tiene que adecuarse al máximo grado de atención a su conservación, la metodología del uso del documento por parte del investigador (incluida la suspensión del uso de un ejemplar por parte de los investigadores), los métodos de difusión (como la denegación de la exposición de un ejemplar), etc., también se han de sujetar a la preservación del documento como patrimonio y como constituyente de un conjunto constitutivo del centro. Por esta causa, las decisiones adoptadas por los especialistas en preservación se convierten en vinculantes y preferentes sobre las que puedan tomar los propios bibliotecarios en numerosas actividades: tiempos de exposición, uso en sala, almacenamiento... (Tacón-Claváin, 2010a; 2010b; 2011a; 2011b).

“ Los bienes patrimoniales sólo se pueden comprender en su integridad y adquirir pleno sentido si están contextualizados ”

La biblioteca

También se ha de tener en cuenta un principio básico: si un objeto patrimonial se analiza en su contexto, aumenta el conjunto de valores que atesora, porque los bienes patrimoniales sólo se pueden comprender en su integridad y adquirir su pleno sentido si están contextualizados.

El documento, salvo que no se haya utilizado ni movido del lugar en el que se depositó tras su fabricación (lo que lo individualizaría por extraño), posee una “vida” y, en su paso de propietario en propietario, de lector en lector, y de depósito documental en depósito documental, se van produciendo interacciones entre el lector, el propietario, el depósito documental y el documento, llegando al momento actual, tras estas actuaciones, modificado y frecuentemente relacionado con un contexto específico. El análisis de esta trayectoria vital del libro es esencial para reconocer la historia del ejemplar y la de las bibliotecas. Si se extrae el ejemplar de ese contexto, se pierde la posibilidad de identificar los factores que intervinieron en hacer que el documento sea en la actualidad como es. Además, cada uno de estos ejemplares ha tenido una función y un porqué dentro de una colección. De esta manera, se puede hablar de una relación íntima entre el documento y la colección que determina la mutación de ambos mediante el acomodo del uno en la otra y de la colección como receptora del documento.

La legislación establece una división de gran interés con respecto al patrimonio bibliográfico: las unidades documentales que poseen unas características especiales se incluyen de forma directa dentro del conjunto de bienes patrimoniales, pero también se incluyen los conjuntos bibliográficos pertenecientes al ámbito público tengan o no esas características. El responsable del patrimonio debe entender que el contexto, el conjunto, la colección (o mejor la biblioteca) es un único bien patrimonial indivisible que se comprende y se entiende así, reunida, preferiblemente en el ámbito espacial en el que se desarrolló. Ya no se trata de analizar el conocimiento almacenado en una obra específica, por muy puntera que sea, ni de analizar una signo de propiedad, ni de observar una bella encuadernación o un grabado determinado; se trata de ver, conocer y comprender qué características posee una biblioteca conventual en el siglo XVIII en un espacio determinado, qué lee la nobleza en el XVIII o qué interés y conformó la biblioteca de un médico en el XVI, qué otros objetos poseía unidos a la biblioteca, qué autores leía, en qué idiomas...

Para tratar correctamente el libro antiguo, el bibliotecario y el investigador deben contar con una bibliografía de apoyo que les ayuden en su labor

Existen bibliotecas que adoptan de forma consciente una posición museográfica conjugándola con la propia bibliotecaria. Son perfectamente conocidas e identificables esas bibliotecas que se pueden denominar bibliotecas-museo. En ellas el contexto y el continente se imbrican en un todo que se transforma en el objeto de estudio y de difusión. La protección patrimonial se amplía a todos los elementos constitutivos de la biblioteca. En estos casos la extracción de alguno de los componentes representa una verdadera amputación de valores patrimoniales, por mucho que se pueda percibir el interés del conjunto. Es evidente que el mostrar “en su contexto” los materiales patrimoniales resulta notablemente complicado en el caso de los libros antiguos, ya que prima la conservación, pero en los casos en los que ha sido posible mantener un nivel de conservación aceptable, teniendo presente que los fondos existentes en la actualidad son también fruto de una historia, es lo deseable y donde se ha de poner el máximo esfuerzo.

Trasciende, por tanto, la información puntual contenida en uno o en varios volúmenes; su alcance (que bien puede identificarse con el conjunto de valores) se ha expandido al conjunto de documentos en su medio ambiente, siendo éste el que hay que proteger. He aquí una nueva relación entre las bibliotecas patrimoniales y los museos: en ellos los objetos patrimoniales se analizan y muestran, siempre que se puede, en su contexto.

No se deberían confundir las bibliotecas patrimoniales, ni siquiera las bibliotecas-museo con los denominados museos del libro o de la imprenta: no sólo las finalidades, sino también los objetos y conocimientos que se pretende transmitir son completamente diferentes. Los museos de la imprenta y del libro muestran cómo se hace el libro, como dice **Grego-**

rio: “un museo del libro nasce per esporre al pubblico questo singolare ‘oggetto’, mostrandone le più segrete sfaccettature e il progressivo divenire nel tempo; portando sotto i riflettori le tecniche di produzione e le innovazioni grafiche, i protagonisti della storia editoriale e le trasformazioni avvenute nelle pratiche di scrittura, di lettura e di mercato” (2006). Baste como ejemplo el de la Comunidad Valenciana situado en El Puig, mientras que las bibliotecas históricas o patrimoniales muestran cómo era el conocimiento en un momento dado.

Un objeto de estudio que presenta esta cierta complejidad exige que el bibliotecario y el investigador tengan acceso a una bibliografía de apoyo, y a unas herramientas de trabajo que les ayuden en su labor, de acceso directo y sin préstamo, en lo que **Carrión-Gútiérrez** (1998, p. 37) denominó como “biblioteca de presencia” (Prasenzbibliothek). La identificación de una edición, de un estilo en la encuadernación, un ex-libris, una firma, la confirmación de la integridad del ejemplar, la existencia de una falsificación, la facsimilización de una página o páginas, la identificación de un pastiche... son aspectos que se deben conocer antes de proceder al tratamiento del volumen concreto. La adquisición de un ejemplar requiere conocer la existencia de otros en la proximidad, las características del mismo, la licitud de su procedencia, la adecuación de su precio al precio de mercado... El préstamo para una exposición sólo se puede realizar si previamente se han evaluado los riesgos (de transporte, de exposición, de reproducción para el catálogo...), se ha estudiado el precio del ejemplar, se ha analizado su valor para la biblioteca y la posibilidad de una reposición en caso de pérdida... Estas actividades requieren unas herramientas, unos instrumentos imprescindibles que conforman una biblioteca de apoyo, que suele ser de uso compartido por parte de los bibliotecarios y de los investigadores, que también precisan estas mismas herramientas. Este fondo contiene también aquellas obras que tratan sobre el centro y sus colecciones, sobre su especialización... por una parte; y sobre el libro antiguo en general, su confección, evolución y estructura... por otra. Fondos similares para el apoyo a los investigadores y técnicos son imprescindibles también en los museos.

El contexto es fundamental para comprender la unidad bibliográfica

En esa conjugación de las bibliotecas con fondos patrimoniales y los museos, se hallan otras identidades: cada vez más las grandes bibliotecas cuentan con tiendas especializadas como los museos y, también, cada vez más, se pueden encontrar museos especializados en las bibliotecas que presentan exposiciones permanentes y temporales que tienen como objeto principal los fondos y colecciones de las propias bibliotecas.

A modo de conclusión

Las similitudes entre las bibliotecas con fondos antiguos y los museos son notables, existen muchos lugares comunes y muchos puntos de proximidad. La separación en el mundo académico no debe ser obstáculo para propiciar la bús-

queda de sinergias de las que ambos mundos pueden salir beneficiados.

El objeto de trabajo de ambas instituciones pertenece al ámbito patrimonial, lo que ofrece como resultado fines y objetivos similares. De hecho, el documento antiguo como documento de investigación o consulta presenta características específicas propias diferentes de las de los libros modernos.

El libro antiguo debido a las vicisitudes de su elaboración y el transcurso del tiempo deviene en ejemplar único, perfectamente diferenciable de los que con él formaron la edición y, por ello, identificado dentro de la colección a la que pertenece. Esta situación se produce igualmente en los objetos conservados en el museo.

De la misma manera que ocurre cuando los objetos custodiados en museos poseen referentes modernos, las diferencias que existen entre los documentos impresos antiguos y los contemporáneos exigen un profesional que posea una especialización en este tipo de fondos.

El libro antiguo requiere un proceso de identificación imprescindible ya que es preciso conocer qué es exactamente lo que se tiene delante para poder llevar a cabo un tratamiento correcto. Este mismo proceso identificativo se produce con las piezas museísticas. Son procesos que afectan a diversos aspectos del libro: textual, material, procedencia...

El contexto es fundamental para comprender la unidad bibliográfica, de hecho, en algunas ocasiones lo que se posee no es un libro sino una biblioteca completa, en cuyo caso se amplía la protección a todo el conjunto y al continente. El contexto es esencial como en el resto de los objetos patrimoniales.

Como se puede observar, en casi todas las particularidades mencionadas se encuentran similitudes con los museos: preponderancia de la preservación, identificación de los documentos, similitud de los mercados anticuarios, necesidad de la identificación previa, convergencia de las técnicas y métodos de difusión de los fondos.

Nota

1. Magritte pintó una pipa y debajo escribió "Ceci n'est pas une pipe" ("esto no es una pipa"), lo que parece una contradicción, pero es realmente cierto: la pintura no es una pipa, es la imagen de una pipa. Cuando a Magritte le preguntaron sobre esta imagen, respondió: "por supuesto no es una pipa: sólo trate usted de llenarla con tabaco".

Bibliografía

Arjona, Martha (2003). *Patrimonio cultural e identidad nacional*. 2ª ed. La Habana: Boloña. ISBN: 9597126133

Abranches, Henrique (1988). *Identidad y patrimonio cultural*. La Habana: Ciencias Sociales.

Bos, Jan (2012). "All books are equal, but some books...: towards a modern vision of special collections". En: Mouren, Raphaële (ed.). *Ambassadors of the book: competences and training for heritage librarians*. The Hague: International Federation of Library Association and Institutions, De

Gruyter, Saur. (IFLA Publications, 160), pp. 15-23. ISBN: 978 3 11 030127 4

Carrión-Gútiérrez, Manuel (1998). "Tres lecciones sobre la Biblioteca Nacional de España". En: Rodríguez-Álvarez, Ramón; Llordén-Miñambres, Moisés (eds.). *El libro antiguo en las bibliotecas españolas*. Oviedo: Universidad de Oviedo, Servicio de Publicaciones, pp. 9-38.

Cataluña (1993). "Ley 9/1993, de 30 de septiembre, del Patrimonio cultural catalán". *Diari oficial de la Generalitat de Catalunya*, n. 1807, pp. 6.748-6.758.

<http://portaldogc.gencat.cat/utillsEADOP/PDF/1807/46350.pdf>

http://noticias.juridicas.com/base_datos/CCAA/ca-I9-1993.html

Clemente, Yolanda (1993). "Análisis, identificación y descripción analítica del libro antiguo: las ediciones y sus variantes: emisiones y estados". En: Pedraza-Gracia, Manuel-José, Clemente-San-Román, Yolanda; De-los-Reyes-Gómez, Fermín. *El libro antiguo*. Madrid: Síntesis, pp. 249-261. ISBN: 84 9756 153 8

Gregorio, María (2006). *Imago libri. Musei del libro in Europa*. Milano: Sylvestre Bonnard. ISBN: 8889609176

La Rioja (2004). "Ley 7/2004, de 18 de octubre, de Patrimonio cultural, histórico y artístico de La Rioja". *Boletín oficial de La Rioja*, n. 136.

<http://goo.gl/X4qoi>

Martín-Abad, Julián (2004a). *Los libros impresos antiguos*. Valladolid: Universidad de Valladolid. ISBN: 978 84 8448 279 6

Martín-Abad, Julián (2004b). *La valoración del libro: el punto de vista del bibliotecario de fondo antiguo*. Documentos de trabajo UCM. Biblioteca Histórica, n. 8. pp. 1-25.

<http://eprints.ucm.es/5698>

Moll, Jaime (1979). "Problemas bibliográficos del libro del Siglo de Oro". *Boletín de la Real Academia Española*, v. 59, n. 216, pp. 49-107.

<http://www.cervantesvirtual.com/obra/problemas-bibliograficos-del-libro-del-siglo-de-oro--0>

Mouren, Raphaële (ed.) (2012). *Ambassadors of the book: competences and training for heritage librarians*. The Hague: International Federation of Library Association and Institutions, De Gruyter, Saur (IFLA Publications, 160). ISBN: 978 11 030127 4

Navarra (2005). "Ley foral 14/2005, de 22 de noviembre, del patrimonio cultural de Navarra". *Boletín oficial de Navarra*, n. 141.

<http://www.lexnavarra.navarra.es/detalle.asp?r=4509>

País Vasco (1990). "Ley 7/1990, de 3 de julio, de patrimonio cultural vasco". *Boletín oficial del País Vasco*. 157, pp. 7062-7092.

<http://goo.gl/sR7HW>

Pearson, David (1998). *Provenance research in book history: a handbook*. London: The British Library. ISBN: 0 7123 0344 8

Pearson, David (2012). *Books as history: the importance of books beyond their texts*. New Castle, Oak Knoll Press. ISBN: 978 1584563150

Pedraza-Gracia, Manuel-José (2008). "La gestión del patrimonio bibliográfico español y el catálogo colectivo de patrimonio bibliográfico: un ejemplo de interrelación entre la administración y el ciudadano". En: Rover, Aires José (ed). *Inclusão digital e governo eletrônico*. Zaragoza: Prensas Universitarias de Zaragoza, Lefis series v. 3, n. 9, pp. 217-246. <http://zaguan.unizar.es/record/4170>

Perú (2004). *Ley general del patrimonio cultural de la nación*. Ley n. 28296. <http://www.ilam.org/ILAMDOC/ley28296%20%20380.pdf>

Prats, Llorenç (1997). *Antropología y patrimonio*. Barcelona, Ariel. ISBN: 978 84 344 42211 7

Racedo, Josefina; Requejo, María-Isabel; Segura, Zulma; Taboada, María-Stella (2004). *Patrimonio cultural e identidad*. Buenos Aires: Cinco/Cercapu. ISBN: 9509693685

Richard, Hélène (2010). *La formation aux questions patrimoniales dans les bibliothèques*. Paris: Ministère de l'Enseignement Supérieur et de la Recherche. Rapport n. 2010-016. http://media.enseignementsup-recherche.gouv.fr/file/2010/52/1/Formation_aux_questions_patrim_def_166521.pdf

Santos-Aramburo, Ana; Torres-Santo-Domingo, Marta (2004). "La Biblioteca Histórica de la Universidad Complutense: una primera aproximación a sus procedencias": En: *La memoria de los libros: estudios sobre la historia del escrito y de la lectura en Europa y América*. Salamanca: Instituto de Historia del Libro y la Lectura, Tomo II, págs. 265-286. ISBN: 978 84 9335 044 4 http://eprints.ucm.es/9836/1/Procedencias_BH.pdf

Serrai, Alfredo (1992), "Le biblioteche storiche". *Il bibliotecario*, n. 33/34, pp. 1-11.

Tacón-Clavaín, Javier (2010a). "Cajas de conservación para libros". *Documentos de trabajo UCM. Biblioteca Histórica*, 2010/4. <http://eprints.ucm.es/10637>

Tacón-Clavaín, Javier (2010b). "El uso de atriles para la apertura de libros antiguos". *Documentos de trabajo UCM. Biblioteca Histórica*, 2010/5. <http://eprints.ucm.es/10638>

Tacón-Clavaín, Javier (2011a). "Procedimiento para la manipulación de códices manuscritos durante la consulta". *Documentos de trabajo UCM. Biblioteca Histórica*, 2011/6. <http://eprints.ucm.es/12460>

Tacón-Clavaín, Javier (2011b). "Procedimiento para la apertura de libros intonsos de fondo antiguo". *Documentos de trabajo UCM. Biblioteca Histórica*, 2011/7. <http://eprints.ucm.es/12472>

Tugores, Francesca; Planas, Rosa (2006). *Introducción al patrimonio cultural*. Gijón: Trea. ISBN: 84 970 4242 5

Unesco (2000). *Definitions for "intangible cultural heritage". Member states (Replies to questionnaires sent to national commissions in February and August 2000)*. Paris: Unesco. <http://www.unesco.org/culture/ich/doc/src/00078-EN.pdf>

Unesco (2009). *Lista representativa del patrimonio cultural inmaterial de la humanidad*. París: Unesco.

Venezuela (1993). "Ley de protección y defensa del patrimonio cultural". *Gaceta oficial de la República de Venezuela*, n. 4.623 extraordinario.

Próximos temas centrales

Noviembre 2013	Formación y aprendizaje
Enero 2014	Documentación audiovisual
Marzo 2014	Políticas de información
Mayo 2014	Visualización de información
Julio 2014	Humanidades digitales
Septiembre 2014	Big data y analítica web
Noviembre 2014	Marketing

Los interesados por favor consulten detenidamente las Normas para autores:

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

y luego envíen sus artículos a través del gestor de manuscritos OJS de la plataforma del Recyt:

<http://recyt.fecyt.es/index.php/EPI/index>

NUEVAS TECNOLOGÍAS EN ANÁLISIS DE INTELIGENCIA COMPETITIVA. CASOS PRÁCTICOS

Pilar Ramírez-Calvo, Ana-Cristina Triviño, Aurelio Berges-García, Juan-Manuel Meneses-Chaus y José-Fernán Martínez



Pilar Ramírez-Calvo, doctora en ciencia y tecnología de alimentos y licenciada en bioquímica, es responsable del área de VT&IC en *e-Intelligent*, gestionando diferentes proyectos en colaboración con organismos públicos y empresas privadas. Analista de inteligencia con más de 10 años de experiencia en entornos de I+D+i y más de 150 informes de estrategia empresarial como: patentes, estados del arte, nuevos mercados, socios, competidores, *benchmarking*, *SWOTs*, etc. Responsable de certificación 166.006:2011 de VT&IC en *e-Intelligent*. Participante en congresos, seminarios y workshops internacionales, con publicaciones científicas y patentes.

<http://orcid.org/0000-0002-4014-686X>

e-Intelligent
Av. de la Albufera, 321, pl. 4, of. 10. 28031 Madrid, España
pilar.ramirez@e-intelligent.es



Ana-Cristina Triviño es gerente y responsable de desarrollo de negocio del área de Vigilancia tecnológica e Inteligencia competitiva (VT&IC): Vicubo Cloud. Consultora y formadora para diferentes unidades de VT&IC de ámbito regional, nacional e internacional. Dirección de proyectos web de: marketing online, usabilidad, e implantación de soluciones de VT&IC, buscadores, gestores documentales, gestores de contenidos, e-learning, business intelligence y data mining. Participa activamente en jornadas, seminarios y máster. Postgrado Executive en inteligencia de negocio (BI) por la *Escuela de Organización Industrial (EOI)* y Licenciada en documentación por la *Universidad Carlos III de Madrid*.

<http://orcid.org/0000-0003-3387-5594>

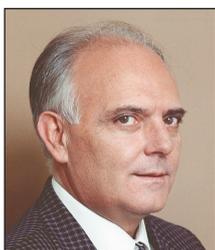
e-Intelligent
Av. de la Albufera, 321, pl. 4, of. 10. 28031 Madrid, España
cristina.trivino@e-intelligent.es



Aurelio Berges-García es director del *Grupo Gridat* e investigador y profesor de telecomunicaciones en la *Universidad Politécnica de Madrid*. Investigador en TICs relacionadas con la recuperación de información inteligente, sistemas de búsqueda, minería de datos y sistemas para la VT&IC. Asesor en el desarrollo e implantación de herramientas de VT&IC, CRM y ERP y experto en sistemas de gestión del conocimiento, como el *Sistema Madri+d*. Posee una larga experiencia como coordinador en más de 80 proyectos de I+D, nacionales e internacionales. Es autor de artículos de revista y ponencias de congreso, así como de varios libros.

<http://orcid.org/0000-0003-0597-6806>

*Univ. Politécnica de Madrid, Escuela Univ. de Ingeniería Técnica de Telecomunicación
Depto. de Ingeniería y Arquitecturas Telemáticas (Diatel)
Carr. de Valencia, Km. 7. 28031 Madrid, España
aurelio.berges@upm.es*



Juan-Manuel Meneses-Chaus es catedrático del *Área de Tecnología Electrónica* de la *Universidad Politécnica de Madrid (UPM)* y director del *Centro de Investigación en Tecnologías SW*. Es experto en síntesis y diseño de arquitecturas digitales de alta velocidad y diseño de sistemas de inteligencia ambiental en aplicaciones de salud. Dirigió la *OTRI* de la *UPM* hasta 2008. Creador y responsable del círculo de innovación en TIC, dedicado a actividades VT&IC. Colabora en programas de IP y es asesor de la *OMPI*. Ha dirigido tesis doctorales y participado en 77 proyectos. Es autor de más de 95 publicaciones en revistas y congresos.

<http://orcid.org/0000-0002-0569-2910>

*Univ. Politécnica de Madrid, Escuela Univ. de Ingeniería Técnica de Telecomunicación
Depto. de Tecnologías del Software y Sistemas multimedia para la Sostenibilidad (Citsem)
Carr. de Valencia, Km. 7. 28031 Madrid, España
juan.meneses@upm.es*



José-Fernán Martínez, de telecomunicación y profesor titular de la *Universidad Politécnica de Madrid (UPM)*, es responsable de proyectos de investigación en IT. Sus áreas de interés son, entre otras, internet de las cosas, ciudades inteligentes, redes inalámbricas de sensores y actuadores (WSAN). Es autor de publicaciones nacionales e internacionales, revisor técnico de eventos internacionales sobre telemática y miembro de comités técnicos y científicos. Ha participado en varios proyectos internacionales y europeos con financiación pública, y también en contratos de investigación con empresas de IT.

<http://orcid.org/0000-0002-0569-2910>

*Univ. Politécnica de Madrid, Escuela Univ. de Ingeniería Técnica de Telecomunicación
Depto. de Tecnologías del Software y Sistemas multimedia para la Sostenibilidad (Citsem)
Carr. de Valencia, Km. 7. 28031 Madrid, España
jf.martinez@upm.es*

Resumen

Las nuevas tecnologías basadas en internet están transformando los modelos de negocio en todo el mundo. Se plantea el uso de las nuevas plataformas de vigilancia tecnológica e inteligencia competitiva en unos casos prácticos en el entorno empresarial. Se repasan las diferentes fuentes de información en combinación con otros datos (como información tecnológica y datos financieros) para la creación de inteligencia. Las herramientas de análisis permiten detectar relaciones entre los datos de manera que las organizaciones pueden identificar cambios, adaptar sus productos a nuevos consumidores o anticipar los movimientos del mercado. Se describen 3 casos prácticos usando la solución *VicuboCloud* en la recolección y análisis de información para la toma de decisiones.

Palabras clave

Vigilancia tecnológica, Inteligencia competitiva, Inteligencia empresarial, Cloud computing, Análisis de datos, Casos prácticos.

Title: New technologies for analysis of business intelligence. Practical cases

Abstract

New internet-based technologies are transforming the business model around the world. The use of new competitive watch and intelligence platforms is discussed. We review the information sources used in combination with other data (e.g., financial and technological data) to create intelligence. Analytical tools allow the detection of connections between data so that the company can identify changes, adapt products to new customers, or anticipate market movements. Three practical cases are presented using the *VicuboCloud* solution to collect and analyze information for decision-making.

Keywords

Technological watch, Competitive intelligence, Business intelligence, Cloud computing, Data analysis, Practical cases.

Ramírez-Calvo, Pilar; Triviño, Ana-Cristina; Berges-García, Aurelio; Meneses-Chaus, Juan-Manuel; Martínez, José-Fernán (2013). "Nuevas tecnologías en análisis de inteligencia competitiva. Casos prácticos". *El profesional de la información*, septiembre-octubre, v. 22, n. 5, pp. 448-454.

<http://dx.doi.org/10.3145/epi.2013.sep.10>

Introducción

La inteligencia competitiva (IC) ha crecido en las últimas décadas para convertirse en una actividad de gran importancia en las empresas (Fuld, 1985; McKinnon; Burns, 1992). Puede definirse como un proceso de creación de relaciones entre información de diferentes fuentes con implicaciones potenciales para la mejora de la gestión de negocio y el apoyo a la toma de decisiones estratégicas en las áreas de marketing, desarrollo de producto, mercado e I+D+i, entre otras. La IC va más allá de la investigación de mercado o la vigilancia tecnológica, centrándose en todos los aspectos del entorno de la empresa para producir inteligencia (conocimiento del entorno para saber cómo actuar de forma óptima) en aspectos clave como la detección temprana de amenazas y oportunidades, identificación de mercados y clientes, monitorización y evaluación de competidores y apoyo a la planificación estratégica.

La cantidad de información disponible generada cada día, la competencia a nivel global, el énfasis en la gestión de la calidad, y la necesidad de disminuir el riesgo en la toma de decisiones ha llevado a utilizar medios tecnológicos en la práctica de la IC. Además, la utilización conjunta de las redes sociales amplía su alcance, incluyendo información adicional táctica y estratégica que puede ser añadida al conocimiento de la compañía.

Tradicionalmente, el proceso de IC se compone de cuatro etapas (Herring, 1999; Fuld & Company, 2002):

1. Planificación y dirección. Determinación de las necesidades de información estratégica o factores críticos de vigilancia (FCV): datos del entorno esenciales para el negocio.
2. Recolección y almacenamiento de datos. Buscar qué fuentes fiables y accesibles se pueden usar y recabar datos de ellas. Filtrar los resultados eligiendo los elementos de interés que se incluirán en la base de datos del proyecto.

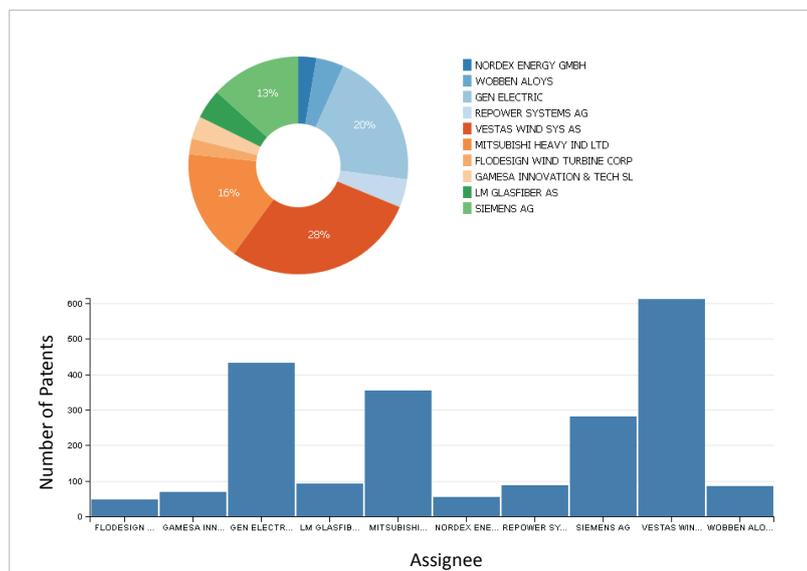


Figura 1. Top 10 líderes de patentes entre los suministradores de energía eólica (fuente: Espacenet)

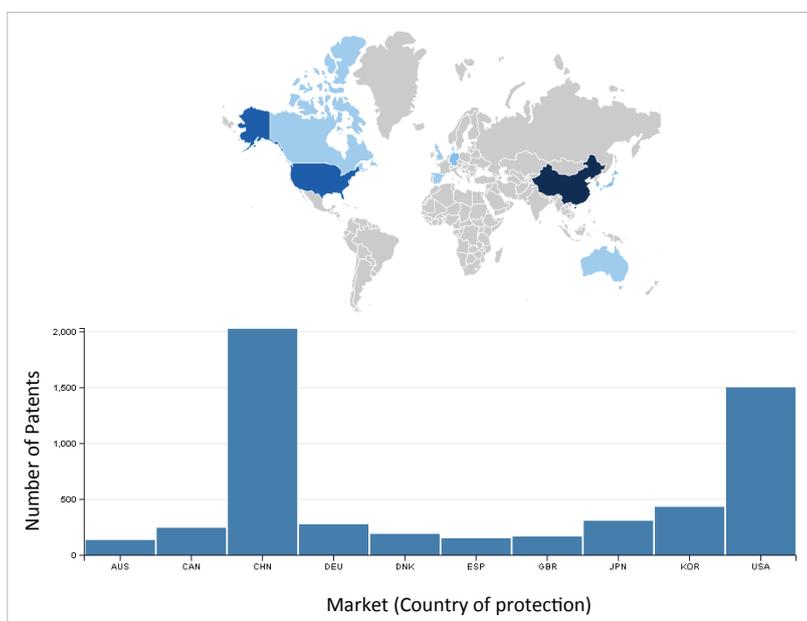


Figura 2. Top 10 países con patentes de energía eólica (fuente: Espacenet)

Tabla 1. Cuota de mercado de productores de energía eólica (fuente: Cleantech, 2012)

Ranking	Nombre empresa	Cuota de mercado	
		2011	2010
1	Vestas	12,7	14,8
2	Sinovel	9,0	11,1
3	Goldwind	8,7	9,5
4	Gamesa	8,0	6,6
5	Enercon	7,8	7,2
6	General Electric	7,7	9,6
7	Suzlon	7,6	6,9
8	Guodian United Power	7,4	4,2
9	Siemens	6,3	5,9
10	MingYang Wind Power	3,6	-

3. Análisis. Esta etapa es crítica, y consiste en interpretar los datos desde un punto de vista estratégico (inteligencia) determinando su relevancia para la toma de decisiones. Para el análisis se usan varios tipos de indicadores y técnicas (indicadores de patentes, análisis semántico, análisis DAFO, benchmarking, las cinco fuerzas de Porter, etc.). Es conveniente que el sistema de inteligencia sea colaborativo pues ello potencia la creación de nuevas ideas.

4. Difusión. Los resultados se ponen a disposición de los mandos superiores que deben tomar las decisiones estratégicas.

El tratamiento de grandes cantidades de datos requiere una potencia de procesamiento y un espacio de almacenamiento enorme. Usando centros de proceso externos (en la nube o *cloud computing*) las compañías han hallado la solución a este problema. Como se sabe, la nube es un modelo de ‘servicio bajo demanda’ que proporciona recursos informáticos (redes, servidores, almacenamiento, aplicaciones y servicios) con un esfuerzo de gestión mínimo por parte de la empresa usuaria, reduciendo el coste de inversión y de operación. Generalmente se paga una suscripción con una tasa mensual (Mell; Grance, 2011). El uso de la nube hace realidad la idea de “todo como un servicio” (*everything as a service*, XaaS) –es decir, las organizaciones clientes se suscriben a un servicio y no tienen que comprar ni hardware ni software para procesar sus datos, ni realizar ningún mantenimiento.

E-intelligent, una empresa spin-off de la UPM que ofrece productos y servicios de vigilancia tecnológica e inteligencia competitiva, tiene el producto *Vicubocloud* como plataforma de inteligencia en la nube para satisfacer las necesidades informativas de

sus clientes. En este artículo se presentan tres ejemplos recientes.

Metodología

El proceso de vigilancia tecnológica e inteligencia competitiva (VT&IC) que presenta este estudio se estructura, tal como recomiendan las normativas europeas en I+D (UNE 166006:2011 y Comité AEN/CTN 166, 2011), en las cuatro etapas descritas anteriormente. *VicuboCloud* se usa para analizar y mostrar la información relevante en gráficos, lo que permite evaluar de forma sencilla la posición de la compañía en el mercado frente a sus competidores.

Caso 1: Sector de energías renovables. Quién es quién en la fabricación de turbinas eólicas

Los fabricantes de turbinas eólicas las diseñan, prueban, fabrican y luego prestan asistencia para su operación y man-

tenimiento. Se trata de un mercado global, cambiante, del que es necesario seguir los movimientos para estar al día de alertas y oportunidades. Se analizó quién es quién entre los suministradores de tecnología eólica y dónde están vendiendo, para lo cual se utilizaron informes de mercado y de patentes, ambos incluidos en las bases de datos de *VicuboCloud*. Los líderes tecnológicos se pueden identificar no sólo por su volumen de producción, sino también por sus patentes.

La figura 1 muestra los 10 líderes en patentes de turbinas eólicas en los últimos tres años. *Vestas*, *General Electric*, *Mitsubishi Heavy*, *Siemens* y *Wobben Alloys* son los principales solicitantes de patentes que aportan desarrollos tecnológicos al mercado. El caso de *Wobben* es digno de tenerse en cuenta ya que es el inventor y no el solicitante el que aparece en la patente, lo cual puede ser parte de una estrategia en la gestión de patentes más amplia por parte de la compañía.

Las patentes pueden cruzarse con los fabricantes de turbinas obtenidos a partir de informes sectoriales actualizados. En 2011, un año difícil y de resultados financieros decepcionantes, la danesa *Vestas*, líder en tecnología, mantuvo su posición también como líder en la fabricación de turbinas, por encima de *Sinovel* (sin embargo, la situación cambió en 2012, ya que *Vestas* perdió su posición frente a *General Electric*).

En cuanto a dónde están vendiendo los suministradores de turbinas eólicas, según el último informe de mercado de *Global Wind Energy Council (GWEC)* (2013), China es el mayor mercado de turbinas, tanto tecnológico como mercantil, con tendencia de



Figura 3. Top 10 autores para el desarrollo de biomarcadores (fuente: *Web of knowledge*)

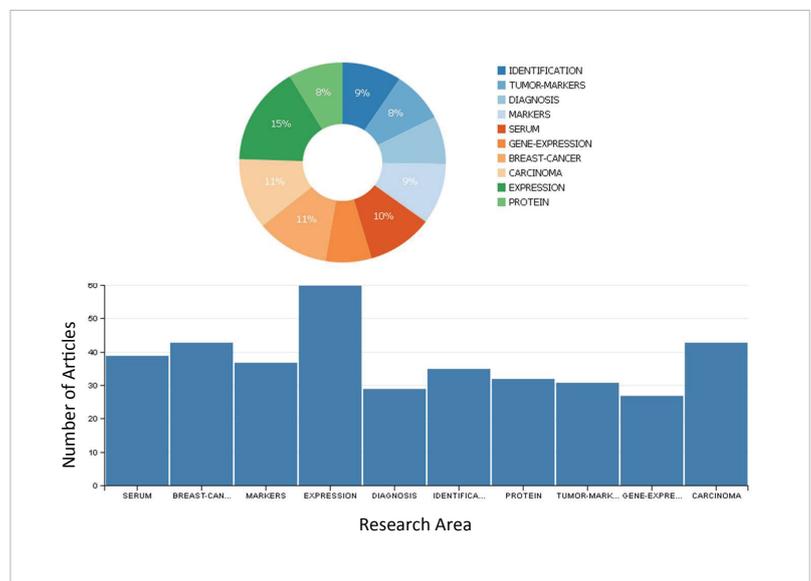


Figura 4. Top 10 áreas de investigación para el desarrollo de biomarcadores (fuente: *Espacenet*)

Tabla 2. Comparativa del proceso de calificación de biomarcadores por parte de *EMA* y *FDA*

	Objetivos y resultados formalmente calificados	
	<i>European Medicines Agency</i>	<i>Food and Drug Administration</i>
Objetivo	Abordar métodos de diseño y elaboración de fármacos innovadores y específicos	Identificar nuevas herramientas que permitan una producción más rápida de los medicamentos para que así lleguen antes a los pacientes
Enfoque	Asesoramiento y colaboración entre el <i>CHMP</i> ^a y las empresas, individuales o consorcios	Asesoramiento y colaboración entre la <i>FDA</i> y las empresas, individuales o consorcios
Calificación formal	Opinión de calificación del <i>CHMP</i> : aceptabilidad de un método definido para un propósito específico para el desarrollo de fármacos. Consejos de calificación del <i>CHMP</i> : orientación sobre futuros estudios	Declaración para la calificación: Adecuación de los datos presentados para obtener la calificación y el contexto de uso de la herramienta de desarrollo de fármacos (<i>drug development tool, DDT</i>) (abierta al público)
Después de la calificación	Proyecto de calificación publicado en la web de la <i>EMA</i> durante 6 semanas (consulta pública)	Notificación del <i>Federal Register</i> publicada en la web como un apéndice a la orientación
Biomarcadores calificados	6 biomarcadores: 2 de nefrotoxicidad y 4 de Alzheimer	3 biomarcadores: 2 de nefrotoxicidad y 1 de daño morfológico cardíaco

a. *Committee for Medicinal Products for Human Use*

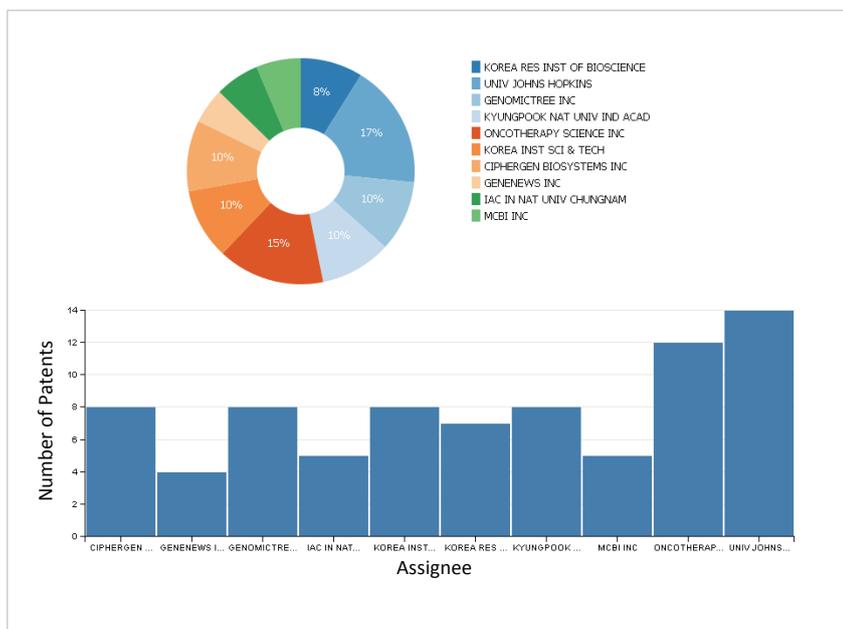


Figura 5. Top 10 líderes de patentes en biomarcadores (fuente: Espacenet)

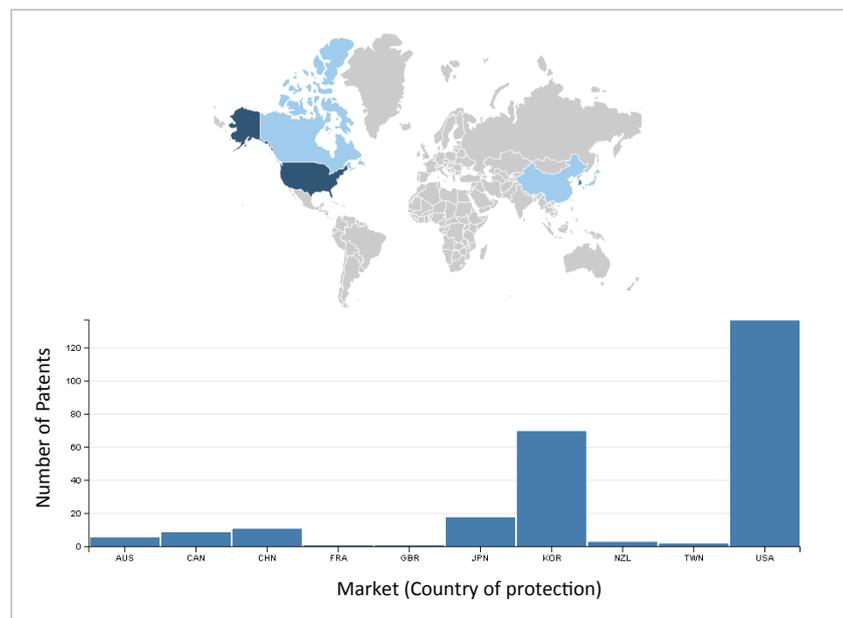


Figura 6. Top 10 países patentadores de biomarcadores (fuente: Espacenet)

crecimiento en mercados extranjeros. China, seguida de Estados Unidos, es el país con más patentes de energía eólica (figura 2).

Toda la información mostrada es parte de un informe más amplio propiedad de *E-intelligent*.

- descubrimiento: identificación del biomarcador;
- desarrollo del prototipo: se crean ensayos inmunológicos;
- comercialización: los ensayos están listos para su aplicación clínica.

Para nuestro estudio se analizará el proceso de calificación de biomarcadores en los mercados de Europa y EUA, el es-

Tabla 3. Posibles socios para la calificación de biomarcadores (fuente: VicuboCloud™)

	Partners information			
	Scientific area	Contact	Source	Date
1	Bladder cancer progression	Aarhus University Hospital, Denmark	Result from the EU funded FP7-Health programme	2013-01-07
2	Cancer diagnosis, prognosis and follow	Spanish National Research Council	Technology market Europe	2013-02-11
4	Respiratory chain dysfunction	Finnish research group	Technology market Europe	2013-01-15
8	Early diagnosis Alzheimer disease	Four Spanish public organisations and the University of Singapore	Technology market Europe	2013-02-22
12	Breast cancer	French Academic Research Laboratory	Technology market Europe	2013-02-25

Caso 2: Sector biotecnológico. Estado de la técnica en biomarcadores y barreras a su comercialización

Un biomarcador es una característica que es medida y evaluada como indicador de procesos biológicos normales, procesos patogénicos, o respuestas farmacológicas a una intervención terapéutica. En medicina, un biomarcador puede ser una sustancia detectable que se introduce en un organismo para examinar la función de un órgano o de otros aspectos de la salud; una sustancia cuya presencia anómala indica un estado de enfermedad; o un cambio en la expresión o el estado de una proteína que se correlaciona con el riesgo o progresión de una enfermedad, o con la susceptibilidad de la enfermedad a un tratamiento dado.

Como ocurre con los fármacos, antes de poder comercializar esas sustancias o esos procedimientos de detección, deben pasar el control de las correspondientes agencias oficiales. La evaluación o calificación de un biomarcador se define como el proceso probatorio, gradual y adaptado, que relaciona un biomarcador con factores biológicos y clínicos.

El mundo de los biomarcadores es complejo y en constante evolución. En los últimos años tienen un papel cada vez más importante en la industria farmacológica y en el diagnóstico clínico.

La introducción de un biomarcador en el mercado requiere tres fases:

tado de la técnica y las barreras comerciales existentes.

El estado de la técnica se obtiene a partir de la bibliografía científica (para detectar nuevo conocimiento) y de patentes (tendencias tecnológicas y de negocio), y las barreras comerciales se identifican analizando la información del proceso de calificación por parte de las agencias europea (EMA, *European Medicines Agency*) y norteamericana (FDA, *Food and Drug Administration*).

Estado de la técnica

Se monitorizaron los artículos publicados por los expertos (figura 3) y se clasificaron según las áreas en las que se están investigando nuevos biomarcadores (figura 4), lo cual permitió establecer el patrón de evolución de la investigación básica. También se buscaron las patentes solicitadas por las empresas e instituciones del sector, cuyo análisis detectó las tendencias actuales en el negocio de los biomarcadores: líderes, mercados y áreas de trabajo tecnológicas. El líder en tecnología de biomarcadores por número de patentes es la universidad norteamericana *Johns Hopkins*, seguida por la compañía japonesa *Oncotherapy*, tal como muestra la figura 5.

La figura 6 muestra los mercados más importantes para la venta de biomarcadores, según los países solicitantes de patentes: USA, Corea y Japón.

Finalmente, se analizaron las áreas tecnológicas de mayor interés en el desarrollo de biomarcadores, y su utilización por parte de los líderes del sector (figura 7).

Barreras comerciales

La comercialización es una tarea costosa pues el biomarcador y su ensayo deben superar la evaluación o calificación de las agencias farmacológicas, y luego hay que preparar su marketing. El procedimiento del ensayo debe seguir normativas establecidas por los organismos reguladores correspondientes.

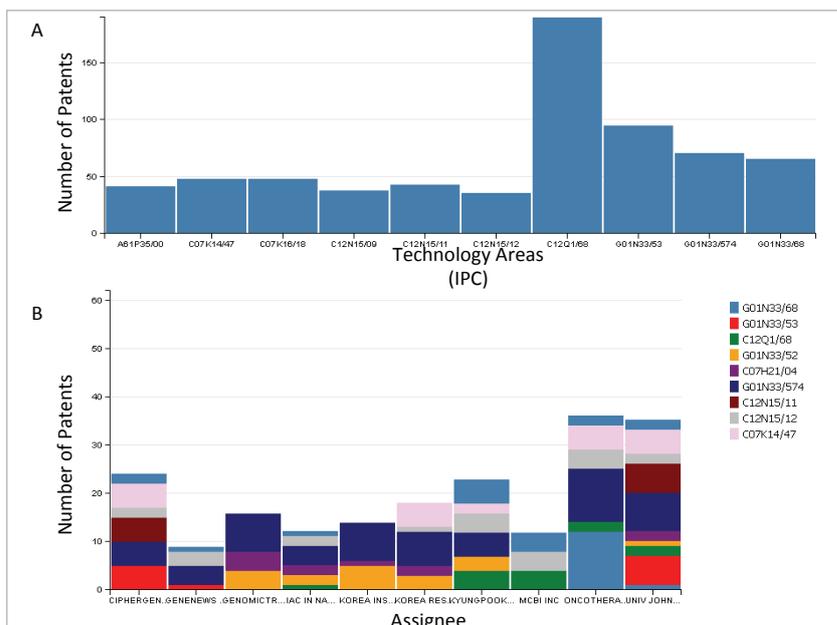


Figura 7. Top 10 áreas tecnológicas y su uso por parte de los principales solicitantes de patentes (fuente: Espacenet)

En la tabla 2 se compara el proceso de calificación seguido por las agencias europea (EMA, *European Medicines Agency*) y norteamericana (FDA, *Food and Drug Administration*).

La solicitud de calificación de un biomarcador es mejor hacerla desde un consorcio o agrupación de instituciones de investigación, pues así se comparten conocimientos, recursos, tecnología, muestras, etc., y se ahorran costes y tiempo. La tabla 3 presenta posibles socios para los proyectos de biomarcadores en diferentes áreas de investigación.

Caso 3: Sector residuos. Análisis competitivo en biofuel

Los biocombustibles son productos –fuentes de energía– generados a partir de biomasa o material ‘recientemente’ vivo. Pueden ser sólidos (madera, cosechas y sus residuos, desechos urbanos, etc.), líquidos (aceite vegetal, aceite residual alimentario, biodiesel, etanol u otros alcoholes, e hidrocarburos biológicos) o gaseoso (biometano, biogas, gases de vertederos o hidrógeno obtenido de biomasa).

Tabla 4. Análisis de posicionamiento competitivo de *Fulcrum Bioenergy* (fuente: *Marketline*)

Características competitivas	Hechos importantes	
	Hecho 1	Hecho 2
Ventajas competitivas	Economía de las materias primas	Proceso propio a escala comercial
Tecnología	Adquirida una licencia para el sistema de gasificación	Patente para integrar los residuos sólidos municipales en el proceso del etanol
Productos	Principalmente etanol	Otros: metanol y propanol
Coste de producción	Biofuel: \$1,30 por galón	El proceso de producción genera electricidad para empresas
Acuerdos de ventas	Empresa grande	Acuerdo de ventas de 3 años
Operaciones financieras recientes (2011-2012)	23-12-2011: Atrae una inversión privada de \$131,21M en sus instalaciones	18-11-2011: otra empresa invierte en ella
Inversores	Un grupo renovable	Firma de capital de riesgo
Socios y alianzas	Dos socios en materias primas y dos en técnica	Un socio vende
Futuros hitos	Construcción de 2 instalaciones adicionales en los EUA	

Tabla 5. Análisis DAFO de *Fulcrum Bioenergy*

Debilidades	<ul style="list-style-type: none"> • Sin fuerza ante posibles competidores (las patentes aún están pendientes) • Concentración geográfica • Los costos del producto sólo son estimados; dependen de múltiples variables
Amenazas	<ul style="list-style-type: none"> • Patente scondedidas a otros • Leyes y reglamentos ambientales • Aumento de la competencia • Cambios en los precios de los bienes de consumo
Fortalezas	<ul style="list-style-type: none"> • Contratos rentables para obtener materia prima • Proceso propietario a escala comercial • No hay nuevas redes de logística para el transporte de la materia prima • Promotor / propietario / operador
Oportunidades	<ul style="list-style-type: none"> • Instalaciones nuevas • Aumento de la demanda de servicios ambientales • Nuevas iniciativas de "energía de los residuos" • Extensión de la patente a nuevos mercados

Informe

Una compañía en el sector del biocombustible quiere analizar la información disponible de diferentes competidores tecnológicos para implementar una estrategia de posicionamiento competitivo en el área de los residuos sólidos urbanos para la producción de biomasa.

Primeramente se analiza la compañía para identificar sus modelos de negocio, acuerdos con otras compañías, desarrollos tecnológicos y noticias financieras. Está situada en California y diseña, desarrolla, posee y gestiona instalaciones que convierten residuos de celulosa en etanol. Se presenta el análisis competitivo de posicionamiento del competidor dentro del sector (tabla 4).

La evaluación final de las ventajas y debilidades competitivas se muestra en el análisis DAFO de la tabla 5, que apoya la toma de decisiones sobre la estrategia a seguir.

Conclusiones

El número de factores que influyen en la estrategia de negocio ha aumentado notablemente debido al fenómeno de la globalización y a la situación económica en Europa y EUA. La apertura de nuevos mercados en todo el mundo no es sólo tarea de grandes empresas, sino de todas, independientemente de su tamaño, con el fin de sobrevivir en este contexto económico sumamente competitivo.

Como el entorno en el que se mueven las compañías se hace cada vez más turbulento, las decisiones de los directivos se complican: deben buscar nuevas oportunidades de negocio en todos los mercados a la vez que lidiar con nuevos competidores en su propio entorno.

Anexo: Características principales de *VicuboCloud*

VicuboCloud es una plataforma que permite a una organización realizar vigilancia tecnológica e inteligencia competitiva de la Web y extraer la información estratégica necesaria. Se caracteriza por ser un servicio en "la nube". Por ello no es necesario ningún tipo de instalación en servidores propios ni licencias de software, con lo que se evitan los costes y molestias asociados al mantenimiento de equipos informáticos. <https://www.vicubocloud.es>

Sus funciones principales son:

- monitorizar cualquier fuente de información;
- centralizar y sistematizar el proceso de vigilancia tecnológica e -inteligencia competitiva (VT&IC);
- reducir el tiempo dedicado a la gestión de información;
- atención personalizada y confidencial;
- facilitar la integración de información de forma segura;
- analizar visualmente la información clasificada;

VicuboCloud incorpora tres extractores de información:

- *Vicubo sectores*: Información de partida clasificada por sectores, alimentada por fuentes de información abierta y contrastadas por un equipo de analistas. Además ofrece licitaciones, ayudas, legislación, normas, proyectos de I+D, informes económicos, análisis de mercado, estadísticas, patentes, publicaciones, artículos, eventos, noticias, etc.
- *Vicubo social*: información de las redes sociales configurable, capaz de rastrear, medir y evaluar en tiempo real la información generada en *Facebook*, *Twitter*, *LinkedIn*, *Google+*, etc.
- *Vicubo reader*: permite recuperar automáticamente las novedades producidas en las fuentes monitorizadas, lo cual elimina la necesidad de acudir a ellas diariamente.

Cada organización puede contratar lo que le interese y durante el tiempo que desee. Se puede alquilar por períodos de tiempo determinados, según la duración del informe. Permite integrar fuentes de información de pago (*Thomson Reuter*, *Factiva*, *Lexis Nexis*, etc.). Su potencia radica en el análisis de información mediante gráficos y la posibilidad de compartir información y recibir alertas.

Bibliografía

Comité Aenor AEN/CTN 166 (2011). *UNE 166.006:2011. Gestión de la I+D+i: Sistema de vigilancia tecnológica e inteligencia competitiva*. <http://www.aenor.es/aenor/normas/normas/fichanorma.asp?tipo=N&codigo=N0046930>

Fuld & Company (2002). *Measuring the value of competitive intelligence: the inadequacies of return-on-investment calculation*. White paper, Fuld & Company, Cambridge, MA. <http://www.fuld.com/wp/measuring-the-value-of-competitive-intelligence>

Fuld, Leonard M. (1985). *Competitor intelligence: how to get it; how to use it*. New York: John Wiley & Sons. ISBN: 978 0471809678

Herring, Jan P. (1999). "Key intelligence topics: a process to identify and define intelligence needs". *Competitive intelligence review*, v. 10, n. 2, pp. 4-14. [http://dx.doi.org/10.1002/\(SICI\)1520-6386\(199932\)10:2<4::AID-CIR3>3.3.CO;2-3](http://dx.doi.org/10.1002/(SICI)1520-6386(199932)10:2<4::AID-CIR3>3.3.CO;2-3)

McKinnon, Sharon; Burns, William (1992). *The information mosaic*. Boston: Harvard Business School Press. ISBN: 978 0875843179

Mell, Peter; Grance, Timothy (2011). *The NIST definition of cloud computing*. NIST special publication 800-145. <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>



UNE-ISO 16175 SOBRE GESTIÓN DE DOCUMENTOS EN OFICINAS ELECTRÓNICAS: ALCANCE Y LIMITACIONES



Vicent Giménez-Chornet



Vicent Giménez-Chornet es profesor titular del área de *Documentación* en la *Universidad Politécnica de Valencia*. Miembro del *Comité 50 de Documentación de Aenor (Asociación Española de Normalización y Certificación)*. Director del máster oficial interuniversitario en gestión cultural. Ha sido presidente del *Col·legi Oficial de Bibliotecaris i Documentalistes de la Comunitat Valenciana* entre 2010 y 2012, secretario de *Fesabid*, archivero en el *Archivo del Reino de Valencia*, y presidente de *Avei (Associació Valenciana d'Especialistes en Informació)*.
<http://orcid.org/0000-0003-1183-9058>

Universidad Politécnica de Valencia
Departamento de Comunicación Audiovisual, Documentación e Historia del Arte
Camino de Vera, s/n. 46022 Valencia
vigicho@har.upv.es

Resumen

Se analiza la norma *UNE-ISO 16175* para determinar su interés y alcance en la gestión de los documentos electrónicos, permitiendo una mejor gestión de los documentos de archivo en las organizaciones. Se analizan también sus carencias o límites, que son generalmente resueltos por otras normas.

Palabras clave

Documentos de archivo, Gestión de documentos electrónicos, Sistema de gestión de documentos electrónicos, Requisitos para la gestión de documentos.

Title: *UNE-ISO 16175* on document management in the paperless office: scope and limitations

Abstract

The *UNE-ISO 16175* is analyzed to determine its scope and merits in the management of electronic records, enabling better archives management in organizations. There are analyzed its shortcomings or limitations that are generally resolved by other standards.

Keywords

Records, Archive documents, Digital records management, Systems, Requirements for record management.

Giménez-Chornet, Vicent (2013). "*UNE-ISO 16175* sobre gestión de documentos en oficinas electrónicas: alcance y limitaciones". *El profesional de la información*, septiembre-octubre, v. 22, n. 5, pp. 455-458.

<http://dx.doi.org/10.3145/epi.2013.sep.11>

1. Introducción

Las normas *ISO (International Organization for Standardization)* son de voluntario cumplimiento, pero se usan porque las entidades encuentran beneficio al hacerlo (**Bustelo-Ruesta, 2012**). Algunos países elaboran normas propias y otros dejan que sus entidades adopten estándares aprobados por organismos internacionales. Situándonos en España, si se quiere afrontar un buen modelo para la gestión de documentos electrónicos, se deben adoptar estándares generados por entidades internacionales, con la excepción del *Esquema nacional de interoperabilidad (PAE)* que se centra en los aspectos de interoperabilidad entre las administraciones.

Sobresalen dos estándares que establecen los requisitos funcionales que se deberían cumplir en la gestión de documentos producidos en la administración electrónica de cualquier entidad:

- *Moreq (Modelo de requisitos para la gestión de documentos electrónicos)*, elaborado en 2001 en la *Comisión Europea* dentro del programa *IDA (Interchange of data between administrations)* e impulsado por la *Fundación DLM Forum*. Lleva dos actualizaciones generadas tras las sugerencias de mejora expresadas por sus usuarios, y su versión más reciente se ha retitulado *Modular requirements for records systems 2010*.
<http://www.dlmforum.eu>

Artículo recibido el 04-06-2013
Aceptación definitiva: 11-07-2013

<http://moreq2010.eu>

- La norma 16175, publicada por ISO y editada en tres partes entre 2010 y 2011, que ha sido adoptada por Aenor como norma UNE-ISO 16175 en 2012, con el título *Información y documentación. Principios y requisitos funcionales para documentos en entornos de oficina electrónica*. Como la gestión de los documentos electrónicos sólo se puede realizar mediante un software (ya no sirven depósitos, estanterías, compactos, carpetas colgantes, cajas, señalética, fichas índice, fichas catalográficas, listados de cuadros de clasificación, listados de inventarios, etc.), cabe trasladar el modelo de gestión manual a unos requerimientos que debe cumplir el software para gestionar los documentos electrónicos producidos por organizaciones o personas en el desarrollo de su actividad, asegurando su integridad, autenticidad y validez a lo largo del tiempo, y que puedan ser recuperados cuando sea necesario.

2. Alcance

La norma UNE-ISO 16175 (2012) tiene tres partes que se venden separadamente.

1. *Generalidades y declaración de principios.*
2. *Directrices y requisitos funcionales para sistemas que gestionan documentos electrónicos.*
3. *Directrices y requisitos funcionales para documentos de la organización.*

Su objetivo principal es la elaboración de unas directrices y requisitos para la producción y gestión de documentos electrónicos. Directrices que debe observar la organización y requisitos que debe cumplir el sistema y el software correspondiente.

En la parte 1 se expone quiénes pueden ser los usuarios potenciales de esta norma, haciendo énfasis en que puede ser de utilidad a:

- responsables de crear un software para la gestión de los documentos electrónicos de archivo;
- organismos gubernamentales, en cuanto les puede permitir tener un modelo para evaluar las prestaciones automatizadas ya implantadas en la gestión de los documentos;
- sector privado, para poder incorporar un modelo de gestión de los documentos electrónicos en sus procesos de negocio.

Trata también de las buenas prácticas cuando las organizaciones deciden introducir las nuevas tecnologías para gestión y control de los documentos. Resalta unos principios rectores que deben considerar las organizaciones que necesitan un sistema de información donde la creación, conservación y uso de los documentos sea la evidencia o prueba de su actividad y satisfaga las necesidades de negocio. Entre ellos se indica, por ejemplo, que la información electrónica tiene que ser gestionada de forma activa y conservada de modo fiable como evidencia auténtica –*prueba o testimonio* en la bibliografía clásica de la archivística española– de la actividad, o que la información de la organización debe estar vinculada a su contexto mediante el uso de metadatos, etc.

Se plantean aspectos sobre la implementación, ya que un software de calidad es sólo un componente en la gestión de los documentos electrónicos de la organización, y señala otros componentes como el marco de políticas, tanto del contexto legislativo como las políticas de la organización, la gestión de proyectos o la gestión del cambio. Asimismo advierte de algunos riesgos que se deben considerar a la hora de seleccionar el software que se implementará en la gestión de los documentos, finalizando con algunos aspectos sobre sostenibilidad financiera, en definitiva una inversión de capital acorde a la estrategia de negocio.

La parte 2 se centra en dos aspectos:

- determinar las directrices para un sistema que gestiona los documentos electrónicos (capítulo 4);
- establecer los requisitos funcionales que debe soportar el software en ese sistema (capítulo 5).

Entre las directrices se señalan los aspectos esenciales que se deben tener en cuenta:

- para que los atributos de los documentos de archivo (autenticidad, fiabilidad, integridad y usabilidad) se mantengan efectivos, para hacer posible la producción de los documentos en su contexto, mantenerlos en el tiempo exigido y poderlos recuperar por sus metadatos;
- riesgos en los que podría caer la organización si no se implementa un sistema de gestión de documentos electrónicos (como incumplir requisitos legales, desacreditación de la dirección ejecutiva si se demuestra su incapacidad de gestionar la información de la organización de manera competente, etc.);
- se señalan los beneficios si implementa el sistema de gestión (protección de los intereses de la organización, capacidad de prestar servicios más eficiente, etc.).

Los requisitos funcionales son un conjunto de requerimientos demandados al software, con dos niveles de exigencia (el “debe”, obligatorio, y el “debería”, aconsejable y a valorar por la organización) para que la gestión de documentos que se realizaba en la archivística tradicional de forma manual ahora se pueda cumplir en la gestión electrónica. Entre ellos están la captura (o el ingreso) de los documentos en el sistema, la vinculación a su contexto, la identificación, la clasificación, los niveles de clasificación, el control del acceso, los controles de seguridad, los metadatos de la gestión –incluidos los referidos a la disposición de los documentos reflejados en las *Tablas de valoración o retención documental*– la trazabilidad, las transferencias, los sistemas de búsqueda o recuperación de la información, etc. Algunos de estos requisitos funcionales del capítulo 5 son coincidentes con el *Moreq* europeo.

La parte 3 se centra también en dos aspectos desde la perspectiva de la organización:

- directrices que ésta debe contemplar (capítulo 2);
- sus requisitos funcionales (capítulo 3).

El objetivo de esta tercera parte es comprender los procesos y requisitos para identificar y gestionar los documentos en los sistemas de la organización y desarrollar los requisitos de la función documental. La norma podrá servir a la orga-

nización para revisar esa función documental o evaluar el grado de cumplimiento de los sistemas ya existentes. Los destinatarios de esta norma son principalmente el personal responsable del diseño, revisión y/o implementación de sistemas en la organización (analistas de negocio, responsables en las adquisiciones de TIC), los profesionales de los documentos que asesoran en los procesos, y los proveedores de software y desarrolladores que deseen incorporar funciones concretas para la gestión de los documentos electrónicos.

Las directrices resaltan la importancia de la gestión de los documentos, y proponen asuntos y procesos que debe considerar la organización a la hora de revisar, diseñar o adquirir funciones para su sistema de gestión. Entre las directrices trata la relevancia de conservar las evidencias de los procesos y actividades de la organización, de la integración de la gestión de documentos en las bases de procesos definidos (flujos de trabajo), y propone una serie de pasos para determinar los requisitos en la gestión de los documentos:

- análisis del proceso de trabajo;
- identificación de los requisitos de las evidencias;
- identificación del contenido y la información asociada a la evidencia, sus vínculos y dependencias (como localización, tamaño, formatos de archivo, seguridad, protección de datos, pistas de auditoría, etc.);
- diseño de estrategias para gestionar los procesos documentales que se consideran esenciales para la organización (por ejemplo, su contenido debe ser inalterable, para considerarlo auténtico y fiable);
- evaluación de riesgos como estrategia adecuada de la organización;
- propuesta e implementación (considerando los roles y las funciones de los usuarios) de un plan que deba ser tratado por equipos del proyecto (formado por expertos externos, propietarios del negocio, profesionales archiveros, expertos en riesgos, etc.) para el desarrollo de una especificación en el diseño del sistema de la organización.

Los requisitos funcionales están sometidos a tres niveles de exigencia

- el “debe”, de imprescindible cumplimiento;
- el “debería”, se debe sopesar su grado de cumplimiento;
- el “puede”, opcional.

Los requisitos son 125 y se estructuran en cuatro apartados:

- creación de los documentos en su contexto: documento fijo con los metadatos del proceso de negocio y el contexto de la organización;
- configuración de los metadatos: debe permitir, entre otras funciones, que el administrador del sistema defina la fuente de datos de cada elemento de los metadatos durante la configuración del sistema, o debería dar soporte al movimiento de los documentos proporcionando mecanismos para la reasignación y reclasificación de los documentos dentro del sistema de la organización);
- soporte a la importación, exportación e interoperabilidad, como importación de la información de las pistas de autoría asociadas a documentos, o la exportación de documentos dentro del sistema de la organización o a una institución archivística que tenga un programa de conservación de documentos a largo plazo;

- conservación/retención de los documentos según la disposición establecida en las *Tablas de valoración/calendarios de conservación* dictadas por órganos autorizados.

3. Limitaciones

Las normas no abarcan, en general, toda la actividad de una entidad, sino que proponen regular algunos procesos concretos que pueden beneficiarse de una buena práctica. En algunos casos estos procesos pueden ser certificables. La gestión de los documentos electrónicos, nativos o incorporados al sistema desde otros formatos, contempla muchas facetas que son tratadas por varios estándares. La *UNE-ISO 16175* ya indica en algunos casos sus límites, aquello que no aborda, y de lo que el usuario podrá encontrar mejor información en otras normas. Algunas limitaciones son:

- Regular la incorporación al sistema de documentos nacidos en papel mediante la digitalización. Las normas *UNE-ISO* sobre digitalización son la *UNE-ISO/TR 13028:2011 IN Información y documentación. Directrices para la implementación de la digitalización de documentos*, y la *UNE-ISO/TR 15801:2008 IN Imagen electrónica. Información almacenada electrónicamente. Recomendaciones sobre veracidad y fiabilidad*.
- Regular el empaquetado de la información, que conserve los metadatos de los documentos electrónicos, para su conservación y difusión. Es el modelo *Open archival information system (OAIS)* que ha sido adoptado como norma *ISO 14721:2012 Space data and information transfer systems. Open archival information system (OAIS). Reference model*, y la *ISO 20652:2006 Space data and information transfer systems. Producer-archive interface. Methodology abstract standard*.
- Regular el intercambio de copias de documentos electrónicos entre sistemas informáticos, tratado extensamente en la *UNE-ISO 17933:2006 GEDI. Intercambio de documentos electrónicos genérico (ISO 17933:2000)*.
- Regular la calidad en el proceso, tratado en las normas *UNE-EN ISO 9000:2005 Sistemas de gestión de la calidad. Fundamentos y vocabulario (ISO 9000:2005)*; *UNE-EN ISO 9001:2008/AC:2009 Sistemas de gestión de la calidad. Requisitos (ISO 9001:2008/Cor 1:2009)*; y *UNE-EN ISO 9004:2009 Gestión para el éxito sostenido de una organización. Enfoque de gestión de la calidad (ISO 9004:2009)*.
- Regular el formato de fichero para la conservación de los documentos a largo plazo. Una propuesta es el formato pdf: *UNE-ISO 19005-1:2008 Gestión de documentos. Formato de fichero de documento electrónico para la conservación a largo plazo. Parte 1: Uso del PDF 1.4 (PDF/A-1)*.
- Regular el proceso de valoración y selección de documentos. Actualmente no existe ninguna norma, y algunos países regulan vía legislación cómo establecer dicho proceso para que queden reflejados en unas *Tablas de valoración / Tablas de retención / Calendarios de conservación* para los documentos generados en las administraciones públicas, y que tengan validez jurídica los documentos que se ha aprobado destruir.
- Regular un diseño básico de recuperación de la información. Actualmente se ha avanzado en la normativa relacio-

nada con la recuperación mediante tesauros: *ISO 25964-1:2011 Information and documentation. Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval*; *ISO 25964-2:2013 Information and documentation. Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies*

- Se complementa con el análisis de los procesos de trabajo desde el punto de vista de la creación, incorporación y control de documentos, tratado en la norma *UNE-ISO/TR 26122:2008 IN Información y documentación. Análisis de los procesos de trabajo para la gestión de documentos*.
- Se complementa con la extensa norma que trata de los metadatos para la gestión de documentos, publicada en tres partes: *UNE-ISO 23081 Información y documentación. Procesos de gestión de documentos. Metadatos para la gestión de documentos*, Parte 1 (2008), Parte 2 (2011), y Parte 3 (2012).
- Se complementa con los criterios generales de las estrategias a seguir para la conservación de la información que ha quedado registrada en los documentos, establecidas en la *UNE-ISO/TR 18492:2008 IN Conservación a largo plazo de la información basada en documentos*.
- Se complementa con la norma que ofrece la metodología para un enfoque sistemático en la creación y gestión de documentos, alineada con los objetivos y estrategias de la organización, proponiendo criterios que debe seguir una organización para la buena gestión de los documentos: *UNE-ISO 30300:2011 Información y documentación. Sistemas de gestión para los documentos*, *UNE-ISO 30301:2011 Información y documentación. Sistemas de gestión para los documentos. Requisitos*. Pronto se publicarán las normas complementarias: *30302* sobre la guía de implementación, *30303* sobre requisitos para los organismos que realicen la auditoría y la certificación, y *30304* sobre guía de evaluación.
- Se complementa con la norma que trata de la gestión de los documentos de archivo: *UNE-ISO 15489-1:2006 Información y documentación. Gestión de documentos. Parte 1: Generalidades*, y *UNE-ISO/TR 15489-2:2006 Información y documentación. Gestión de documentos. Parte 2: Directrices (ISO/TR 15489-2:2001)*.

4. Conclusión

La norma *UNE-ISO 16175* establece en sus tres partes los principios de buenas prácticas mediante una propuesta de directrices que deben cumplir el sistema y la organización, y una propuesta de requisitos funcionales que debe cumplir

el software que controlará el sistema de gestión de los documentos en un entorno de oficina electrónica. La norma *UNE-ISO 16175* es más amplia que el *Moreq*, ya que éste se limita a los requerimientos funcionales del sistema o software, mientras que la norma abarca también las directrices que debe cumplir la organización para el correcto funcionamiento de la oficina electrónica y garantizar la gestión de unos documentos auténticos, íntegros y fiables. Otras normas pueden resolver las limitaciones de la *UNE-ISO 16175* en las diversas actividades que se puedan plantear en una entidad.

“ La norma *UNE-ISO 16175* es más amplia que el *Moreq*, ya que éste se limita a los requerimientos funcionales del sistema o software ”

Bibliografía

Bustelo-Ruesta, Carlota (2012). “La normalización internacional en información y documentación: ¿una historia de éxitos? El caso de la normalización ISO en gestión de documentos”. *Métodos de información (MEI)*, II época, v. 3, n, 4, pp. 39-46.
<http://dx.doi.org/10.5557/IIMEI2-N2-039046>

DLM Forum (2013). *About MoReq2010. Modular requirements for records systems*. DLM Forum Foundation.

PAE, Portal de Administración Electrónica. *Esquema Nacional de Interoperabilidad*.
<http://administracionelectronica.gob.es/ctt/verPestanaGeneral.htm?idIniciativa=en>

UNE-ISO 16175-1:2012 (2012). *Información y documentación. Principios y requisitos funcionales para documentos en entornos de oficina electrónica. Parte 1: Generalidades y declaración de principios*. Madrid: Aenor.

UNE-ISO 16175-2:2012 (2012). *Información y documentación. Principios y requisitos funcionales para documentos en entornos de oficina electrónica. Parte 2: Directrices y requisitos funcionales para sistemas que gestionan documentos electrónicos*. Madrid: Aenor.

UNE-ISO 16175-3:2012 (2012). *Información y documentación. Principios y requisitos funcionales para documentos en entornos de oficina electrónica. Parte 3: Directrices y requisitos funcionales para documentos en los sistemas de la organización*. Madrid: Aenor.

INDICADORES



THE NEED FOR END-USER CUSTOMIZATION OF THE JOURNAL-SETS OF THE SUBJECT CATEGORIES IN THE SCIMAGO JOURNAL RANKING DATABASE FOR MORE APPROPRIATE LEAGUE LISTS. A CASE STUDY FOR THE LIBRARY & INFORMATION SCIENCE FIELD



Péter Jacsó

Nota: Este artículo puede leerse traducido al español en:
http://www.elprofesionaldeinformacion.com/contenidos/2013/sept/12_esp.pdf



Péter Jacsó is a professor in the *Department of Computer and Information Sciences* at the *University of Hawaii*. He has been teaching courses on library automation, digital libraries, advanced database searching, and computerizing serials management. He has years of practical experience in running computer centers, managing library and information services, and evaluating database software and content. He has published extensively on these topics in journals, conference proceedings, and books, and received awards for his teaching and database reviews.
<http://orcid.org/0000-0003-1409-2404>

Univ. of Hawaii, Dept. of Information and Computer Sciences, Library and Information Science Program
2550 The Mall, Hamilton Bldg. # 2-I. Honolulu, HI 96822, USA
<http://www.jacso.info>
jacso@hawaii.edu

Abstract

The open access *SCImago Journal & Country Rank (SJR)* service -built from data in the subscription-based *Elsevier's Scopus* database- offers bibliographic information and bibliometric indicators for nearly 20,000 journals and other serial publications for 1996-2011. Journals are assigned to 300+ pre-defined subject categories and 26 broad subject areas to produce league lists. The set of 134 journals assigned to the Library and Information Sciences (LIS) subject category and a sample set of 50 other LIS journals, scattered across several other categories, were analyzed in the latest edition (*SJR-2011*). Recommendations are made on how *SJR* could be enhanced by its developers to include simple customization options in the interface to facilitate a more efficient look-up and more valid comparison of the standing of journals and other serials and to reflect the prevailing preferences at the institutions where the rank positions in journal league lists are used to inform career advancement and funding decisions. It is argued that extending the current three-year citation window to five years would be more appropriate for all disciplines, and would also make *SJR* comparable to metrics-based journal league lists produced from the open access *Microsoft Academic Search (MAS)*, *Google Scholar Metrics for Publications (GSMP)*, and *Eigenfactor.org (EF)* data and the subscription-based *Journal Citation Reports (JCR)*, which pioneered the idea of a metrics-based journal league list.

Keywords

Journal ranking, League lists, *SCImago Journal Rank*, *Scopus*, Journalology, Metrics-based publication assessment, Bibliometrics, Scientometrics, Informetrics.

Título: Necesidad de que los usuarios finales puedan personalizar los conjuntos de revistas de las categorías temáticas en la base de datos *SCImago Journal Ranking* para obtener rankings más apropiados. Estudio de caso del campo Biblioteconomía y Documentación

Resumen

El servicio en acceso abierto *SCImago Journal & Country Rank (SJR)* –elaborado a partir de la base de datos de pago *Scopus*, de *Elsevier*- ofrece información bibliográfica e indicadores bibliométricos de casi 20.000 revistas y otras publicaciones periódicas en el período 1996-2011. Para producir los rankings, a las revistas se les asignan más de 300 categorías temáticas pre-definidas, así como 26 áreas temáticas más amplias. En este trabajo se ha analizado el conjunto de 134 revistas asignadas a la categoría Biblioteconomía y Ciencia de la Información (LIS) y una muestra de otras 50 revistas –que también tratan LIS, pero dispersas en varias otras categorías-, todo ello en la última edición *SJR-2011*. Se hacen recomendaciones sobre cómo los desarrolladores del *SJR* podrían mejorar el producto si incluyeran opciones de personalización simples en la interfaz para

Article received on 08-07-2013

Final acceptance: 28-07-2013

facilitar una visualización más eficiente y una comparación mejor de la situación de las revistas y otras publicaciones periódicas. Así también se reflejarían las preferencias de las instituciones para las que la posición de las revistas en el ranking la utilizan para informar la promoción profesional y las decisiones de financiación. Se argumenta que la ampliación de la ventana de citación actual de tres a cinco años sería más adecuada para todas las disciplinas, y también haría *SJR* comparable con los rankings en acceso abierto producidos por *Microsoft Academic Search (MAS)*, *Google Scholar Metrics for Publications (GSMP)*, *Eigenfactor.org (EF)* y de los *Journal Citation Reports (JCR)* –éste de acceso mediante suscripción de pago-, que fue pionero en idear un ranking de revistas basado en métricas.

Palabras clave

Ranking de revistas, *SCImago Journal Rank*, *Scopus*, Estudios de revistas, Evaluación de publicaciones basada en métricas, Bibliometría, Cienciometría, Informetría.

Jacsó, Péter (2013). "The need for end-user customization of the journal-sets of the subject categories in the *SCImago Journal Rank* database for more appropriate league lists. A case study for the Library & Information Science field". *El profesional de la información*, septiembre-octubre, v. 22, n. 5, pp. 459-473.

<http://dx.doi.org/10.3145/epi.2013.sep.12>

Introduction

SCImago Journal & Country Rank (SJR) was the first service to offer an open access alternative –from a different source- to the subscription-based *Journal Citation Reports (JCR)* of the *Institute for Scientific Information, ISI* (now *Thomson-Reuters*). Among the open access journal league lists, *SJR* is the largest, best designed, and richest in indicators. However, *SJR* does not provide all of the *JCR* features, such as the highly informative lists of citing and cited journals with graphs and tables of the number of citations received from and given to other journals and themselves (i.e. self-citations), their impact factors (if covered by *JCR*) and the yearly distribution of citations. While the *JCR* has been available for the Sciences and the Social Sciences, *SJR* extends its coverage to include the Arts & Humanities journals, –where journals still matter-.

SJR also includes a *Country Ranking* module for assessing the scholarly publication productivity and impact by country, based on the affiliations of authors. There is also a standalone application, the *SCImago Institutional Ranking Service*, based on the authors' institutional affiliations. It would have been better to give each component a unique acronym, such as *SJR*, *SCR*, and *SIR* to avoid confusion. Henceforth *SJR* will be used only for the *SCImago Journal Ranking* component itself.

SJR was launched as a free service in 2007, based on data produced by *Elsevier* for its *Scopus* database, which in turn was meant to compete with the *Web of Science (WoS)* database. The deficiencies of *SJR* are inherited from *Scopus*, which still has inconsistent and odd assignment of journals to subject categories, as well as significant gaps in the coverage of many journals. This paper proposes a solution for the former problem, and a follow-up paper will discuss the implications of the latter, both from the perspective of journal ranking and league lists.

Adding some relatively simple software features to *SJR*, as outlined in this paper, for customizing the disciplinary league lists of nearly 20,000 journals could provide a very comprehensive and still easy-to-use resource for renewal and cancellation decisions in collection development. It would

also help in the assessment of published research related to promotion, tenure and funding decisions where the standing of journals in which applicants have published is of high importance.

There have been several substantial reviews of *SJR* itself ever since its debut and the interest in its use apparently keeps growing, as *SJR* (and *Scopus*) are compared against other services which are used to create journal, institution and/or country league lists (**Bergman**, 2012; **Delgado-López-Cózar et al.**, 2013; **Falagas et al.**, 2008; **González-Pereira et al.**, 2009, 2010; **Jacsó**, 2010; **Oppenheim**, 2008; **Torres-Salinas et al.**, 2010). Especially relevant is the paper where the authors recommend improving the subject classification in *SJR* on the basis of reference analysis (**Gómez-Núñez**, 2011).

“The interest in the *SJR* use apparently keeps growing”

For the background of metrics-based ranking and comparison of journals, and about the digital sources from which they are created, the following papers can provide further information to the readers (**Bar-Ilan**, 2010; **De-Moya-Ane-gón et al.**, 2007; **Deis**; **Goodman**, 2005; **Dess**, 2006; **Garfield**, 1999; 2005; **Jacsó**, 2007, 2008, 2009a; **Moed**, 2010; **Pudovkin**; **Garfield**, 2004, 2009; **Raj**; **Zainab**, 2012; **Wagner**, 2009).

The content profile of *SJR*

The 2011 edition of *SJR (SJR-2011)* has bibliographic information and bibliometric indicators for more than 19,700 serial publications, including scholarly and professional journals, bulletins, newsletters, book series and conference proceedings. For the sake of simplicity, serials and journals will be used synonymously in the rest of the paper as a group term for all these document types.

SJR covers 1996-2011, providing yearly detail from 1999 onward. For 1996, 1997 and 1998 only the 3-year cumulative indicators are given, as the purpose of the *SJR* indicator and

other related indicators has been to report the productivity and citedness of the journals in the previous 3-year period, vis-a-vis the classic 2-year window through the traditional *Journal Impact Factor (JIF-2)* introduced by **Eugene Garfield** in print and then microform and digital formats as the annual *Journal Citation Reports (JCR)*. The only exception for the 3-year citation window in *SJR* is the set of citations/document (citation rate) indicators, which are provided for 2, 3 and 4 year windows, although the last one appears only in the data table and not in the league lists of *SJR*.

For *SJR*-2011 there are 8.85 million master records for documents. The distribution of records by source type is not specified but by running a no-holds-barred search in *Scopus* for 2008-2011 to retrieve all the records, it can be estimated that 80.3% of the records are from academic journals, 1.7% from trade publications, 15% from conference proceedings, and 3% from book series.

By document type, the distribution is 66% journal articles, 6% review articles (of the research literature), and 19% conference papers. These make up the so-called citable items (91%), the term introduced originally for the *JCR*. In reality, some of the other document types (editorials, short notes, letters to the editors, errata) are also cited occasionally, but to a minimum extent. As the assignment of document types has not been accurate and consistent in databases, it is worth mentioning because *SJR* and *JCR* use the distinction in the league tables, *MAS* and *GSMP* do not, and *Eigenfactor* uses only the citable document types for measuring their productivity and impact.

SJR takes into account a maximum or 33% of the total references which cite the journal itself

The *SJR* league lists can be looked up for each year by simply selecting any year as the “census year” from 1999 to 2011. The citation window is fixed to 3 years. For example, *SJR*-2011 has the census year of 2011 (Y1), and produces most of its indicators by calculating the number of citations received from papers published in 2011 by the documents published in 2008 (Y1-1), 2009 (Y1-2) and 2010 (Y1-3). Some other indicators use a different calculation method. The cited documents indicator refers to those documents which received at least one citation in any year within the assessment time frame of 2008, 2009, 2010 or 2011.

The *SJR* indicator considers not just the number of citations but also the prestige of the citing sources, so the citations received in the year selected by papers published in the previous three years are counted and weighted. This is the

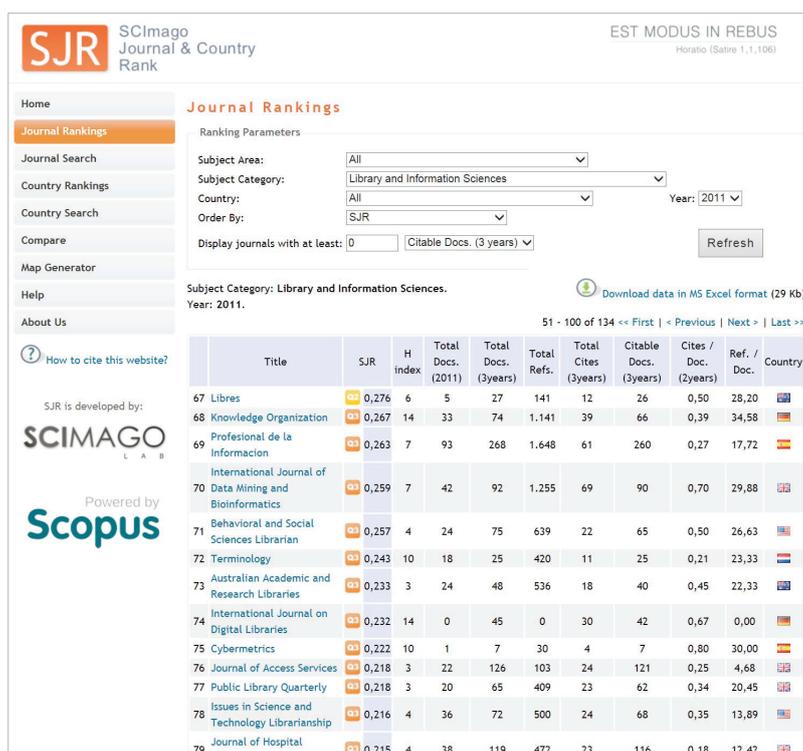


Figure 1a. Excerpt from the league table of journals in the LIS subject category <http://www.scimagojr.com/journalrank.php?category=3309>

same approach used in the *Eigenfactor* database, but the latter automatically excludes all self-citations at the journal level. *SJR* takes into account a maximum of 33% of the total references which cite the journal itself (confirmed by **González-Pereira**, personal communication, 2011) This unusual treatment of journal-level self-citation is not mentioned in the help file, but should be, and prominently, as most users are not aware of this policy.

The *SJR* 2012 edition was expected to be published by the first quarter of 2013. This is not a critical issue from the perspective of the current research, which focused on the most recent edition (*SJR*-2011 as of mid-July, 2013) for subsequent comparison with competing league lists that calculate bibliometric indicators for papers published between 2007-2011, from databases such as the 2011 editions of the *JCR*, *Eigenfactor*, *GSMP*, and *MAS* databases, which offer indicators for –among others- a 5-year window.

Data are imported from the *Scopus* database. If after the import the *Scopus* database is updated with records for documents in its still much needed back-filling process to eliminate or reduce the gaps, these changes are not reflected by the latest *SJR* indicators until its next update. That is the reason that a current search from the *Scopus* database (updated several times a week) may show for 2008-2011 more documents published in a journal, and more citations received by it than reported in *SJR*. The follow-up of this paper will focus on the consistency of the breadth of coverage of LIS journals.

The recommended 5-year citation window

Extending the 3-year citation window to a 5-year period would be a useful enhancement for several reasons. One

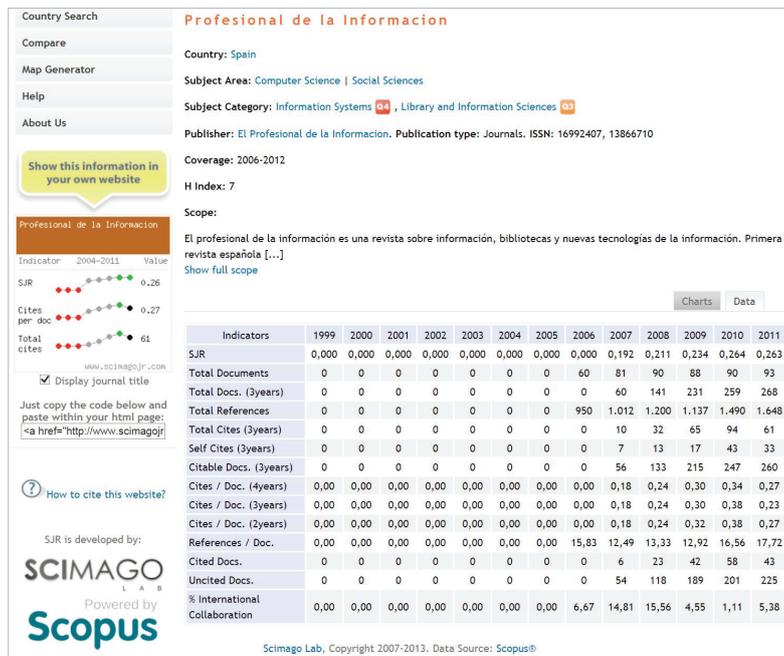


Figure 1b. The data table of a journal with additional indicators about its performance <http://www.scimagojr.com/journalsearch.php?q=16992407&tip=iss>

is that the 3-year period is too short in most disciplines. Journals may reach their citation peak by the 3rd year after publication (González-Pereira *et al.*, 2010) but citations received in the 4th and 5th and subsequent years still could significantly increase most of the indicator scores of the journals' productivity and impact. As we can learn from *JCR-2011*, in the LIS field the aggregated Cited Half-Life was 7.1 years, i.e. the median age of the papers that were cited in the census year. In *JCR-2012* (which was just released as this manuscript went to press) the Cited Half-Life is 7.3 years. In the Computer Science-Information Systems subcategory these indicators were 6.7 and 7.0 for 2011 and 2012, respectively. This could be an argument for using a longer citation window, but the other reasons still make the 5-year window better for comparing databases that offer indicators for 5-year window.

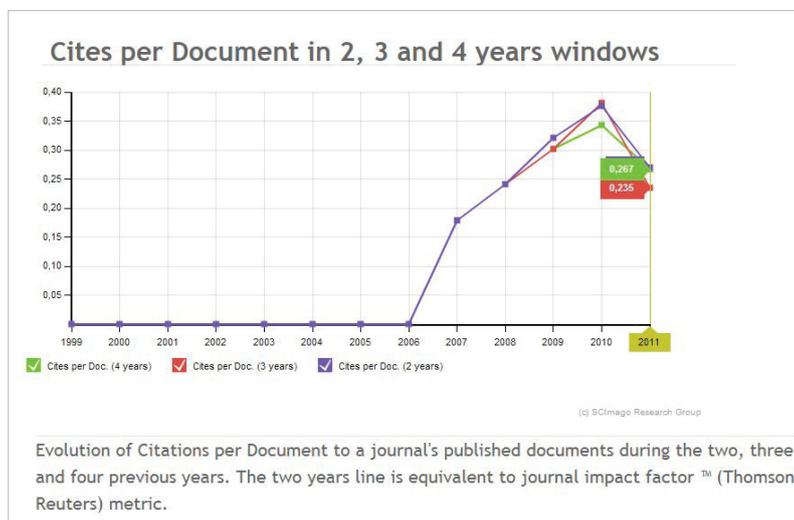


Figure 1c. The impact factors of a journal for a 2, 3 and 4-year citation window <http://www.scimagojr.com/journalsearch.php?q=16992407&tip=iss>

The 5-year citation window would be also more appropriate for promotion/tenure application support and post-tenure assessment where the 5-year time span and repeating cycle are very common. All of the above-mentioned databases use a 5-year window (sometimes with other options) versus the original 2-year window used in the classic *JCR*, which I refer to as *JIF-2* for clarity and to distinguish it from the 5-year impact factor (*JIF-5*) added to *JCR* in 2009.

It is recommended to calculate and display in the *SJR* league table and Details table the indicators for a 5-year window for **only** those journals covered by *Scopus* consistently every year between 2007 and 2011 (for *SJR* 2011). This restrictive inclusion principle should also be applied if the developers wouldn't extend the 3-year window. It is to be noted that having comprehensive data for the chosen window (i.e. without gaps at the volume and issue level) can be more important than the choice of the width of the time frame, as will be discussed in the follow-up paper).

The content of the league table for the LIS subject category, a journal Data table and one of the many standalone charts from *SJR-2011* are illustrated in figure 1a, 1b and 1c. The content, layout and legibility of the league tables is very good. It will be recommended, nevertheless, that in the additional alternate league tables the large logos be moved to another area, in order to accommodate more indicators (currently present in the journal's data table) that are very informational for end-users, such as the ratio of Cited Documents, and the suggested External Citation and Cites/Doc. rate for 5 years.

The competitive context

The journal set of *SJR-2011* is of almost twice the size of the number of serials covered by the most current edition of *Thomson-Reuters' Journal Citation Reports (JCR 2011)*: 194 versus 83 – at first sight. However, the journal set size information must be qualified and interpreted appropriately in the comparable subject areas and categories, rather than just taken at face value. The set of 134 journals assigned to the LIS category in *SJR* was analyzed in detail for coverage along with 50 other, undoubtedly LIS-related, journals scattered across a dozen other subject categories and not included in the LIS subject category.

GSMP limits the league list to the top 20 journals in every subcategory even if it has information about many other journals in the subject category. As usual, *Google* does not reveal the number of journals, let alone the titles included in *GSMP*. The attitude of giving as the fox gives (soup to the stork) – Editor's note: to give something but with

limitations-, has been a trademark of the developers of *Google Scholar*. The two indicators, the 5-year h-index and the median citations, are already calculated automatically for many more journals in order to generate the top-20 list, so displaying the top 50 or even the top 100 journals in LIS would have not required additional effort. Of course, this may give away too much – by *Google's* standard- about how well or how badly the *GSMP* software crawls and parses digital journal collections.

The *GSMP* service was launched April 1, 2012 (ominously, on April Fools' Day), from a cleaned-up version of *Google Scholar*. To the credit of the *GSMP* developers, the clean-up was good compared to regular *Google Scholar* data, although still not sufficiently reliable, showing duplicate entries with minor differences in accented and other special characters (Jacsó, 2012, 2013; Delgado-López-Cózar et al., 2012). In the original announcement of the service it was promised that the *GSMP* data would not change for one year ("Scholar Metrics are based on our index as it was on April 1st, 2012. For ease of comparison, they are NOT updated as the Scholar index is updated."). Actually, *GSMP* was re-launched seven months later after some additional clean-up. The 5-year h-index and the h-5 median indicators are not informative enough without knowing the total number of documents retrieved from and citations matched for the journal by *GSMP*. Again, it would reveal too much about the (in)comprehensiveness of the source coverage in *GSMP*. *Google* developers should have a simple utility to check duplicate entries in league lists, a sort option by journal name, an increased limit to at least 50 journals, and should include the number of documents and number of citations used for each journal to calculate the two indicators.

In *SJR* there are 50 undoubtedly LIS-related journals scattered across a dozen other subject categories not included in the LIS subject category

The *Eigenfactor Service (EFS)* –based on *JCR* data- provides information on only 45 of the 83 LIS journals included in the *JCR*-2011 Information Science and Library Science (ISLS) category.

<http://www.eigenfactor.org/rankings.php?bsearch=NU&searchby=category&orderby=Eigenfactor>

It "adds insult to injury" that some of the most widely known and most influential ISLS journals are excluded. This massively distorts the real rank positions of many of the 45 journals included in the ISLS category by the *Eigenfactor* database. Some of the serials missing from this pseudo-*JCR*/ISLS league list are *ASLIB proceedings*, *Information research*, *Information technology and libraries*, *JASIS&T*, *Journal of information science*, *Journal of information technology*, and *Scientometrics*. The ontology created by the *Eigenfactor* developers is somewhat better, having indicators for 53 journals in its Information Science category (it does not have an LIS category), but still much too limited for a sufficiently comprehensive league list of LIS journals: it has only 3 in-

140	AUSTRIAN J EARTH SCI ISSN: 0251-7493	EF: 12 AI: 40
141	J OVONIC RES ISSN: 1584-9953	EF: 12 AI: 29
142	TURK ONLINE J EDUC T ISSN: 1303-6521	EF: 9 AI: 15
143	INFORM DEV ISSN: 1741-6469	EF: 9 AI: 23
144	ENFERM EMERG ISSN: 1575-4723	EF: 8 AI: 21
145	ASIAN BIOMED ISSN: 1905-7415	EF: 8 AI: 5

Figure 2. Journals erroneously assigned to the Infectious Diseases category of the *Eigenfactor* "ontology" in the 2010 edition of the database are still not corrected in that edition as of mid-2013.

<http://www.eigenfactor.org/rankings.php?search=118&year=2010&searchby=efcat&orderby=Eigenfactor>

dicators at a precision of 7 decimal places, which are not readily comprehensible by mere mortal end-users. The idea of weighting the citation counts by the prestige of the citing source was a very good one, and the use of percentile bar-charts is very good, but the implementation needs much more spot-checking to minimize non-sensical information.

The other absurdity in the handling of league lists in the *Eigenfactor* database is the assignment of journals to completely unrelated categories. For example, some of the journals that ended up in the Infectious Diseases subject category within the *Eigenfactor* ontology in the 2010 edition of the database were the following titles: *Austrian journal of earth sciences*, *Education science and technology*, *Energy*, *Information development*, *Journal of ovonic research*, *Financial research letters*, *Miskolc mathematical notes*, *Platinum metals review*, and *Turkish online journal of education technology*. These were corrected in the 2011 edition after the critical review (Jacsó, 2012b), but in mid-2013 they still appear in the 2010 edition of the *Eigenfactor* database.

MAS has a Library Science category with 72 journals in 2013 (62 for the 2007-2011 window). This is a realistic set because the category is LS, not LIS. *MAS* uses the term "current 5-year", so as time passes by for many journals the 5-year time span is shifting. In *MAS* there are some journals already where the 5-year span is from 2008 to 2012. It would have been better from Microsoft to have separate edition(s) for each year, e.g. *MAS*-2010, *MAS*-2011 to avoid ambiguity by the partially shifting 5-year window.

The subscription-based *Scopus* database itself did not create league lists, but started to directly provide journal-level bibliometric data through its Journal Analyzer module, including now the *SJR* and *SNIP* (Source-normalized impact per paper) indicators. The latter makes it possible to compare the prestige/impact of journals in different disciplinary areas by correcting for their very different citation patterns. This is highly relevant for some subject areas, such as LIS, where information science journals usually dominate the

1700	COMPUTER SCIENCE Subject area	
code	# of JNs	Subject categories
1701	165	Computer Science (miscellaneous)
1702	114	Artificial Intelligence
1703	127	Computational Theory and Mathematics
1704	78	Computer Graphics and Computer-Aided Design
1705	140	Computer Networks and Communications
1706	204	Computer Science Applications
1707	40	Computer Vision and Pattern Recognition
1708	124	Hardware and Architecture
1709	40	Human-Computer Interaction
1710	132	Information Systems
1711	45	Signal Processing
1712	184	Software

Figure 3a. The suggested hierarchically browsable subject category index for the Computer Science broad subject area

code	# of JNs	Subject categories
3501	93	Dentistry (miscellaneous)
3502	→ 1	Dental Assisting
3503	→ 1	Dental Hygiene
3504	9	Oral Surgery
3505	8	Orthodontics
3506	→ 1	Periodontics
3601	41	Health Professions (miscellaneous)
3602	5	Chiropractics
3603	→ 3	Complementary and Manual Therapy
3604	→ 1	Emergency Medical Services
3605	12	Health Information Management
3606	→ 1	Medical Assisting and Transcription
3607	12	Medical Laboratory Technology
3608	→ 0	Medical Terminology
3609	→ 2	Occupational Therapy
3610	→ 3	Optometry
3611	11	Pharmacy
3612	70	Physical Therapy, Sports Therapy and Rehabilitation
3613	→ 1	Podiatry

Figure 3b. The abuse of the miscellaneous category in Dentistry and the irrational ones in Dentistry and Health Professions with 0-4 journals

code	# of JNs	Subject categories
2001	169	Economics, Econometrics and Finance (miscellaneous)
2002	388	Economics and Econometrics
2003	125	Finance

Figure 3c. Elsevier’s miserly and reticent treatment of the Economics, Econometrics and Finance broad subject area by parroting its components as subject categories – except for Econometrics

top positions and library science journals often compete with journals in communication and education in the various databases used for creating league lists.

Surprisingly, *Scopus* does not even use the subject category terms developed by Elsevier for the journal title list of –well, *Scopus*. One feature that was not well implemented in the otherwise state-of-the-art *Scopus* software was the summarization and visualization of the journals’ performance indicators (Jacsó, 2007, 2008, 2009c). Even today, the *Scopus*

Journal Analyzer module looks as if it had been an afterthought. This functionality is further limited by the restriction that a maximum of 10 journals can be handled by the *Journal Analyzer* module.

Subject areas and categories in SJR

SJR relies not only on the data content of *Scopus* but also on the classification of the source publications into 26 broad subject areas and 300+ subject categories (which are like subject descriptors of controlled vocabularies in many databases).

Subject areas

Elsevier assigns serials to one or more of the 26 (technically, 27) broad subject areas, ranging from Agricultural and Biological Sciences to Veterinary Sciences. The 27th broad subject area is labeled “Undefined,” with 778 items in *Scopus*. Volunteering this information is appreciated, but it is not obvious why these items could not be assigned to one of the 26 other *Scopus* subject areas. *SJR* does not have this category. A review of some of the records in the Undefined category suggests that these entries are simply incomplete records, missing the metadata element that assigns a record to a subject area.

By far the largest number of journals is assigned to the broad subject area of Medicine. There are 5,312 serials assigned to that subject area, about 27% of the actively covered serials for 2008-2011 in *SJR*. On average, journals are assigned to 1.3 broad subject areas. This high share is also reflected by the 2.5 million records within the Medicine subject area, a similar percentage of the 8.8 million total records for documents processed by Elsevier for the same time period in *Scopus*, which served as the document base for *SJR-2011*.

Of course, it is not expected that the documents would be evenly distributed among the 26 subject areas. However, the imbalance in the coverage by broad subject areas is quite clear, which in turn suggests some changes. Nursing, Health Professions, Dentistry and Decision Sciences are also listed at the top level among the 26 broad subject areas, in the company of Computer Science, Physics/ Astronomy, Biochemistry/Genetics, and Materials Science. The first three, which -combined- barely exceed 3.6% of the 8.8 million records, could have been assigned to a lower-level subject category, such as Complementary Medicine or Allied Health. Decision Sciences (with fewer than 80,000 documents) could have been one of the lower-level subject categories of the Business, Management and Accounting or the Computer Science subject areas.

Apparently, the design of the subject hierarchy was driven more by the pre-eminent databases of Elsevier in medicine and healthcare, such as *EMbase* and *EMcare*, and the *EM-tree Thesaurus*, which was enhanced by 1,500 terms for the nursing field alone.

The mean number of journals per subject areas is 954, the median is 606, and there are areas where the number of journals is excessively large, such as Engineering (1,767), Agricultural and Biological Sciences (1,622), Biochemistry, Genetics, and Molecular Biology (1,506).

The number of journals assigned to a subject area does not imply breadth of coverage at the record level. For example, the 3rd largest subject area by number of journals is Arts & Humanities (2,002), but there are only 240,000 records in that broad subject area, which seemed to have been an afterthought to make *Scopus* appealing to Arts & Humanities libraries (Jacsó, 2007). On the other hand, there is no *JCR* category for Arts & Humanities, so that is a definite asset of *SJR*, despite its relatively small volume of documents.

Subject categories

Elsevier also identifies, in its publicly accessible journal list, the subject categories assigned to each journal. There are more than 300 subject category terms. A journal may be assigned to several subject categories. No perfect assignment can be expected for nearly 20,000 journals in 26 broad subject areas, but browsing through several hundred journals and checking their assignment to subject categories indicated that the ontology is mostly good, logical and uses common terminology, not the strained and indirect language of many controlled vocabularies. However, there are serious shortcomings for some subject areas and subject categories that must be corrected by *Elsevier* and then updated in *SJR*.

The absurdities in some of the subject category levels –discussed below– may be the reason why *Scopus* does not use these subject descriptors in the records, let alone make them search criteria. Only the 26 very broad areas appear in the *Scopus* records and are offered for searching and filtering the results. These are not appropriate for creating league lists, which require much more specific subject categories.

This is where *SJR* and another similarly appealing and standalone service, the *CWTS Journal Indicators* enter the scene. <http://www.journalindicators.com>

They demonstrate the importance and power of the natural intelligence and bibliometric experience of the members of the *SCLImago Research Group* and *Leiden University's Centre for Science and Technology Studies (CWTS)*. The latter implemented, integrated and at the end of 2012 updated **Henk F. Moed's** original *SNIP* indicator algorithm (reflecting the significant differences among the citation practices in the various disciplinary fields) to produce an open access league list of nearly 20,000 journals from the *Scopus* database (Moed, 2010; De-Moya-Anegón *et al.*, 2007; Waltman *et al.*, 2012).

One of the most apparent advantages of *SJR* and the *CWTS* services (beyond being open access) is that their designers were aware of the existing classification of the journals covered by *Scopus* into 300+ subject categories and built them into their services, which was not done in *Scopus* itself. These subject area terms make it possible to produce league lists by practical and mostly well-identified research fields, but the vocabulary requires some clean-up.

Assignment of subject categories to journals

There are weaknesses in the subject term assignments in *SJR*, which are inherited from *Elsevier's* journal list in *Scopus*. The following examples illustrate the major deficiencies in the assignment of journals to subject categories.

Scientometrics, the first journal dedicated to this subject is –correctly– assigned to LIS, along with two other categories: Computer Science Applications, and Computational Theory and Mathematics. *Journal of informetrics* is assigned to 5 subject categories (Applied Mathematics, Computer Science Applications, Management Science and Operation Research, Modeling and Simulation, Statistics and Probability), but not to LIS. Similarly, *Law library journal* is assigned only to the Law category, and *Legal reference services quarterly* only to the LIS category. It would be very appropriate to assign these two journals to both the LIS and Law subject categories. *Journal of digital information* is assigned to the Information Systems category only, and not to LIS, while *Journal of digital information management* is assigned to LIS and two other subject categories.

It is not clear why *College & research libraries* (entered with “and” as the conjunction rather than with the ampersand symbol used by this and hundreds of other journals) is assigned correctly to the LIS subject category, but *College & research libraries news* is assigned to the Education subject category and not to LIS.

Some categories are redundant, as evidenced when one category covers almost all of the journals also assigned to the other. For example, the subject category of Gerontology (with 10 journals) overlaps Geriatrics and Gerontology (with 84 journals). Only 3 journals in the former category are not assigned to the latter. In addition, there is a subject category of Aging. This is a symptom of the inconsistent assignment of journals to subject categories by *Elsevier* in its journal list compiled for the *Scopus* database, and hence imported into *SJR*.

Conceptually and linguistically there does not seem to be much of a difference between the subject terms Psychiatric Mental Health and Psychiatry and Mental Health. Still, both appear on the subject category list. The former is assigned to 18 journals, the latter to 329 journals, and 8 of them appear in both subject categories. There are even more complicated situations. For example, there are subject categories for Management Information Systems (54 journals), Information Management Systems (46 journals), Information Systems (132 journals), Health Information Management (54 journals) and Health Informatics (31 journals).

This confuses even to experienced librarians who recall from the 1970s the displeasure of searching multiple databases using phrase-indexed descriptor terms from the thesauri of half a dozen databases that assigned practically the same terms but with slightly different punctuation and structure. The search engine did not tolerate an extra space, or a semi-colon instead of a colon. For example, the user who wants to know the standing of journals related to medical information systems would need to jump through many subject categories to learn about the availability of such journals because they are scattered across several subject categories for no obvious reason. With a good add-on utility these anomalies could be minimized.

The inconsistency and sloppiness in the assignment of journals to subject categories undermines the purpose of creating disciplinary league lists of journals. Assigning journals to several categories is not a bad idea, if “there is a method to

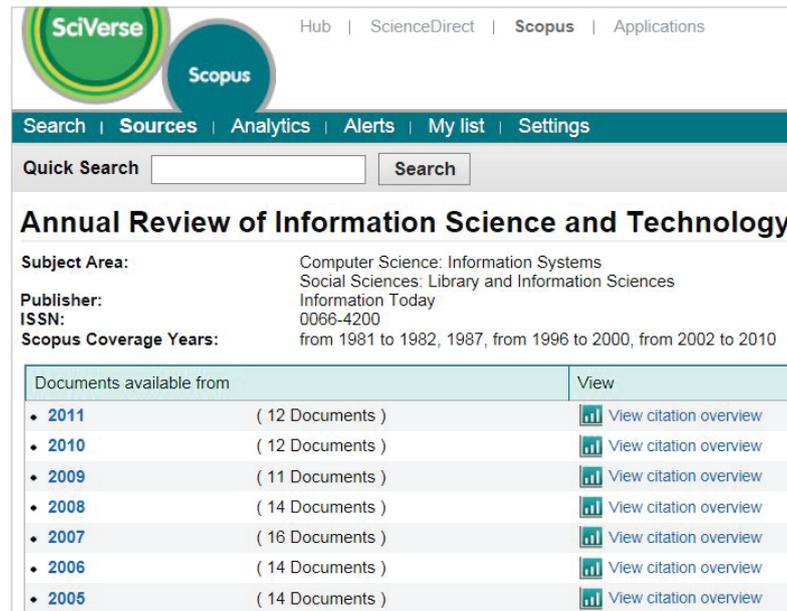


Figure 4a and 4b. Complete absence of *Arist* in *SJR*, complete coverage in *Scopus* for the past 10 years.

it". It is appropriate and informative to see that the *Health information and libraries journal* is assigned to Health Informatics, Health Information Management, and LIS, and is in the 2nd quartile in all these subject areas. Its assignment to the Medicine (Miscellaneous) subject category is questionable, and its position in the top quartile is somewhat surprising, until one realizes that there are 1,563 journals in the Medicine (Miscellaneous) subject category, and it is listed as the 360th journal by the h-index and 304th by its *SJR* score among the 390 journals in the top quartile for this subject area.

“Everything is miscellaneous”

Each subject area has a miscellaneous subject category, often with a very large number of journals, proving the adage made famous by the bestseller book that argued “everything is miscellaneous” (Weinberger, 2007), i.e. there is no universally agreeable method for classifying information.

The 26 subject areas and the 300+ subject categories are browsable – a very useful feature. It could be further improved by offering an option to browse the subject terms in a hierarchical way, e.g. to look up all the subject categories under the Computer Science broad subject area. The subject areas have a 4-digit code and the subject categories

code have the same first 2 characters. The miscellaneous subject category is always the first one. Currently these are not used in *SJR*; they were picked up from the *Scopus* journals list for this paper. The following series (figure 3) is not screenshots from the browsable subject index but excerpts from my mock-up sheets to illustrate how the enhanced browsable subject category index should look (i.e. alphabetical and shows the category code, and the number of journals assigned to a particular subject category). Each should start with the main category code and all upper-case subject area entry, such as Computer Science. It would give the user an instant feel for the number of journals within the subject area and its subject categories. The number of journals per category should not be calculated, as a journal may be assigned to more than one of the subject categories. The average number of journals in a subject category is 102; the median number is 65 across *SJR*-2011.

The redundant subject category terms and those with no associated journals –such as Medical Terminology, Nurse Assisting- should be deleted. The 28 other categories with only a couple of journals assigned to them –such as Dental Assisting, Dental Hygiene, Periodontics, Podiatry, Optometry- should be merged into other, semantically closest, subject categories. The inadequate and unprofessional treatment of the Health Profession and Dentistry subject areas clearly stand out at this level, supporting the earlier observation that they should not be designated among

the 26 broad subject areas. They also illustrate the overuse of the miscellaneous category and the irrational subject categories with 0-3 journals (figure 3b). One would expect that *Elsevier* could have come up with a few dozen additional and more specific subject categories rather than merely parroting the name of the broad subject area with 606 journals assigned to the Economics, Econometrics and Finance subject areas (figure 3c). The thesaurus of the *EconLit* database could have given some inspiration for *Elsevier's* indexing specialists.

The treatment of some subject categories also reinforces the subtitle of the above mentioned book (“the power of the new digital disorder”). In *Elsevier's* journals classification, it is madness and there is **no** method to it - to paraphrase Hamlet. This is not the venue to philosophize about ontologies, but rather to recommend that *SJR* developers add a relatively simple software solution to alleviate or even solve the problems in *Elsevier's* “ontology” for the subject classification of journals. For the *El profesional de la información* readership, the best way to help drive home these points about the inaccuracies and inconsistencies in the imported subset of the *Scopus* database may be to illustrate the implications through the prism of (re)assigning LIS journals from different subject categories in a flexible,

non-prescriptive way. The recommended software enhancements could be implemented by the *SJR* developers, and easily used by librarians and other interested end-users for any other subject categories to create more appropriate league lists for any of the subject categories.

The LIS subject category and its source base

There is a subject category in *SJR* for Library and Information Sciences. Almost all the 134 journals in it are LIS-related, although to different extent. Only three journals don't seem to be relevant for the LIS⁴ subject area (which is used here in a broad sense to include Library and Information Sciences/Systems/Services/Studies). These are the *Canadian journal of program evaluation, Development and learning in organisations*, and the *International journal of hospitality and tourism administration*. Just because they may have a few papers which discuss evaluation of a LIS Program or the role of a library as a learning organization, or the comparison of the far less stellar, no tuxedo required, library facilities on cruise ships, they should not be included in the LIS league list. *Terminology* and the *International journal of lexicography* may not seem to belong to LIS, but it is appropriate to assign them to it because both journals have papers related to natural language searching, controlled vocabularies, indexing and abstracting.

It is surprising at first sight that several core LIS journals do not appear in the LIS category league list of *SJR*, such as *Journal of academic librarianship, Information processing & management, Journal of documentation, Online information review, Information technology and libraries, Internet research, Law library journal, Canadian journal of information and library science, Journal of scholarly publishing*, and *School library media research*.

The good news is that all of the above-mentioned serial sources are included in *SJR*. However, they are assigned (by Elsevier for the *Scopus* journal list) to the Education, Information Systems, Information Management Systems, Management Information Systems, Law, Business, Management, Accounting, Communications and Media Technology subject categories, but are not assigned (also) to the LIS category. They should be and could be, because multiple category assignment is possible, and some journals are –reasonably– assigned to more than three subject categories in the journal list of *Scopus*, and hence in *SJR*. The process of re-assigning journals to subject categories by librarians and other sufficiently interested end-users could be very easy if the *SJR* developers would add a simple utility to the *SJR* software.

The bad news is that some highly relevant and active LIS journals – especially from the School Librarianship and the Archiving fields– are not covered by *Scopus* and hence by *SJR* at all. These include *School libraries worldwide, Teacher-librarian, Journal of information literacy, Archivaria, Archives*, and *Archival issues*. Important journals from other LIS fields are also totally absent in *Scopus* and hence in *SJR*, such as *Information technologies, Library technology reports, Evidence-based library and information practice, New review of information behavior research*, and *Journal of education for library and information science*. Only *Scopus* can help in this regard.

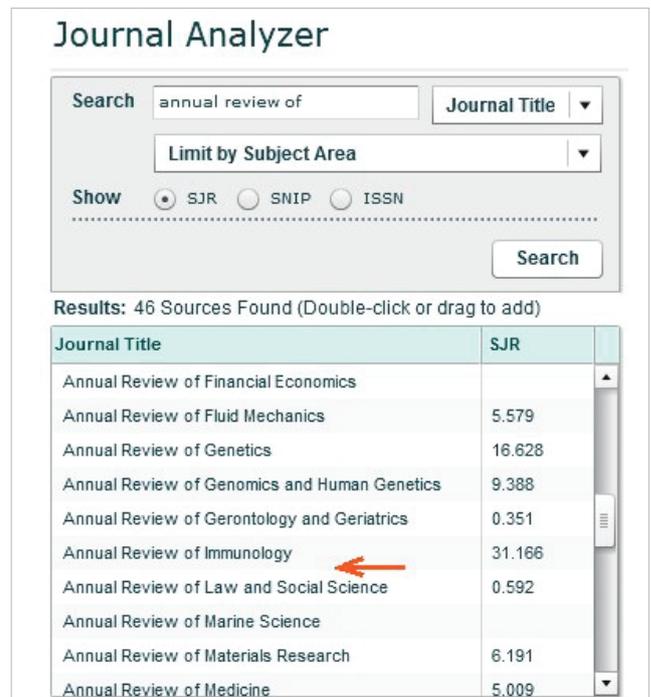


Figure 4c. Enigmatic absence of *Arist* among other annual reviews in *Scopus* Journal Analyzer

Oddly, the *Annual review of information science & technology (Arist)*, which would have had the top ranking position by several indicators for its 2008-2011 volumes, is not included in the LIS category, nor in any other categories of the most current edition of *SJR* (figure 4a).

Scopus has had continuous coverage (at least for the past decade) of the 12-15 *Arist* chapters (figure 4b) typically issued every year, so its complete omission from the *SJR* is enigmatic. It was included in *SJR*'s earlier editions, and in a review four years ago (Jacsó, 2009b) –ironically– it was used as an illustration for the content-rich and very well presented *SJR* league lists. An attempt to use the Journal Analyzer module of *Scopus* to trace this oddity suggested that the problem is on the *Scopus* side, which does not show *Arist* among the 46 Annual Reviews series (figure 4c). The concern is not merely about this particular publication, but other serials which may not show up in any of the league lists for some reason.

Journals to be excluded from the *SJR*-2011 for league lists

There are more than a thousand journals that should have not been included in the 2011 edition simply because data are lacking for one or more years in the 2008-2011 volumes. That should have been a primary filtering criterion for importing *Scopus* data. Such journals do not qualify for *SJR*-2011 if the gaps were caused by inconsistent indexing in *Scopus*.

The most handicapped journals in this scenario are those for which data are not at all available for the first year of the target time period, i.e. 2008 for *SJR*-2011. Papers published in 2008 had the best chance to accumulate citations over the longest time and therefore to increase the journals' indicator scores. Practically, restricting the import to

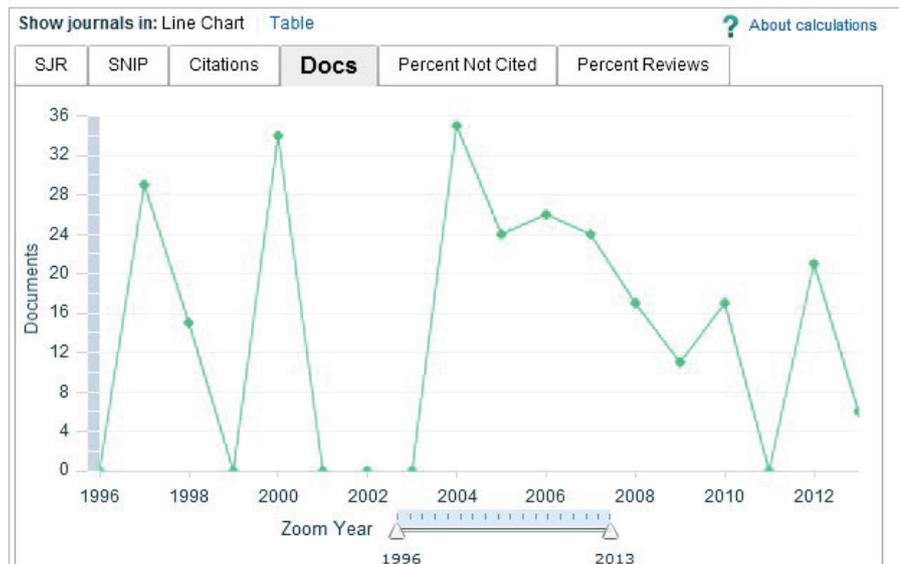


Figure 5. Multiple missing years in the run of the *International journal on digital libraries*

journals with complete data for each year in the assessed time period, i.e. 2008-2011 in case of *SJR-2011*, would be the fairest process. For the LIS league, such exclusion would reduce the number of qualifying journals by 25 and the set of LIS journals to 109.

The *SJR* indicator is based on the algorithm which calculates a size dependent score for each journal's performance. Even if "only" the data for a single year are missing for a journal, this will distort the first step of the calculation and misrepresent the indicator scores if the indexing/abstracting service (*Scopus*) missed an entire year and *SJR* inherits the incomplete data. The exception is if the journal did not publish any papers in one or more years (2001-2003, 2011) –as happened with the *International journal on digital libraries*, for example, in spite of its very appealing title and stellar advisory board. It gives the impression that the journal had a roller-coaster publishing pattern, without the fun of it (figure 5). In such case the journal should not be removed from the league list, as the problem is not caused by missing issues and volumes in *Scopus*. The technical aspects of verifying the "culprit" of such gaps will be discussed in the follow-up paper.

It may be an unpopular, "unpatriotic" suggestion from an LIS professor to remove journals from the LIS subject categories, but it could be gently done by the end-user(s) with the minor additional software features recommended below. At least 50-60 journals could be moved or added to the LIS subject category from several other categories, in addition to those that were mentioned earlier as assigned to several categories but not LIS. This is a scenario where the need to reassign the stray journals is obvious, as they undoubtedly belong to LIS, such as the *Journal of documentation*, *Online information review* or *Journal of information systems*, etc.

In this case, most of them would come from the categories of Information Systems (38), Information Systems and Management (5), and Management Information Systems (6), with a few more from Education, Communication, Media Technology, Law, Business Management and Account-

ing. This is an ultra-conservative count. Of the 54 journals assigned to the Management Information System category, only those with "information management" in the title, rather than "management information", were included, and of the 204 journals in the Computer Science Applications subject category only those dealing directly –in this author's experience- with information science, systems, service, studies (i.e. IS⁴ analogously to LIS⁴ discussed earlier) were counted.

If the time period is extended to 5 years, i.e. 2007-2011, an additional 9 journals would need to be removed from the current LIS subject category for not having

any data for either or both of the two additional years, 2006 and 2007. As discussed earlier, this 5-year time span would allow a comparison with the indicators used in *JCR-2011*, *Eigenfactor-2011*, and *GSMP-2011*. Despite the removal of journals, *SJR* still would have the largest LIS set among those competitors which have the LIS category and show the entire league list of the categories, not just a teaser as *GSMP* does. Adding LIS journals scattered across many other subject categories would give *SJR* by far the largest LIS category, with about 200 active journals. The process of customizing the journal list of categories until a consensus is reached by all parties or by a supermajority of those involved in the decision-making process could be made very efficient by enhancing the *SJR* software.

Customizing the journal set of subject categories

Looking up key indicators of the status of journals is just the first step. The next one is to create a league list from the set of journals needed by the institution for collection development and for the assessment of the publishing performance (productivity and impact) of its research faculty. The option for librarians and other information professionals of modifying the journal lists in existing categories and even creating new categories, permanently or on the fly, would be very useful for producing customized league lists to match the list of preferred or expected publishing outlets of a university, college or department for decisions related to tenure, promotion and grant applications, as well as for serials collection development.

It cannot be expected that the members of the *SCImago Research Group* would fix the inconsistent and inappropriate subject classification errors found in *Scopus*, and no classification system could please all the users all the time. There still would remain journals that should be added to or removed from existing "pret-a-porter" league lists to match the "official" preferences of the different units of an institution. However, the *SCImago Research Group* could offer a built-in software solution for this problem as an add-on utility that would re-rank the league list, keeping the very appealing

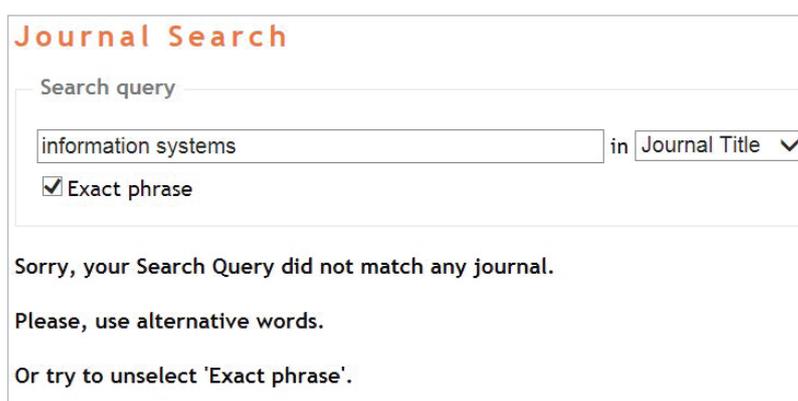
layout and filtering features while providing a customized version of the *SJR* that would please most of the users most of the time at an institution that prefers to create some tailor-made disciplinary league lists from the nearly 20,000 journals.

For example, a Library Studies (LS) category created by the end-users of *SJR* could produce a shorter, much more focused list of perhaps 60-65 journals, removing IS⁴ (information sciences/systems/services/studies), communications, and other journals and adding some LS journals currently in other *SJR*-2011 categories to avoid comparing apples to oranges. Similarly, creating a category for IS⁴ by removing LS journals, and adding IS journals from other categories, would produce a better league list for those institutions where library science is not taught, but courses related to IS⁴ are.

The case of the *University of Hawaii (UH)* may well illustrate this need for upward and downward customization, i.e. broadening and narrowing the journal set. It would require somewhat different league lists for assessing the publishing performance of the faculty and/or doctoral students of the *Library and Information Sciences (LIS) Program*, the *Department of Information & Computer Sciences (ICS)*, and the *Communication and Information Sciences (CIS)* interdisciplinary Ph.D. program. The latter is a joint effort of 4 units, including *LIS*, *ICS*, the *School of Communications* within the *College of Social Sciences*, and the *Department of Information Technology* in the *College of Business*. Substantial differences exist between the different colleges and departments concerning which serials are preferred.

“The *SJImago Research Group* could offer a built-in software solution for this problem as an add-on utility that would re-rank the league list”

In spite of the good cooperation among the representatives of the 4 units in their regular meetings to assess the progress of doctoral students in terms of publishing efforts and achievements, the process is quite time-consuming, as expectations from the candidates differ – depending on their dissertation topic- regarding the preferred publishing venues. These meetings would be more efficient if a customized league list of the preferred journals and conference proceedings could be modified and displayed ad-hoc to check their status indicators in the context of related serials. This would be even more important at the meetings of the system-level tenure and promotion review committees (TPRC) where applications from Astrophysics to Zoology faculty are reviewed and discussed and none of the members may be sufficiently familiar with the standing of the journals in those fields. They may bring up various metrics from different services to support their pro or con arguments without



Journal Search

Search query

information systems in Journal Title ▾

Exact phrase

Sorry, your Search Query did not match any journal.

Please, use alternative words.

Or try to unselect 'Exact phrase'.

Figure 6. Dysfunctional option for exact phrase searching

knowing or explaining the limitations and biases of the widely popular, but non-transparent, services such as *GSMP*.

Software enhancements for customizing journal sets in *SJR*

There are two essential steps for customizing the journal sets assigned to the subject categories for the *Scopus* database and imported into *SJR*. One is to (re)move journal(s) from an inappropriate subject category, such as the *International journal of hospitality and tourism administration* from the LIS category, which needs to be moved to the Tourism, Leisure and Hospitality Management category where it is not included, so simply removing it would make the journal disappear from *SJR*. The other essential step is to locate a journal which is not included in the expected subject category, such as the *Journal of informetrics*, which is assigned to 5 categories but not to LIS, as was discussed above. Searching is critical as the users may not be expected to keep jumping from one category to several others, until they find the journal, or give up the “wild goose chase”.

Locating journals in *SJR*

Finding quickly some journals that don't appear in the league list of the subject area where the user expected to find them is very important, especially considering the deficiencies in the assignment of journals to subject categories. This was the case in the test phase with the LIS category for about 60 journals that are undoubtedly related to the LIS disciplinary field, but were not listed there. The elegance and simplicity of the current user interface for the league tables, the filtering, sorting options and the league list presentation of *SJR* are as appealing as flamenco dancing by the late Antonio Gades. These should be enhanced by a search box with a smart and powerful search engine directly on the template page originally designed for looking up/browsing league tables.

There is a Journal Search module in *SJR* but it should not be used for several reasons. It is a disappointingly primitive component, as if it had been outsourced to an undergraduate student who quit the undergraduate course on Computer Programming 101 after the first session. In many regards this search module performs as poorly as elderly tourists trying their first flamenco steps while watching a show during a sightseeing tour of Seville.



Figures 7a and 7b. Metamorphosis of & into **and** from the cover page in *ScienceDirect* to *Scopus* and *SJR*

The inferior search module merely searches for the character string entered in the search box, bringing up odd results—actually, not searching but scanning for a character string match. For example, for the search term **tax** it retrieves *Taxon*, *Zoofaxa*, *Mycotaxon*, *Syntax*, etc., beyond the relevant journals. To find journals with the title word **reference** it also retrieves journals such as *Food quality and preferences*, *Patient preference and adherence*. There is no explicit truncation symbol, but entering the search term **libr** will find library, libraries, librarian(s), and librarianship. It can be convenient, but because of string matching rather than title word searching, it also retrieves *Fluid phase equilibria*, *Equilibrium research*, and several other journals. The term **online** finds a number of journals with nonlinear in their title, such as *Nonlinear dynamics*, *Nonlinearity*, *Journal of nonlinear science*. Users would still feel that they went for a wild-goose chase as *Online*, the excellent LIS trade journal, is not available in *SJR* even if it has been covered well in *Scopus*.

There is an exact phrase search option using a check box, but it does not work. With the check box activated, the search term **information systems** should find only the journal with that two-word title without additional word(s), but whenever the check box is activated it triggers an error message. Unselecting the exact phrase option will produce a result list of 32 journals which match the character string, but include additional word(s). Searching by ISSN will bring back the wanted journal, but it is not an efficient solution when trying to locate quickly the journals with the term **information systems** in their title, as the result list is sorted by the h-index and cannot be re-sorted by title. Searching by publisher will work, but the data is not current (reflecting the time before *Elsevier*, *Taylor and Francis*, *Springer* went on a shopping spree and acquired dozens of publishers), and for league list purposes this metadata is of no primary importance.

There are no Boolean operators. It is impossible to formulate queries like **age OR aging OR aged** and it adds insult to injury that the string-matching module retrieves many journals with the words image, imaging, language, management, manager, managing, managed, etc. in their title. The query **information AND management** will retrieve two journals where the title includes exactly that 3-letter character string, such as *Information and management* (although the journal itself uses the “&” sign). The journal *Information processing & management* is not retrieved because information and management are not adjacent words in the journal’s title. In other words, AND is not a Boolean AND operator but a character string which should appear in upper or lower case in the journal title to match it. The search term **Information & Management** does not yield any result because the ‘&’ symbol is transformed to the character string ‘and’ in *Scopus*, and *SJR* inherits this spelling “feature” (figures 7a and 7b). These bad traits of the search module—which hinder the *SJR* software—are reminiscent of the first efforts of library automation in the former Soviet Union in the early 1960s.

None of the first efforts of library automation in the former Soviet Union in the early 1960s.

No sorting option is offered for results produced by the current search module. The results lists are by default sorted by the h-index of the journals. This is not clear for the typical end-user because the h-index is not displayed. A good search module should offer all of the sort options available in the league list module (Title, *SJR* score, Total documents, Total cites, Citable documents and Cites/Documents). A smart search engine box located on the league table template could replace the inferior search module. The only feasible way to implement the recommended changes is to empower even the casual user to customize the list and feel/look like a regular ballroom dancer.

Filtering and sorting the search results

The *SJR* league list template offers good options to restrict the search beyond the subject categories. The option for filtering by the country of the journal publishers’ headquarters is not really as relevant as in the *SCR (SCImago Country Report)* database, where country refers to the authors’ affiliation(s), and can provide an approximation (in the context of other countries) for the productivity and impact of the document published by authors’ affiliation. In *SJR* the country name refers to the journal’s place of publication.

Limiting by language of publication could be useful because end-users may want to limit the search in this way (e.g., to only English and/or Spanish language journals), to exclude journals in the other 50+ languages which are unknown and irrelevant for the vast majority of the researchers, universities and libraries around the world. The other option for finding journals is by ISSN and publisher name. Both of these options should be kept in the new smart search engine on the league list template, rather than as a separate function as it is now.

Adding, moving and removing journals as an end-user function

The results list of the league tables should be enhanced in the new search engine by offering an ADD TO, MOVE and REMOVE button next to each item, allowing end-users to make a customized league list to their liking. End-users could just click on the appropriate button and choose from a scroll-down list the subject categories to add or remove the journal, such as adding the *Journal of documentation* to the LIS category, or removing *Development and learning in organisations* from the LIS category. Removing journals from their current category would just require a click on the check-box. It would be the responsibility of the end-users to realize when removing (i.e. deleting) is not the appropriate option for a journal that is assigned only to a single subject category, but a wrong one, as in the case of the *Canadian journal of program evaluation*, which appears only under the LIS category. In such case it must be moved to the appropriate category. It is important to have an optional log-in function or a cookie to keep the customized league lists for later use/updating.

The competent and careful re-assignment of journals to subject categories can make the league lists more fair and credible from the perspective of editors and authors. There is nothing wrong with assigning a journal to more than one subject category, but mis-assignment of journals based solely on *Elsevier's* judgment can unjustly handicap a journal, because its rank position depends much on the company of the journals in the debatable category, especially if that is the only category to which the journal is assigned. The journal keeps the same indicator set independent of the category where it ends up, but it may be similar to assigning a boxer to the Bantamweight category when he qualifies to the Flyweight category.

A summary was created to compare the key metrics indicators of the current LIS category and of the selected subset of 38 journals from the 182 journals assigned to the Information Systems (ISY) category and intended to be added to the LIS category. The mean, median and maximum scores clearly indicate that the journals in the ISY category have significantly higher scores, more like Flyweight versus Heavyweight contenders. If an LIS journal is assigned to the ISY category only, its rank position is much handicapped by all eight metrics.

Adding the 38 journals from the ISY category to the LIS category would produce better rank positions for them. For example, the *Journal of electronic publishing* currently assigned to the ISY subject category within the Computer Science broad subject area is in the bottom quartile (Q4) by *SJR* score; it would be in the lower quartile (Q3) if it were moved or added to the LIS category within the Social Sciences broad subject area. The same is true for the *Canadian journal of information and library science*. The one quartile change is from Q3 to Q2 (upper quartile) for *Technical services quarterly*.

SJR uses Q1 for top, Q2 for upper, Q3 for lower and Q4 for bottom quartile, which is a less common notation than the reverse (Q4 for top, Q3 for upper quartile, etc.). A flip-flop option or a cookie setting notation for this as well as for the decimal fraction and thousand separator notation from the Continental to the US system would be useful as part of the one-time customization process.

For most of the indicators, a similar option to change from the quartile to the percentile system could provide a closer distinction among the journals' standings within a league list as another layout option for those who are less interested in the absolute unit values than in the easily interpreted, directly comparable percentile ranks. Showing the measures of central tendency and the range on the top of the league list could also offer a set of traditional statistical indicators that would help the user get a better feel for the context.

Conclusions

SJR is the largest open access directory, covering nearly 20,000 journals and other serial publications. It is among the best-designed services, with an intuitive interface, browsable indexes, sorting and filtering options, and compact but information-rich journal league lists in 300 subject areas. The entire content can be downloaded in one fell swoop into an Excel file for further processing by any spreadsheet software.

It needs a far better search engine to help in finding journals that are not always assigned by *Elsevier* to the most reasonable, most likely or expected categories. A journal may be ranked far lower by many of the metrics indicators in one subject category than in the other much more appropriate category because of the syndrome of comparing apples to oranges. End-users should be empowered to customize the

Figure 8. Measures of central tendency for key metrics for the journal in the LIS versus the ISY subject categories

Title	SJR score	H index	Total docs. (2011)	Total docs. (3 yrs)	Total refs. 2011	Total cites (3 yrs)	Cites/Doc. (2 yrs)	Refs./Doc
Library and Information Science (LIS) alone: 134 journals								
mean	0.419	11	37	105	891	89	0.65	24.13
median	0.276	6	29	82	523	30	0.44	22.74
max	3.652	78	274	805	9,370	1,614	4.01	61.91
Information Systems alone: 38 journals								
mean	1.058	20	50	123	2,111	349	2.04	46.74
median	0.555	14	29	69	1,045	95	1.50	38.92
max	10.411	94	409	1,051	14,892	4,670	14.43	263.00

journal lists easily online to meet their expectations/preferences, while retaining all the existing browsing and search features and the superb design and implementation of the *SJR* layout. It has given the Spanish *SCImago Research Development Group*, and LIS community, a good reputation. Relatively minor software enhancements by the software developers could make the service even more popular in collection development and in metrics-based research evaluation through proxy movement, where the league list position of the publishing outlets of individuals, groups, departments and entire universities strongly influence decisions related to career advancements, accreditation and funding applications.

The ontology of *SJR* subject areas and categories has weak points and shows negligence, as does the assignment of journals to subject categories. These are inherited from the content and structure of the journal list compiled by *Elsevier* for the *Scopus* database, which has had many innovative software features and many shortcomings in content. There are still significant gaps in the breadth of coverage of the source publications for *Scopus*, which are inherited by *SJR*. These may have a significant effect on the rank position of journals. A follow-up paper will discuss this issue.

References

- Bar-Ilan, Judit** (2010). "Rankings of information and library science journals by JIF and by h-type indices". *Journal of informetrics*, v. 4, n. 2, pp. 141-147.
<http://dx.doi.org/10.1016/j.joi.2009.11.006>
- Bergman, Elaine M. Lasda** (2012). "Finding citations to social work literature: the relative benefits of using *Web of Science*, *Scopus*, or *Google Scholar*". *The journal of academic librarianship*, v. 38, n. 6, pp. 370-379.
<http://dx.doi.org/10.1016/j.acalib.2012.08.002>
- De-Moya-Anegón, Félix; Chinchilla-Rodríguez, Zaida; Vargas-Quesada, Benjamín; Corera-Álvarez, Elena; Muñoz-Fernández, Francisco-José; González-Molina, Antonio; Herrero-Solana, Víctor** (2007). "Coverage analysis of Scopus: A journal metric approach". *Scientometrics*, v. 73, n. 1, pp. 53-78.
<http://dx.doi.org/10.1007/s11192-007-1681-4>
- Deis, Louise; Goodman, David** (2005). "Web of Science (2004 version) and Scopus". *The Charleston advisor*, v. 6, n. 3, pp. 5-21.
- Delgado-López-Cózar, Emilio; Cabezas-Clavijo, Álvaro** (2012). "Google Scholar Metrics: an unreliable tool for assessing scientific journals". *El profesional de la información*, v. 21, n. 4, pp. 419-427.
<http://dx.doi.org/10.3145/epi.2012.jul.15>
- Delgado-López-Cózar, Emilio; Cabezas-Clavijo, Álvaro** (2013). "Ranking journals: could Google Scholar Metrics be an alternative to Journal Citation Reports and SCImago Journal Rank?". *Learned publishing*, v. 26, n. 2, pp. 101-113.
<http://dx.doi.org/10.1087/20130206>
- Dess, Howard M.** (2006). "Database reviews and reports. Scopus". *Issues in science and technology librarianship*, Winter.
<http://www.istl.org/06-winter/databases4.html>
- Falagas, Matthew E.; Kouranos, Vasilios D.; Arencibia-Jorge, Ricardo; Karageorgopoulos, Drosos E.** (2008). "Comparison of SCImago Journal Rank indicator with Journal Impact Factor". *The FASEB journal*, v. 22, n. 8, pp. 2623-2628.
<http://dx.doi.org/10.1096/fj.08-107938>
- Garfield, Eugene** (1999). "Journal impact factor: a brief review". *Canadian Medical Association Journal*, v. 161, n. 8, pp. 979-980.
<http://www.garfield.library.upenn.edu/papers/journalimpactCMAJ1999.pdf>
- Garfield, Eugene** (2005). "The agony and the ecstasy - The history and the meaning of the Journal Impact Factor". In: *The intl congress on peer review and biomedical publication*. Chicago, USA. September 16.
<http://www.garfield.library.upenn.edu/papers/jifchicago2005.pdf>
- Gómez-Núñez, Antonio J.; Vargas-Quesada, Benjamín; De-Moya-Anegón, Félix; Glänzel, Wolfgang** (2011). "Improving SCImago Journal & Country Rank (SJR) subject classification through reference analysis". *Scientometrics*, v. 89, n. 3, pp. 741-758.
<http://dx.doi.org/10.1007/s11192-011-0485-8>
- González-Pereira, Borja; Guerrero-Bote, Vicente P.; De-Moya-Anegón, Félix** (2009). "The SJR indicator: A new indicator of journals' scientific prestige".
<http://arxiv.org/ftp/arxiv/papers/0912/0912.4141.pdf>
- González-Pereira, Borja; Guerrero-Bote, Vicente P.; Moya-Anegón, Félix** (2010). "A new approach to the metric of journals' scientific prestige: The SJR indicator". *Journal of informetrics*, v. 4, n. 3, pp. 379-391.
<http://dx.doi.org/10.1016/j.joi.2010.03.002>
- González-Pereira, Borja** (2011). Personal communication.
- Jacsó, Péter** (2007). "Scopus", *Gale: Péter's digital reference shelf*, 2007, November.
<http://jacso.info/PDFs/Jacso-Scopus-2007-11-Gale.pdf>
- Jacsó, Péter** (2008). "The pros and cons of computing the h-index using Scopus". *Online information review*, v. 32, n. 4, pp. 524-535.
<http://dx.doi.org/10.1108/14684520810897403>
- Jacsó, Péter** (2009a). "Errors of omission and their implication for computing scientometric measures in evaluating the publishing productivity and impact of countries". *Online information review*, v. 33, n. 2, pp. 376-385.
<http://dx.doi.org/10.1108/14684520910951276>
- Jacsó, Péter** (2009b). "SCImago, Worldmapper, Atlapedia". *Online*, v. 33, n. 3, pp. 50-54.
- Jacsó, Péter** (2009c). "Scopus", *Gale: Péter's digital reference shelf*, June.
<http://www.gale.cengage.com/reference/peter/200906/scopus.html>
- Jacsó, Péter** (2010). "Comparison of journal impact rankings in the SCImago Journal & Country Rank and Journal Citation Reports databases". *Online information review*, v. 34, n. 4, pp. 642-657.
<http://dx.doi.org/10.1108/14684521011073034>

Jacsó, Péter (2012). "Google Scholar Metrics for Publications – The software and content feature of a new open access bibliometric service". *Online information review*, v. 36, n. 4, pp. 604-609.
<http://dx.doi.org/10.1108/14684521211254121>

Jacsó, Péter (2013). "The role of information professionals in interpreting the results of journal league lists for research performance evaluation" (Prologue), *Anuario ThinkEPI*, v. 7, pp. 11-15.
<http://thinkpi.net>

Moed, Henk F. (2010). "Measuring contextual citation impact of scientific journals". *Journal of informetrics*, v. 4, n. 3, pp. 265-277.
<http://dx.doi.org/10.1016/j.joi.2010.01.002>

Oppenheim, Charles (2008). "Out with the old and in with the new: the RAE, bibliometrics and the new REF". *Journal of librarianship and information science*, v. 40, n. 3, pp. 147-149.
<http://dx.doi.org/10.1177/0961000608092550>

Pudovkin, Alexander I.; Garfield, Eugene (2004). "Rank-normalized impact factor: A way to compare journal performance across subject categories". *Proceedings of the ASIST annual meeting*, v. 41, pp. 507-515.

Pudovkin, Alexander I.; Garfield, Eugene (2009). "Percentile rank and author superiority indexes for evaluating individual journal articles and author's overall citation performance". Presented at: *Fifth intl conf on webometrics & scientomet-*

rics (WIS). Tenth Collnet Meeting, Sept.13-16, Dalion, China.

Raj, Ram-Gopal; Zainab, Awang N. (2012). "Relative measure index: a metric to measure the quality of journals". *Scientometrics*, v. 93, n. 2, pp. 305-317.
<http://dx.doi.org/10.1007/s11192-012-0675-z>

Torres-Salinas, Daniel; Jiménez-Contreras, Evaristo (2010). "Introducción y estudio comparativo de los nuevos indicadores de citación sobre revistas científicas en *Journal Citation Reports* y *Scopus*". *El profesional de la información*, v. 19, n. 2, pp. 201-207.
<http://dx.doi.org/10.3145/epi.2010.mar.12>

Wagner, A. Ben (2009). "Percentile-based journal impact factors: a neglected collection development metric". UB Faculty and Staff Publications.
<http://ubir.buffalo.edu/xmlui/handle/10477/1318>

Waltman, Ludo; Van-Eck, Nees-Jan; Van-Leeuwen, Thed N.; Visser, Martijn S. (2012). "Some modifications to the SNIP journal impact indicator". *Journal of informetrics*, v. 7, n. 2, pp. 272-285.
<http://dx.doi.org/10.1016/j.joi.2012.11.011>
<http://arxiv.org/ftp/arxiv/papers/1209/1209.0785.pdf>

Weinberger, David (2007). *Everything is miscellaneous: the power of the new digital disorder*. Henry Holt and Company, LLC. ISBN: 0 8050 8811 3

4ª Conferencia sobre calidad de revistas de ciencias sociales y humanidades



CRECS 2014
8-9 mayo
<http://www.thinkpi.net/crecs2014>

Salón de Actos de la Casa del Lector
(Antiguo Matadero Madrid, Paseo de la Chopera, 10)

El profesional de la
información

Fundación Germán
Sánchez Ruipérez



EXCELLENCE WITH LEADERSHIP: THE CROWN INDICATOR OF SCIMAGO INSTITUTIONS RANKINGS IBER REPORT

Veljko Jeremić, Marina Jovanović-Milenković, Zoran Radojičić, and Milan Martić

Nota: Este artículo puede leerse traducido al español en:
http://www.elprofesionaldelainformacion.com/contenidos/2013/sept/13_esp.pdf



Veljko Jeremić is an assistant professor at the *Department of Operational Research and Statistics* at the *Faculty of Organizational Sciences, University of Belgrade*. He was born in 1985, finished his PhD thesis in the field of computational statistics in the 2012. He has published over 60 scientific papers with emphasis on *ISI* indexed journals (over 20 papers). He is co-author of one university textbook. He has acted as a reviewer for many *ISI* indexed journals such as: *Jasist*, *Scientometrics*, *Journal of applied statistics*.

<http://orcid.org/0000-0002-7761-1841>

jeremic.veljko@fon.bg.ac.rs



Marina Jovanović-Milenković is an assistant professor at the *Department of Information Technologies* at the *Faculty of Organizational Sciences, University of Belgrade*. She was born in 1976, finished her PhD thesis in the field of information and communication technology in health care management in 2011. She published over 30 scientific papers with emphasis on *ISI* indexed journals (5 papers). She has participated in the organization of the doctoral degree study program *Management of the health system*. Her major interests are e-health, ICT implementations in health systems and business decision making.

<http://orcid.org/0000-0003-3687-5282>

marinaj@fon.bg.ac.rs



Zoran Radojičić is an associate professor at the *Department of Operational Research and Statistics* at the *Faculty of Organizational Sciences, University of Belgrade*. He defended his PhD thesis in the field of computational statistics in 2007. Main research interests include: statistics, applied statistics, social sciences and biostatistics. He has published over 40 papers at *SCI Thomson Reuters* database. He has participated in *Tempus* and numerous projects supported by the *Ministry of Science of Republic of Serbia*. He was included in international projects financed by *World Bank* and *EBRD* in the region.

<http://orcid.org/0000-0002-1536-8591>

radojicic.zoran@fon.bg.ac.rs



Milan Martić is a dean and full professor at the *Department of Operational Research and Statistics* at the *Faculty of Organizational Sciences, University of Belgrade*. He defended his PhD thesis in the field of operational research in 1999. He has published over 130 scientific papers with emphasis on *ISI* indexed journals (over 12 papers). He is co-author of seven university textbooks. He has acted as a reviewer for many *ISI* indexed journals. His major interests are data envelopment analysis, efficiency models and operational research.

<http://orcid.org/0000-0002-0289-6678>

martic.milan@fon.bg.ac.rs

University of Belgrade, Faculty of Organizational Sciences, Jove Illica 154, 11000 Belgrade, Serbia

Abstract

Although there are many models for ranking higher education institutions, the *SCImago Institutions Rankings* methodology stands out for its ability to present quantitative and qualitative indicators of scientific output. Besides *Total number of published papers*, several indicators are concerned with quality dimensions of published papers, such as *International collaboration*, *Scientific leadership* or *High quality publications*. However, official rankings are provided solely on the basis of one indicator: *Output (total number of published papers)*. This paper presents a statistical I-distance method that integrates all the indicators into one value, which therefore represent a rank and show which of the input indicators is the most important for the process of ranking. Our results clearly showed that *Excellence with Leadership* occupies the most significant spot.

Article received on 22-04-2013

Final acceptance: 01-07-2013

Keywords

Ranking of universities, I-distance method, *SCImago Institutions Rankings Iber reports*, Statistical methods.

Título: Excelencia con liderazgo: el indicador rey del *SCImago Institutions Rankings Iber report*

Resumen

Aunque hay muchos modelos para clasificar instituciones de educación superior, la metodología de *SCImago Institutions Rankings* destaca por su capacidad de presentar indicadores cuantitativos y cualitativos de la producción científica. Además del *Número total de artículos publicados*, varios indicadores se refieren a aspectos de calidad de los trabajos, como *Colaboración internacional*, *Liderazgo científico* o *Publicaciones de alta calidad*. Sin embargo, las clasificaciones oficiales se ofrecen únicamente en base a un indicador: *Resultados* (número total de artículos publicados). Este trabajo presenta el método de I-distancia estadística que integra todos los indicadores en un valor, lo que facilita el establecimiento de una lista o ranking, y muestra cuál de los indicadores es el más importante para el proceso de clasificación. Los resultados muestran claramente que *Excelencia con liderazgo* es el más influyente.

Palabras clave

Rankings de universidades, Método I-distancia, *SCImago Institutions Rankings Iber reports*, *SIR*, Métodos estadísticos.

Jeremić, Veljko; Jovanović-Milenković, Marina; Radojičić, Zoran; Martić, Milan (2013). "Excellence with Leadership: the crown indicator of *SCImago Institutions Rankings Iber report*". *El profesional de la información*, September-October, v. 22, n. 5, pp. 474-480.

<http://dx.doi.org/10.3145/epi.2013.sep.13>

1. Introduction

The increasing number of methodologies for ranking higher education institutions (HEI) has attracted many different stakeholders, especially students. Consequently, those rankings are quite often used as an indicator of a university's reputation and performance (Agasisti; Pérez-Esparrells, 2010; Altbach, 2013; Bonaccorsi; Daraio, 2008; Bowman; Bastedo, 2011; Hazelkorn, 2011; Hien, 2010; Jeremić *et al.*, 2011; Marginson, 2011; Sadlak, 1978; Salmi, 2003; Torres-Salinas *et al.*, 2011; García *et al.*, 2012). One can argue that "University rankings are very appealing, in that they provide a single number that allows, at a glance, to situate a given university in the worldwide context. However, this very simplicity of use can be highly misleading in that most rankings are based on a simple formula that aggregate subjectively chosen indicators" (Saisana; D'Hombres, 2008). Almost certainly, the most cited ranking list is the *Academic Ranking of World Universities (ARWU)*, which has been the focus of researchers (Paruolo *et al.*, 2013; Saisana; D'Hombres; Saltelli, 2011) since its first creation in 2003 (Aguillo *et al.*, 2010; Dehon; McCathie; Verardi, 2010; Jovanović *et al.*, 2012; Docampo, 2008, 2011; Docampo *et al.*, 2012). Yet, almost immediately after the release of its first ranking, the *ARWU* attracted a great deal of criticism (Docampo, 2012, 2013; Billaut; Bouyssou; Vincke, 2010). One of the potential weaknesses frequently highlighted (Nishy *et al.*, 2012; Prathap, 2012a, 2013; Radojičić; Jeremić, 2012) is the absence of scientific quality indicators such as high quality papers (those ranked in the first quartile ~ 25% ~ in their categories), etc. Thus, the latest release of the *SCImago Institutions Rankings (SIR) Iber reports (SIR, 2013)*, which quantifies the research performance of 1,600 leading research institutions of Ibero-American countries, brings even more to the table (Prathap, 2012b).

<http://www.scimagoir.com>

The *SIR* approach integrates one quantitative and various qualitative variables. The *Output (O)* indicator is a measure of the quantity of an institution's publication output. It represents the total number of documents published in scholarly journals indexed in Scopus (Romo-Fernández *et al.*, 2011).

Seven other variables represent the quality dimension of scientific output: *International collaboration (IC)*, *Normalized impact (NI)*, *High quality publications (Q1)*, *Specialization index (SI)*, *Excellence rate (ER)*, *Scientific lead (Lead)* and *Excellence with leadership (Ewl)*.

The *IC* variable represents an institution's output ratio produced in collaboration with foreign institutions. The values are computed by analyzing output with affiliations including more than one country address (Guerrero-Bote; Olmeda-Gómez; De-Moya-Anegón, 2013; Lancho-Barrantes; Guerrero-Bote; De-Moya-Anegón, 2013). However, the question raised is whether this characteristic actually provides insight into scientific output, since cooperation of Serbian and Montenegrin universities, for instance, is international while these countries were until recently one state union. Further, *NI* compares the average scientific impact of the institution with the world average (taken as 1). Thus, a score of 0.8 implies a performance 20% below average, whereas a score of 1.3 means the institution is considered 30% above average (González-Pereira; Guerrero-Bote; De-Moya-Anegón, 2010). One can argue that since "citation is counted only if it is made to an item published in the three previous years", some important citations have been excluded (Geta-chew-Dinku, 2011).

Also, *Q1* is the ratio of publications that the institution publishes in what the *SCImago* team takes as the most influential scholarly journals of the world: those ranked in the first quartile (25%) in their categories as calculated by *SCImago*

Journal Rank. Since this is reported as a percentage, the ratio (Q1/25) is a crude normalized proxy for quality of publication, with a value of 1 taken as the world average (Miguel; Chinchilla-Rodríguez; De-Moya-Anegón, 2011). It is essential to mention that *SJR* doesn't include journals' self-cites (as done by Thomson Reuters *IF*) and the weight of citations depends upon the "prestige" of the citing journal (Falagas et al., 2008).

The *SI* variable indicates the extent of thematic concentration/dispersion of an institution's scientific output. The values of this indicator range between 0 and 1, indicating more generalist or specialized institutions, respectively (López-Illescas; De-Moya-Anegón; Moed, 2011). On the other hand, *ER* indicates the percentage of an institution's scientific output that is included in the set formed by 10% of the most-cited papers in their respective scientific fields. This indicator serves as a measure of the high-quality output of research institutions. Again, the ratio *ER/10* allows one to normalize this indicator so that the world average becomes 1 (Bornmann; De-Moya-Anegón; Leydesdorff, 2012). *Lead* indicates an institution's "output as main contributor", that is the number of papers in which the corresponding author belongs to the institution (De-Moya-Anegón, 2012). Finally, *Ewl* indicates the amount of documents in the *Excellence rate* in which the institution is the main contributor (*SIR*, 2013).

Nonetheless, although the *SIR Iber report* presents all the valuable data, the official rankings are presented based only on the number of *Total published papers* (indicator *Output - O*). Having said this, it is essential to provide a potential upgrade of current framework and create a synthesised indicator that will incorporate both the quantitative and qualitative dimensions of *SIR Iber report*. Further, it is vital to establish which of these dimensions provides better insight into scientific excellence of a HEI. As a possible remedy to the issue, the statistical I-distance method is elaborated and applied.

2. I-distance method

Quite frequently, the score obtained in a specific league list can seriously affect the process of taking exams, entering competitions, UN projects participation, medicine selection and many other areas (Jeremić; Radojičić, 2010; Al-Lagilli et al., 2011). I-distance is a metric distance in an n-dimensional space. It was originally proposed and defined by Branislav Ivanović, and has appeared in various publications since 1963 (Ivanović, 1977). Ivanović devised this method to rank countries according to their level of development on the basis of several indicators; many socio-economic development indicators had been considered and the problem was how to use all of them in order to calculate a single synthetic indicator which would thereafter represent the rank.

For a selected set of variables $X^I=(X_1, X_2, \dots, X_k)$ chosen to characterize the entities, the I-distance between the two entities $e_r=(X_{1r}, X_{2r}, \dots, X_{kr})$ and $e_s=(X_{1s}, X_{2s}, \dots, X_{ks})$ is defined as

$$D(r, s) = \sum_{i=1}^k \frac{|d_i(r, s)|}{\sigma_i} \prod_{j=1}^{i-1} (1 - r_{j1.12\dots j-1})$$

where $d_i(r, s)$ is the distance between the values of variable X_i for e_r and e_s , e.g. the discriminate effect,

$$d_i(r, s) = x_{ir} - x_{is}, i \in \{1, \dots, k\}$$

σ_i the standard deviation of X_i , and $r_{j1.12\dots j-1}$ is a partial coefficient of the correlation between X_i and X_j ($j < i$), (Bulajić et al., 2012; Dobrota; Jeremić; Marković, 2012).

The construction of the I-distance is iterative; it is calculated through the following steps:

- Calculate the value of the discriminate effect of the variable X_1 (the most significant variable, that which provides the largest amount of information on the phenomena that are to be ranked).
- Add the value of the discriminate effect of X_2 which is not covered by X_1
- Add the value of the discriminate effect of X_3 which is not covered by X_1 and X_2
- Repeat the procedure for all variables (Jeremić et al., 2012; Radojičić et al., 2012).

Sometimes, it is not possible to achieve the same sign mark for all variables in all sets, and, as a result, a negative correlation coefficient and a negative coefficient of partial correlation may occur (Jeremić et al., 2011; Maletić et al., 2012). This makes the use of the square I-distance even more desirable. The square I-distance is given as:

$$D^2(r, s) = \sum_{i=1}^k \frac{d_i^2(r, s)}{\sigma_i^2} \prod_{j=1}^{i-1} (1 - r_{j1.12\dots j-1}^2)$$

In order to rank the entities (in this case, universities), it is necessary to have one entity fixed as a referent in the observing set using the I-distance methodology (Jeremić et al., 2012; Jovanović et al., 2012). The entity with the minimal value for each indicator or a fictive minimal entity should be utilized as the referent entity, as the ranking of the entities in the set is based on the calculated distance from the referent entity (Seke et al., 2013).

3. Results of the I-distance method

For this study, the latest release of the *SCImago Institutions Rankings (SIR) Iber reports (SIR, 2013)* was analyzed. Out of the 1,600 leading research institutions ranked in the *SIR 2013 Iber reports*, 148 universities with the *Output* indicator larger than 1,000 papers were selected and the I-distance method was performed on that sub dataset. The results achieved by means of the squared I-distance method for the first 20 HEI are shown below in table 1 (full list is available upon request). The construction of the I-distance uses an iterative approach, and the crucial idea is to reduce the duplication of information. It is done by partially integrating the variables into the creation of one value which will therefore represent the rank. Particularly interesting is that variables with different types of measurements (percentages, GDP, student enrolment rate, etc.) could easily be integrated into one variable.

As can be seen from table 1, *Universidade de São Paulo* tops the I-distance method. This university has an impressive *Output* indicator, with 47,833 published papers. On the other hand, all the other indicators are solid but far from the best. One should note that the *Universitat Rovira i Virgili* is highly placed (2nd spot), although it has a rather small number of published papers – just 4,186 (less than 10% of *São Paulo*'s output). Precisely this information is crucial because it is essential to elaborate other variables in which the *Universitat Rovira i Virgili* impresses. For instance, in a quality indicator such as *Excellence with leadership* ~ *Ewl* (10.6) and *Excellence rate* ~ *ER* (17.17), the *Universitat Rovira i Virgili* is one of the best universities (number one concerning the indicator “*Excellence with leadership*”, third in the “*Excellence rate*”). A similar conclusion has been noted by **Radojčić & Jeremić** (2012), for *Rockefeller University*, a postgraduate-only institution with a small number of published papers but fully impressive in quality indicators such as *Q1* (88.6%) and *ER* (48.8).

Consequently, it is essential to determine which of the eight input indicators is the most important for the process of ranking. Thus, this data set has been further examined and the correlation coefficients of each variable with the I-distance values have been determined. The results shown in table 2 demonstrate that the most significant variable for the calculated I-distance value is *Excellence with leadership (Ewl)*, highly correlated with the I-distance value ($r = 0.676$, $p < 0.01$). *Total number of published papers (Output)*, *Excellence rate (ER)*, *Normalized impact (NI)* and *High quality publications (Q1)* are also very important indicators, with each correlation larger than 0.5 ($p < 0.01$). This find-

Table 1. Results of the squared I-distance method for HEI provided in SIR 2013 Iber report (first 20 placed HEI)

Rank	Higher education institution	State	I-distance
1	<i>Universidade de São Paulo</i>	BRA	90.858
2	<i>Universitat Rovira i Virgili</i>	ESP	45.542
3	<i>Universidad Politécnica de Valencia</i>	ESP	42.071
4	<i>Universitat de Lleida</i>	ESP	39.112
5	<i>Universidade de Vigo</i>	ESP	38.452
6	<i>Universitat Politècnica de Catalunya</i>	ESP	37.253
7	<i>Universitat Jaume I</i>	ESP	36.711
8	<i>Universidad Pública de Navarra</i>	ESP	34.533
9	<i>Universidad Carlos III de Madrid</i>	ESP	34.283
10	<i>Universidade do Minho</i>	PRT	34.034
11	<i>Universitat d'Alacant</i>	ESP	33.583
12	<i>Universidad Nacional del Litoral</i>	ARG	33.401
13	<i>Universidad Politécnica de Cartagena</i>	ESP	33.298
14	<i>Universitat de Barcelona</i>	ESP	33.035
15	<i>Universidad Técnica Federico Santa María</i>	CHL	32.803
16	<i>Universidad de Sevilla</i>	ESP	32.323
17	<i>Universidad de Castilla-La Mancha</i>	ESP	32.287
18	<i>Universitat de les Illes Balears</i>	ESP	32.052
19	<i>Universidade de Aveiro</i>	PRT	31.984
20	<i>Universidad de Zaragoza</i>	ESP	31.686

Table 2. The correlation between input variables and I-distance values

	I-distance
Excellence with leadership (Ewl)	0.676**
Output (O)	0.589**
Excellence rate (ER)	0.561**
Normalized impact (NI)	0.533**
High quality publications (Q1)	0.522**
Scientific lead (Lead)	0.444**
International collaboration (IC)	0.302*
Specialization index (SI)	0.286*

** $p < 0.01$, * $p < 0.05$

ing clearly shows that the qualitative dimension of scientific output must not be neglected at any cost.

4. Concluding remarks

The increasing number of ranking methodologies used to identify the world's best universities is pushing the academic world into becoming even more concerned with the assessment of higher education. With these rankings often used as a marketing tool for universities to show their educational or research excellence, the need to provide rankings as accurate as possible becomes exceptionally important (**Radojčić; Jeremić**, 2012). The analysis presented here stresses potential improvements in the *SCImago Ranking* methodology, with emphasis on the quality indicator of universities' scientific output. The idea is consistent with the approach taken by **Torres-Salinas et al.** (2011), which presents the IFQ2A index as an integrated index of qualitative and quantitative scientific indicators. In addition, the *Leiden ranking* (Leiden, 2013) has some similarities with *SCImago* in terms of bibliometric data, and rankings focused on the research performance of institutions. However, in *Leiden ranking* the journals that do not have a strong international scope (i.e., not published in English or articles with authors concentrated in one or a few countries) or have a small number of references to other journals in the *Web of Science* database are being excluded from the analysis (**Waltman et al.**, 2012). Perhaps this could be a way to go for following *SIR* lists. Moreover, by default the *Leiden ranking* reports size-independent indicators (average statistics per publication, such as a university's average number of citations per publication). The advantage of size-independent indicators is that they enable comparisons between smaller and larger universities (Leiden, 2013). As an alternative to size-independent indicators, the *Leiden ranking* can also report size-dependent indicators, which provide overall statistics of the publications of a university (the total number of citations of the publications of a university). Size-dependent indicators are strongly influenced by the size of a university and therefore tend to be less useful for comparison purposes (**Waltman et al.**, 2012). Furthermore, our approach could contribute to the emerging efforts to map regions of academic excellence and scientific output (**Bornmann; Leydesdorff**, 2011; **Bornmann; Waltman**, 2011; **Prathap**, 2011). Additionally, I-distance methodology could embrace the postulates of bootstrapping and we hope to encourage potential researchers in this area.

Acknowledgment

The authors would like to thank the editor and anonymous reviewers for their valuable comments and suggestions, which significantly improved the quality of the paper.

5. References

Agasisti, Tommaso; Pérez-Esparrells, Carmen (2010). "Comparing efficiency in a cross-country perspective: the case of Italian and Spanish state universities". *Higher education*, v. 59, n. 1, pp. 85-103.

<http://dx.doi.org/10.1007/s10734-009-9235-8>

Aguillo, Isidro F.; Bar-Ilan, Judit; Levene, Mark; Ortega, José-Luis (2010). "Comparing university rankings". *Scientometrics*, v. 85, n. 1, pp. 243-256.

<http://dx.doi.org/10.1007/s11192-010-0190-z>

Altbach, Philip (2013). *The international imperative in higher education*. Rotterdam: SensePublishers.

<https://www.sensepublishers.com/media/1661-the-international-imperative-in-higher-education.pdf>

Al-Lagilli, Sliman; Jeremić, Veljko; Seke, Kristina; Jeremić, Danko; Radojčić, Zoran (2011). "Evaluating the health of nations: a Libyan perspective". *Libyan journal of medicine*, v. 6, 6021.

<http://dx.doi.org/10.3402/ljm.v6i0.6021>

Billaut, Jean-Charles; Bouyssou, Denis; Vincke, Philippe (2010). "Should you believe in the Shanghai ranking: an MCDM view". *Scientometrics*, v. 84, n. 1, pp. 237-263.

<http://dx.doi.org/10.1007/s11192-009-0115-x>

Bonaccorsi, Andrea; Daraio, Cinzia (2008). "The differentiation of the strategic profile of higher education institutions. New positioning indicators based on microdata". *Scientometrics*, v. 74, n. 1, pp. 15-37.

<http://dx.doi.org/10.1007/s11192-008-0101-8>

Bornmann, Lutz; De-Moya-Anegón, Félix; Leydesdorff, Loet (2012). "The new excellence indicator in the World Report of the SCImago Institutions Rankings 2011". *Journal of Informetrics*, v. 6, n. 2, pp. 333-335.

<http://www.leydesdorff.net/scimago11>

<http://dx.doi.org/10.1016/j.joi.2011.11.006>

Bornmann, Lutz; Leydesdorff, Loet (2011). "Which cities produce more excellent papers than can be expected? A new mapping approach —using Google Maps— based on statistical significance testing". *Journal of the American Society of Information Science and Technology*, v. 62, n.10, pp. 1954-1962.

<http://dx.doi.org/10.1002/asi.21611>

Bornmann, Lutz; Waltman, Ludo (2011). "The detection of 'hot regions' in the geography of science: a visualization approach by using density maps". *Journal of informetrics*, v. 5, n. 4, pp. 547-553.

<http://arxiv.org/pdf/1102.3862.pdf>

<http://dx.doi.org/10.1016/j.joi.2011.04.006>

Bowman, Nicholas A.; Bastedo, Michael N. (2011). "Anchoring effects in world university rankings: exploring biases in reputation scores". *Higher education*, v. 61, n. 4, pp. 431-444.

<http://dx.doi.org/10.1007/s10734-010-9339-1>

Bulajić, Milica; Knezević, Snezana; Jeremić, Veljko; Zarkić-Joksimović, Nevenka (2012). "Towards a framework for evaluating bank efficiency". *International journal of agricultural and statistical sciences*, 2012, v. 8, n. 2, pp. 377-384.

De-Moya-Anegón, Félix (2012). "Liderazgo y excelencia de la ciencia española". *El profesional de la información*, v. 21, n. 2, pp. 125-128.

<http://dx.doi.org/10.3145/epi.2012.mar.01>

Dehon, Catherine; McCathie, Alice; Verardi, Vincenzo (2010). "Uncovering excellence in academic rankings: A closer look at the Shanghai ranking". *Scientometrics*, v. 83, n. 2, pp. 515-524.

<http://dx.doi.org/10.1007/s11192-009-0076-0>

Dobrota, Marina; Jeremić, Veljko; Marković, Aleksandar (2012). "A new perspective on the ICT Development Index". *Information development*, v. 28, n. 4, pp. 271-280.

<http://dx.doi.org/10.1177/0266666912446497>

Docampo, Domingo (2008). "International rankings and quality of university systems". *Revista de educación*, Special issue, pp. 149-176.

http://www.revistaeducacion.mec.es/re2008/re2008_07.pdf

Docampo, Domingo (2011). "On using the Shanghai ranking to assess the research performance of university systems". *Scientometrics*, v. 86, n. 1, pp. 77-92.

<http://dx.doi.org/10.1007/s11192-010-0280-y>

Docampo, Domingo (2012). "Adjusted sum of institutional scores as an indicator of the presence of university systems in the ARWU ranking". *Scientometrics*, v. 90, n. 2, pp. 701-713.

<http://dx.doi.org/10.1007/s11192-011-0490-y>

Docampo, Domingo (2013). "Reproducibility of the Shanghai academic ranking of world universities". *Scientometrics*, v. 94, n. 2, pp. 567-587.

http://sci2s.ugr.es/rankinguniversidades/prensa/2012/reproducibility_docampo_2012.pdf

<http://dx.doi.org/10.1007/s11192-012-0801-y>

Docampo, Domingo; Herrera, Francisco; Luque-Martínez, Teodoro; Torres-Salinas, Daniel (2012). "Aggregate ranking of Spain's universities in the Shanghai Ranking (ARWU): Effect on autonomous communities and campuses of international excellence". *El profesional de la información*, v. 21, n. 4, pp. 428-432.

<http://hdl.handle.net/10481/21539>

<http://dx.doi.org/10.3145/epi.2012.jul.16>

Falagas, Matthew; Kouranos, Vasilios; Arencibia-Jorge, Ricardo; Karageorgopoulos, Drosos (2008). "Comparison of SCImago journal rank indicator with journal impact factor". *The FASEB journal*, v. 22, n. 8, pp. 2623-2628.

<http://www.fasebj.org/content/22/8/2623.full.pdf+html>

<http://dx.doi.org/10.1096/fj.08-107938>

García, José A.; Rodríguez-Sánchez, Rosa; Fernández-Valdivia, Joaquín; Torres-Salinas, Daniel; Herrera, Francisco (2012). "Ranking of research output of universities on the

- basis of the multidimensional prestige of influential fields: Spanish universities as a case of study". *Scientometrics*, v. 93, n. 3, pp. 1081-1099.
<http://dx.doi.org/10.1007/s11192-012-0740-7>
- Getachew-Dinku, Godana** (2011). *Communication studies theses, dissertations, and student research*. SCImago, Paper 14. Measuring Scholarly Metrics.
<http://digitalcommons.unl.edu/commstuddiss/14>
- González-Pereira, Borja; Guerrero-Bote, Vicente; De-Moya-Anegón, Félix** (2010). "A new approach to the metric of journal's scientific prestige: the SJR indicator". *Journal of informetrics*, v. 4, n. 3, pp. 379-391.
<http://dx.doi.org/10.1016/j.joi.2010.03.002>
- Guerrero-Bote, Vicente P.; Olmeda-Gómez, Carlos; De-Moya-Anegón, Félix** (2013). "Quantifying the benefits of international scientific collaboration". *Journal of the American Society for Information Science and Technology*, v. 64, n. 2, pp. 392-404.
<http://dx.doi.org/10.1002/asi.22754>
- Hazelkorn, Ellen** (2011). *Rankings and the reshaping of higher education, The battle for world-class excellence*. Palgrave MacMillan. ISBN: 978 0230243248
- Hien, P. D.** (2010). "A comparative study of research capabilities of East Asian countries and implications for Vietnam". *Higher education*, v. 60, n. 6, pp. 615-625.
<http://dx.doi.org/10.1007/s10734-010-9319-5>
- Ivanović, Branislav** (1977). *Classification theory*. Belgrade: Institute for Industrial Economics.
- Jeremić, Veljko; Radojičić, Zoran** (2010). "A new approach in the evaluation of team chess championships rankings". *Journal of quantitative analysis in sports*, v. 6, n. 3.
<http://dx.doi.org/10.2202/1559-0410.1257>
- Jeremić, Veljko; Bulajić, Milica; Martić, Milan; Radojičić, Zoran** (2011). "A fresh approach to evaluating the academic ranking of world universities". *Scientometrics*, v. 87, n. 3, pp. 587-596.
<http://dx.doi.org/10.1007/s11192-011-0361-6>
- Jeremić, Veljko; Bulajić, Milica; Martić, Milan; Marković, Aleksandar; Savić, Gordana; Jeremić, Danko; Radojičić, Zoran** (2012). "An evaluation of European countries' health systems through distance based analysis". *Hippokratia*, v. 16, n. 2, pp. 170-174.
<http://goo.gl/CINtkA>
- Jovanović, Milica; Jeremić, Veljko; Savić, Gordana; Bulajić, Milica; Martić, Milan** (2012). "How does the normalization of data affects the ARWU ranking?". *Scientometrics*, v. 93, n. 2, pp. 319-327.
<http://dx.doi.org/10.1007/s11192-012-0674-0>
- Lancho-Barrantes, Bárbara S.; Guerrero-Bote, Vicente P.; De-Moya-Anegón, Félix** (2013). "Citation increments between collaborating countries". *Scientometrics*, v. 94, n. 3, pp. 817-831.
<http://dx.doi.org/10.1007/s11192-012-0797-3>
- López-Illescas, Carmen; De-Moya-Anegón, Félix; Moed, Henk F.** (2011). "A ranking of universities should account for differences in their disciplinary specialization". *Scientometrics*, v. 88, n. 2, pp. 563-574.
<http://dx.doi.org/10.1007/s11192-011-0398-6>
- Leiden. "CWTS Leiden 2013 Rankings list and methodology".
<http://www.leidenranking.com>
- Maletić, Predrag; Kreca, Marija; Jeremić, Veljko; Bulajić, Milica; Djoković, Aleksandar** (2012). "The ranking of municipalities in Serbia through the development level of SME in agribusiness". *International journal of agricultural and statistical sciences*, v. 8, n. 1, pp. 7-13.
- Marginson, Simon** (2011). "Higher education and public good". *Higher education quarterly*, v. 65, n. 4, pp. 411-433.
<http://dx.doi.org/10.1111/j.1468-2273.2011.00496.x>
- Miguel, Sandra; Chinchilla-Rodríguez, Zaida; De-Moya-Anegón, Félix** (2011). "Open access and Scopus: a new approach to scientific visibility from the standpoint of access". *Journal of the American Society for Information Science and Technology*, v. 2, n. 6, pp. 1130-1145.
<http://eprints.rclis.org/16100>
<http://dx.doi.org/10.1002/asi.21532>
- Nishy, P.; Panwar, Yatish; Prasad, Suresh; Mandal, G. K.; Prathap, Gangan** (2012). "An impact-citations-exergy (iCX) trajectory analysis of leading research institutions in India". *Scientometrics*, v. 91, n. 1, pp. 245-251.
<http://dx.doi.org/10.1007/s11192-011-0594-4>
- Paruolo, Paolo; Saisana, Michaela; Saltelli, Andrea** (2013). "Ratings and rankings: voodoo or science?". *Journal of the Royal Statistical Society: Series A (Statistics in society)*, v. 176, n. 3, pp. 609-634.
<http://arxiv.org/pdf/1104.3009.pdf>
<http://dx.doi.org/10.1111/j.1467-985X.2012.01059.x>
- Prathap, Gangan** (2011). "Geographic distribution of Indian academic research". *Current science*, v. 101, n. 6, pp. 715.
<http://www.currentscience.ac.in/Volumes/101/06/0715.pdf>
- Prathap, Gangan** (2012a). "The quality-quantity-quasity and energy-exergy-entropy exegesis of expected value calculation of citation performance". *Scientometrics*, v. 91, n. 1, pp. 269-275.
<http://dx.doi.org/10.1007/s11192-011-0516-5>
- Prathap, Gangan** (2012b). "The research performance of top Indian and Chinese higher education institutions compared". *Current science*, 2012b, v. 102, n. 6, p. 827.
<http://www.currentscience.ac.in/Volumes/102/06/0827.pdf>
- Prathap, Gangan** (2013). "Top Indian higher education institutions and the Leiden SCImago rankings". *Current science*, 2013, v. 104, n. 4, pp. 407-408.
<http://www.currentscience.ac.in/Volumes/104/04/0407.pdf>
- Radojičić, Zoran; Isljamović, Sonja; Petrović, Natasa; Jeremić, Veljko** (2012). "A novel approach to evaluating sustainable development". *Problemy ekorozwoju - Problems of sustainable development*, v. 7, n. 1, pp. 81-85.
- Radojičić, Zoran; Jeremić, Veljko** (2012). "Quantity or qual-

ity: What matters more in ranking higher education institutions?". *Current science*, v. 103, n. 2, pp. 158-162.
<http://www.currentscience.ac.in/Volumes/103/02/0158.pdf>

Romo-Fernández, Luz M.; López-Pujalte, Cristina; Guerrero-Bote, Vicente P.; De-Moya-Anegón, Félix (2011). "Analysis of Europe's scientific production on renewable energies". *Renewable energy*, v. 36, n. 9, pp. 2529-2537.
<http://dx.doi.org/10.1016/j.rser.2012.10.020>

Sadlak, Jan (1978). "Efficiency in higher education – concepts and problems". *Higher education*, v. 7, n. 2, pp. 213-220.
<http://dx.doi.org/10.1007/BF00129419>

Saisana, Michaela; D'Hombres, Beatrice; Saltelli, Andrea (2011). "Rickety numbers: volatility of university rankings and policy implications". *Research policy*, v. 40, n. 1, pp. 165-177.
<http://dx.doi.org/10.1016/j.respol.2010.09.003>

Saisana, Michaela; D'Hombres, Beatrice (2008). *Higher education rankings: robustness issues and critical assessment*. EUR 23487, Joint Research Centre, Publications Office of the European Union, Italy. ISBN: 978 82 79 09704 1
http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/12694/1/eur23487_saisana_dhombres.pdf
<http://dx.doi.org/10.2788/92295>

Salmi, Jamil (2003). "Constructing knowledge societies: new challenges for tertiary education". *Higher education in Europe*, v. 28, n. 1, pp. 65-69.

Seke, Kristina; Petrović, Natasa; Jeremić, Veljko; Vukmirović, Jovanka; Kilibarda, Biljana; Martić, Milan (2013). "Sustainable development and public health: rating European countries". *BMC public health*, v. 13, n. 77.
<http://dx.doi.org/10.1186/1471-2458-13-77>

SCImago Institutions Rankings (SIR) Iber reports.
<http://www.scimagoir.com>

Torres-Salinas, Daniel; Moreno-Torres, José G.; Delgado-López-Cózar, Emilio; Herrera, Francisco (2011). "A methodology for institution-field rankings based on a bidimensional analysis: the IFQ2A index". *Scientometrics*, v. 88, n. 3, pp. 771-786.
<http://dx.doi.org/10.1007/s11192-011-0418-6>

Waltman, Ludo; Calero-Medina, Clara; Kosten, Joost; Noyons, Ed; Tijssen, Robert; Van-Eck, Jan; Van-Leeuwen, Thed; Van-Raan, Anthony; Visser, Martijn; Wouters, Paul (2012). "The Leiden Ranking 2011/2012: data collection, indicators, and interpretation". *Journal of the American Society for Information Science and Technology*, v. 63, n. 12, pp. 2419-2432.
http://sticonference.org/Proceedings/vol2/Waltman_Leiden_791.pdf
<http://dx.doi.org/10.1002/asi.22708>

AGENDA <http://www.elprofesionaldelainformacion.com/agenda.html>

Octubre-noviembre

15-17 de octubre de 2013

CONF INTL SOBRE ACCESO ABIERTO, COMUNICACIÓN CIENTÍFICA Y PRE-SERVACIÓN DIGITAL (BIREDIAL)
VIII SIMPOSIO INTL DE BIBLIOTECAS DIGITALES (SIBD)
Costa Rica
Universidad de Costa Rica
<http://biredial2013.ucr.ac.cr>

16-18 de octubre de 2013

2013 INTL OPEN ACCESS CONF / MEDOANET CONF
Atenas
Greek National Documentation Centre; Medoanet Project
http://openaccess.gr/conferences/conference2013/?language_id=1

16-18 de octubre de 2013

V ENCUENTRO LATINOAMERICANO DE BIBLIOTECARIOS, ARCHIVISTAS Y MUSEÓLOGOS (EBAM). A dimensão social e educativa dos arquivos, bibliotecas e museus na América Latina
São Luís, Maranhão, Brasil
Centro de Ciências Sociais/Universidade Federal do Maranhão
<http://portais.ufma.br>

16-18 de octubre de 2013

ENCUENTRO DE CIENCIAS DE LA INFORMACIÓN (ECIM)
Resistencia (Chaco), Argentina
Universidad Nacional del Nordeste (UNNE). Facultad de Humanidades. Departamento de Ciencias de la Información
<http://www.ecim2013.com.ar>

17-18 de octubre de 2013

II JORNADES VALENCIANES DE DOCUMENTACIÓ
Valencia
Col·legi Oficial de Bibliotecaris i Documentalistes de la Comunitat Valenciana (Cobdcv)
<http://www.cobdcv.es/jornades>
jornades@cobdcv.es

17-18 de octubre de 2013

INTL CONF AND HACKDAYS ON DIGITAL LIBRARIES AND REPOSITORIES
Praga
National Library of Technology
<http://eod2013.techlib.cz>

18-19 de octubre de 2013

7TH INTL CONF EUROPEAN GUIDELINES FOR COOPERATION BETWEEN LIBRARIES, ARCHIVES AND MUSEUMS
Sarajevo
<http://www.bam.ba/index.php/en/konferencija-bam/46-bam-konferencije/153-bam-konferencija-2013>

18-20 de octubre de 2013

6º CONGRESO NACIONAL DE BIBLIOTECAS MÓVILES
Burgos
Roberto Soto, aclebim@yahoo.es
<http://www.bibliobuses.com/laasocongresos6congreso.htm>

21-25 de octubre de 2013

JORNADAS TÉCNICAS RedIRIS
Madrid
Universidad Politécnica de Madrid, Escuela Técnica Superior de Ingenieros Industriales
<http://www.rediris.es/jt/jt2013>

22-25 de octubre de 2013

EUROPEAN CONF ON INFORMATION LITERACY (ECIL)
Estambul
Department of Information Management, Hacettepe University; Department of Information and Communication Sciences, Zagreb University
<http://www.ilconf.org>

24-25 de octubre de 2013

CLASSIFICATION AND VISUALIZATION: Interfaces to knowledge. Intl UDC seminar 2013

- La Haya
National Library of the Netherlands
<http://seminar.udcc.org/2013>
- 25-26 de octubre de 2013**
XVII JORNADAS BIBLIOTECARIAS DE ANDALUCÍA
Jaén
Asociación Andaluza de Bibliotecarios
<http://www.aab.es/jornadas/xvii-jornadas-bibliotecarias-de-andalucia>
- 28 de octubre-1 de noviembre de 2013**
DIGITAL HERITAGE 2013 INTERNATIONAL CONGRESS
Marsella
Unesco, VSMM Society, IEEE, Eurographics, et al.
<http://www.digitalheritage2013.org>
- 4-6 de noviembre de 2013**
VI ENCUENTRO DEL GRUPO REGIONAL IBÉRICO DE EDICIÓN
Oporto
José-Antonio Frías, frias@usal.es
http://www.edicic.org/grupos_regionales.html
- 5 de noviembre de 2013**
IMPLEMENTING OPEN ACCESS POLICY
Londres
<http://www.westminsterforumprojects.co.uk/forums/event.php?eid=641>
- 6-8 de noviembre de 2013**
II JORNADAS SOBRE BIBLIOTECAS DE MUSEOS. Estrategias e innovación
Madrid
Red de Bibliotecas de Museos (Bimus)
<http://bimus.mcu.es>
- 7-8 de noviembre de 2013**
6^{AS} JORNADAS ARCHIVANDO. Valoración documental
León
Fundación Sierra Pambley
<http://jornadasarchivando.sierrapambley.org>
Javier González-Cachafeiro
archivo@sierrapambley.org
- 7-9 de noviembre de 2013**
XI CONGRESO DEL CAPÍTULO ESPAÑOL DE ISKO
I CONGRESO DEL CAPÍTULO PORTUGUÉS DE ISKO
Oporto
José-Antonio Frías, frias@usal.es
<http://www.ugr.es/~isko>
- 7-9 de noviembre de 2013**
COMMUNICATION AND EDUCATION IN KNOWLEDGE SOCIETY
Timisoara, Rumanía
<https://sites.google.com/site/cescuvt/english>
- 13-15 de noviembre de 2013**
JORNADAS SOBRE ACCESO, FORMACIÓN E INVESTIGACIÓN EN INFORMACIÓN Y EDICIÓN CIENTÍFICAS
Buenos Aires
Centro Argentino de Información Científica y Tecnológica (Caicyt-Conicet)
http://www.caicyt.gov.ar/index.php?option=com_content&view=article&id=364
- 14-15 de noviembre de 2013**
TECHNOLOGICAL ECOSYSTEMS FOR ENHANCING MULTICULTURALITY (TEEM). Learning & education
Salamanca
Universidad de Salamanca
<http://teemconference.eu>
- 18-20 de noviembre 2013**
ALIGNING NATIONAL APPROACHES TO DIGITAL PRESERVATION (ANADP 2013). An action assembly
Barcelona
Biblioteca de Catalunya, Institut d'Estudis Catalans
<http://educopia.org/events/ANADP11>
- 18-20 de noviembre de 2013**
V CONGRESO INTL DE CIBERPERIODISMO Y WEB 2.0
Bilbao
Fac de Ciencias Sociales y de la Comunicación, Univ del País Vasco
koldo.meso@ehu.es
<http://ciberpebi.wordpress.com>
- 19-20 de noviembre de 2013**
ONLINE INFORMATION
Londres
<http://www.online-information.co.uk>
- 19-22 de noviembre de 2013**
7TH METADATA AND SEMANTICS RESEARCH CONF (MTRSR 2013)
Tsalónica
Alexander Technological Educational Institute
<http://mtrsr2013.teithe.gr>
- 20-22 de noviembre de 2013**
I CONGRESO INTL INFOXICACIÓN. Mercado de la información y la psique
Sevilla
Facultad de Comunicación, Universidad de Sevilla.
<http://ladecom.es/infoxicacion>
- 21-23 de noviembre de 2013**
LA COLABORACIÓN CIENTÍFICA: una aproximación multidisciplinar
Valencia
Universidad de Valencia
<http://www.colaboracioncientifica.es>
- 25-27 de noviembre de 2013**
SEMANTIC WEB IN LIBRARIES CONF (SWIB13)
Hamburgo
<http://swib.org>
- 27-29 de noviembre de 2013**
IX INTL CATALOGUER MEETING. Catalogação: do real ao virtual
II National Cataloguer Meeting
Rio de Janeiro
Fundação Biblioteca Nacional do Brasil
<http://www.enacat.ufscar.br>
- 27-30 de noviembre de 2013**
VII ENCUENTRO INTL DE INVESTIGADORES Y ESTUDIOSOS DE LA INFORMACIÓN Y LA COMUNICACIÓN (ICOM)
La Habana
Facultad de Comunicación, Universidad de La Habana
icom2013@fcom.uh.cu
- 28-29 de noviembre de 2013**
III JORNADAS DE INTERCAMBIO Y REFLEXIÓN ACERCA DE LA INVESTIGACIÓN EN BIBLIOTECOLOGÍA
La Plata, Argentina
Depto. de Bibliotecología, Fac. de Humanidades y Ciencias de la Educación, Univ. Nacional de la Plata
<http://jornadabibliotecologia.fahce.unlp.edu.ar/jornadas-2013>
- 3-5 de diciembre de 2013**
V CONGRESO INTL LATINA DE COMUNICACIÓN SOCIAL. La sociedad ruido, entre el dato y el grito
La Laguna, Tenerife
http://www.revistalatinacs.org/13SLCS/2013_convocatoria_5_congreso.html
- 4-6 de diciembre de 2013**
IX WORKSHOP INTERNACIONAL DE CIÊNCIA DA INFORMAÇÃO (WICI)
Brasília
Universidade de Brasília
<http://soac.bce.unb.br/index.php/ixwici/2013/index>
- 15-16 de diciembre de 2013**
REDESIGNING WORLDWIDE CONNECTIONS. Theoretical and pragmatical systemic approaches to the complexity management
Tenerife
<http://www.wcsaglobal.org>
- 5-8 de enero de 2014**
9TH INTL CONF ON INFORMATION VISUALIZATION THEORY AND APPLICATIONS (IVAPP 2014)
Lisboa
<http://www.ivapp.visigrapp.org>
- 21-24 de enero de 2014**
IV CONGRESO INTL DE LA ASOCIACIÓN ESPAÑOLA DE INVESTIGACIÓN DE LA COMUNICACIÓN (AE-IC 2014). Espacios de comunicación
Bilbao
Universidad del País Vasco
<http://www.aeic2014bilbao.org>

PUBLICACIONES EL PROFESIONAL DE LA INFORMACIÓN

REVISTA EL PROFESIONAL DE LA INFORMACIÓN

<input type="checkbox"/>	Suscripción anual institucional	202 €
<input type="checkbox"/>	Suscripción individual	94 €
<input type="checkbox"/>	Suscripción sólo online.....	117 €

2013 v. 22

<input type="checkbox"/>	n. 5: Gestión de contenidos	36 €
<input type="checkbox"/>	n. 4: Economía de la información	36 €
<input type="checkbox"/>	n. 3: Bibliotecas y documentación de museos	36 €
<input type="checkbox"/>	n. 2: Educación y biblioteca	36 €
<input type="checkbox"/>	n. 1: Soportes digitales.....	36 €

2012 v. 21

<input type="checkbox"/>	n. 6: Bibliotecas universitarias	36 €
<input type="checkbox"/>	n. 5: Información y derecho	36 €
<input type="checkbox"/>	n. 4: Comunicación digital	36 €
<input type="checkbox"/>	n. 3: Organización del conocimiento	36 €
<input type="checkbox"/>	n. 2: Publicación científica y acceso abierto	36 €
<input type="checkbox"/>	n. 1: El futuro de la web	36 €

2011 v. 20

<input type="checkbox"/>	n. 6: Fuentes de información 2.0	25 €
<input type="checkbox"/>	n. 5: Inteligencia competitiva	25 €
<input type="checkbox"/>	n. 4: Fotografía y bancos de imágenes	25 €
<input type="checkbox"/>	n. 3: Información de las administraciones públicas	25 €
<input type="checkbox"/>	n. 2: Archivos administrativos e intranets	25 €
<input type="checkbox"/>	n. 1: Psicología y sociología de la información.....	25 €

2010 v. 19

<input type="checkbox"/>	n. 6: Medios de comunicación en internet	20 €
<input type="checkbox"/>	n. 5: Cooperación de bibliotecas en red	20 €
<input type="checkbox"/>	n. 4: Arquitectura de la información	20 €
<input type="checkbox"/>	n. 3: Información biomédica	20 €
<input type="checkbox"/>	n. 2: Doc. en publicidad y comunic. empresarial	20 €
<input type="checkbox"/>	n. 1: Sector editorial	20 €

Formulario de compra EPI

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

ANUARIO THINKEPI

Tarifas institucionales

<input type="checkbox"/>	Anuario 2013	97 €
<input type="checkbox"/>	Anuario 2012	50 €
<input type="checkbox"/>	Anuario 2011	40 €
<input type="checkbox"/>	Anuario 2010	35 €
<input type="checkbox"/>	Anuario 2009	30 €
<input type="checkbox"/>	Anuario 2008	30 €
<input type="checkbox"/>	Anuario 2007	30 €
<input type="checkbox"/>	Anuario 2013+2012	139 €
<input type="checkbox"/>	Anuario 2013+2012+2011	165 €
<input type="checkbox"/>	Anuario 2013 a 2007	260 €

Tarifas individuales (particulares)

<input type="checkbox"/>	Anuario 2013	43 €
<input type="checkbox"/>	Anuario 2012	30 €
<input type="checkbox"/>	Anuario 2011	23 €
<input type="checkbox"/>	Anuario 2010	21 €
<input type="checkbox"/>	Anuario 2009	21 €
<input type="checkbox"/>	Anuario 2008	21 €
<input type="checkbox"/>	Anuario 2007	21 €
<input type="checkbox"/>	Anuario 2013+2012	60 €
<input type="checkbox"/>	Anuario 2013+2012+2011	76 €
<input type="checkbox"/>	Anuario 2013 a 2007	130 €

Formulario de compra Anuario ThinkEPI

<http://www.thinkepi.net/pedido>

LIBROS EL PROFESIONAL DE LA INFORMACIÓN (Editorial UOC)

<input type="checkbox"/>	1. Tecnologías de la web semántica	12,00 €
	Juan-Antonio Pastor	
<input type="checkbox"/>	2. La revolución del libro electrónico	11,50 €
	José-Antonio Cordón	
<input type="checkbox"/>	3. Sistemas de información en la empresa	11,50 €
	Josep Cobarsi-Morales	
<input type="checkbox"/>	4. Información en el móvil	12,00 €
	Natalia Arroyo-Vázquez	
<input type="checkbox"/>	5. Acceso abierto a la ciencia	12,00 €
	Ernest Abadal	
<input type="checkbox"/>	6. Fuentes de información médica	12,00 €
	Pablo Medina-Aguerreberre	
<input type="checkbox"/>	7. Gestión de la reputación online	11,50 €
	Javier Leiva-Aguilera	
<input type="checkbox"/>	8. Wikipedia de la A a la W	12,50 €
	Tomás Saorín	
<input type="checkbox"/>	9. Etiquetar en la web social	12,00 €
	Raquel Gómez-Díaz	
<input type="checkbox"/>	10. Mejorar las búsquedas de información	12,00 €
	Silvia Argudo y Amadeu Pons	
<input type="checkbox"/>	11. Clubes de lectura	11,50 €
	Óscar Carreño	
<input type="checkbox"/>	12. Plan social media y community manager	13,50 €
	Julián Marquina-Arenas	
<input type="checkbox"/>	13. Documentación audiovisual en televisión	12,00 €
	Jorge Caldera y Pilar Arranz	
<input type="checkbox"/>	14. Gestión de documentos en la e-administración	12,00 €
	Elisa García-Morales	
<input type="checkbox"/>	15. El film researcher	12,00 €
	Iris López-de-Solis	
<input type="checkbox"/>	16. Preservación digital	12,00 €
	Miquel Térmens	
<input type="checkbox"/>	17. Gestión de contenidos	13,00 €
	Ricardo Eito-Brun	
<input type="checkbox"/>	18. Documentación fotográfica	14,00 €
	Juan-Miguel Sánchez-Vigil y Antonia Salvador-Benitez	
<input type="checkbox"/>	19. Documentación cinematográfica	13,00 €
	Elena De la Cuadra	
<input type="checkbox"/>	20. Archivos	14,00 €
	Ramón Alberch-Fugueras	

Formulario de compra libros EPI-UOC

<http://www.elprofesionaldelainformacion.com/form.php>



En todos los casos, excepto en los Anuarios, hay que añadir los gastos de envío

Información y pedidos: Isabel Olea / epi.iolea@gmail.com / Tel.: +34-608 491 521

<http://www.elprofesionaldelainformacion.com>
<http://www.thinkepi.net>



Te entendemos”

Sistemas de gestión de bibliotecas Open Source

Interfaces interactivas y OPACs

Repositorios OAI

Gestión documental y de archivos

Digitalización

Outsourcing de servicios documentales

Desarrollo de sitios web / multimedia / e-learning

Comunicación y e-marketing de servicios de información



Oficinas Centrales:
C/Garcilaso 15-B
46003 Valencia

Tel.: 96 369 41 23
Fax: 96 369 34 39
info@masmedios.com



Deseo recibir todos los números de la revista *EPI* a partir del mes de enero del año 20 Las suscripciones van por años naturales, de enero a diciembre

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 Código de seguridad CVC2

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 CAIXESBBXXX**

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

PayPal a la cuenta EPISCP@gmail.com

Enviar el boletín relleno por correo electrónico, o postal al APARTADO 32.280 - 08080 BARCELONA - ESPAÑA.

Consultas: suscripciones@elprofesionaldelainformacion.com o +34 609 352 954

Open choice. Los autores pueden liberar su artículo en open access en las webs de *EPI* mediante el pago de 300 € (400 € si el artículo es en inglés, porque además se publica online una versión en español)

Precios 2013

Suscripción normal: 194,23 € + IVA = 202 €

Suscripción personal: (sólo a domicilios particulares) 90,38 € + IVA = 94 €

Gastos de envío fuera de España: Europa: 40 € Américas: 60 €

Suscripción sólo online: 96,69 € + IVA = 117 €

Número suelto: 34,61 € + IVA = 36 €

Gastos de envío fuera de España: Europa: 9 € Américas: 15 €

Fuera de Europa no se cobra el IVA



NORMAS PARA LOS AUTORES

El profesional de la información tiene dos secciones principales:

ARTÍCULOS: Trabajos de investigación y temas analizados en profundidad.

ANÁLISIS: Experiencias, estudios de casos, análisis de productos, reseñas, etc.

Las contribuciones han de ser originales e inéditas, no pueden haberse publicado previamente en soporte papel o electrónico. El tamaño ideal es de 3.500 palabras, aunque en algunos casos la Redacción puede autorizar una mayor extensión.

El texto ha de enviarse en Word, rtf u odt. Las tablas deberán ir pegadas en el mismo documento. Todos los materiales gráficos (diagramas, fotografías, capturas de pantalla, etc.) deben pegarse en el Word y además enviarse en ficheros independientes (en formatos xls, jpg, pdf, etc.). Las imágenes jpg deben tener una resolución de al menos 300 pp (unos 200 KB cada una).

El texto debe presentarse completamente plano, sin autoformatos ni automatismos de Word (subsecciones, viñetas, citas enlazadas, pies de página, sangrías, tabulaciones, colores, etc.), pero debe seguir el estilo de *EPI* en cuanto a **negritas** (nombres de los autores citados), *cursivas* (instituciones, títulos de revista, marcas) y mayúsculas. Los urls deben estar sin hipervínculo.

Las citas bibliográficas en el texto se realizarán de la forma: (**Apellido**, año) o (**ApellidoAutor1**; **ApellidoAutor2**, año).

La redacción debe ser concisa y precisa, evitando la retórica.

Los trabajos deben incluir: a) título, b) resumen de 100-150 palabras, c) 5-10 palabras clave, d) title, e) abstract de 100-150 palabras, f) 5-10 keywords.

Aparte se incluirá el nombre de los autores, su lugar de trabajo y dirección (postal y electrónica), su foto tipo carnet superior a 100 KB en jpg, y un currículum de unas 70 palabras.

EVALUACIÓN

Los trabajos son revisados según el sistema tradicional "peer review" en doble ciego por al menos dos expertos en el tema, del Consejo Asesor de la revista y/o externos. La revista se compromete a informar del resultado a los autores.

REFERENCIAS BIBLIOGRÁFICAS

Ordenadas alfabéticamente por autor, se limitarán a las obras citadas en el artículo. No se acepta bibliografía de relleno.

Artículos de una publicación periódica:

Apellido, Nombre; Apellido2, Nombre2 (año). "Título del artículo". *Título de la publicación periódica*, mes, v., n., pp. xx-yy. *Dirección url iniciada en nueva línea sin barra y sin punto finales* <http://dx.doi.org/10.xxxx/doi>

Ponencia presentada en un congreso:

Apellido, Nombre; Apellido2, Nombre2 (año). "Título de ponencia". En: *Nombre del congreso*, pp. xx-yy. *Dirección url iniciada en nueva línea sin barra y sin punto finales* <http://dx.doi.org/10.xxxx/doi>

Monografías e informes:

Apellido, Nombre; Apellido2, Nombre2 (año). *Título del trabajo*. Lugar de publicación: editor. ISBN: espacios, sin guiones *Dirección url iniciada en nueva línea sin barra y sin punto finales* <http://dx.doi.org/10.xxxx/doi>

Capítulo de una monografía:

Apellido, Nombre; Apellido2, Nombre2 (año). "Título del capítulo". En: *Apellido, Nombre; Apellido2, Nombre2*. *Título de la monografía*. Lugar de publicación: editor, pp. xx-yy. ISBN: espacios, sin guiones *Dirección url iniciada en nueva línea sin barra y sin punto finales*

Recurso en línea:

Apellido, Nombre; Apellido2, Nombre2. *Título del recurso*. *Dirección url iniciada en nueva línea sin barra y sin punto finales*

Todas las contribuciones se tienen que enviar a la sección *EPI* de la plataforma OJS del Repositorio Español de Ciencia y Tecnología (Recyt) de la Fecyt: <http://recyt.fecyt.es/index.php/EPI/index>

Previamente los autores deben registrarse en: <http://recyt.fecyt.es/index.php/EPI/user/registerJournal>