



Observatorio

**Knowledge organization:
some trends in an emergent domain**

Richard P. Smiraglia

Artículos

Tesaurus: estándares y recomendaciones

S. Sánchez-Cuadrado, M. J. Colmenero y J. A. Moreiro

Linked data for open vocabularies and HIVE's global framework

Eva Méndez y Jane Greenberg

Aplicación de SKOS para la interoperabilidad de vocabularios en el entorno de LOD

José-Antonio Pastor-Sánchez, F. Javier Martínez-Méndez y José-Vicente Rodríguez-Muñoz

Evaluación de sitios web multilingües: metodología

Mar Andreu-Vall y Mari-Carmen Marcos

Consumo de información periodística online en Colombia

Liliana Gutiérrez, Javier-Andrés Gómez, Andrea Salgado, Juliana Estrada y María-José Ramírez-Méndez

Carencias en los sistemas de gestión del conocimiento

Laura P. Pinto, Luis E. Becerra y Luis C. Gómez-Flórez

Promoción de la lectura en las bibliotecas públicas de Cataluña

Mònica Baró, Teresa Mañà, Maite Barrios y Júlía Baena

Análisis

Clasificación de crowdsourcing basada en tareas

E. Estellés-Arolas y F. González-Ladrón-De-Guevara

LD y LOD: su implantación en Europeana

Ana Ríos-Hilario, Diego Martín-Campo y Tránsito Ferreras

Medición de la influencia digital: Klout y Peerindex

Javier Serrano-Puche

Software documental

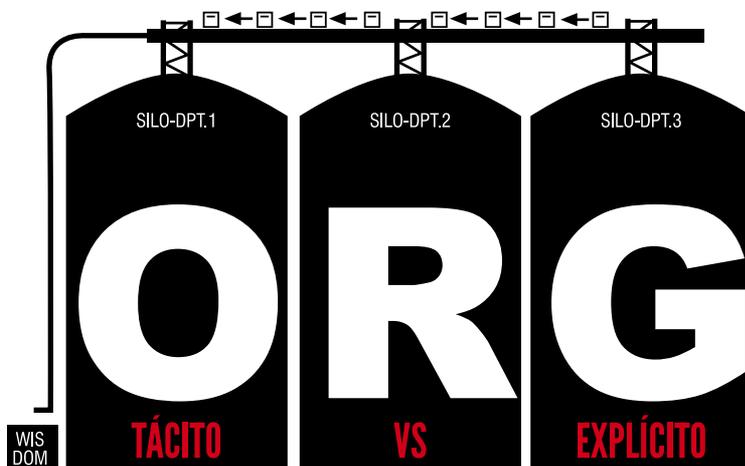
TemaTres: software para gestionar tesauros

Audilio Gonzales, María Ramírez y Diego Ferreyra



EL PROFESIONAL DE LA INFORMACIÓN

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



ORGANIZACIÓN DEL CONOCIMIENTO

07:11 AM -FROM SILOS TO WISDOM



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

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

Sanvergrafic

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://www.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

Lluís Codina

Universitat Pompeu Fabra

<http://www.lluiscodina.com>

Ricardo Eito-Brun

Grupo GMV

reito@gmv.es

Elea Giménez-Toledo

Inst. de Estud. Document. sobre Ciencia y Tecnología

elea.gimenez@cchs.csic.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

José-Antonio Millán

Libros y bitios

<http://jamillan.com>

Fernanda Peset

Universidad Politécnica de Valencia

mpesetm@upv.es

Jorge Serrano-Cobos

MASmedios

jorgeserrano@gmail.com

Daniel Torres-Salinas

Universidad de Navarra

torressalinas@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.

Ramon Alberch

Generalitat de Catalunya, Barcelona.

Adela d'Alòs-Moner

Doc6, Barcelona.

Ricardo Baeza-Yates

Univ. de Chile, Santiago, Chile. Yahoo! Research, Barcelona.

Carlos B. Amat

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

Jesús Bustamante

Biblioteca, Cedefop, Salónica, Grecia.

Carlota Bustelo-Ruesta

Consultora, Madrid.

Emilio Delgado-López-Cózar

Universidad de Granada, Granada.

Javier Díaz-Noci

Universitat Pompeu Fabra, Barcelona.

Assumpció Estivill

Universitat de Barcelona, Barcelona.

Antonia Ferrer-Sapena

Univ. Politècnica de Valencia, Valencia.

António Fidalgo

Universidade da Beira Interior, Portugal.

Francisco-Javier García-Marco

Universidad de Zaragoza, Zaragoza.

Paola Gargiulo

Caspar, Roma, Italia.

Johannes Keizer

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

Thomas Krichel

Palmer School of Libr. & Inform. Sci. LIU, NY, USA.

Victoria Manglano

Ovid Technologies, Madrid.

Mari-Carmen Marcos

Universitat Pompeu Fabra, Barcelona.

Pere Masip

Blanquerna, Univ. Ramon Llull, Barcelona.

Charles McCathieNeville

Opera Software, Oslo, Norway.

Marcos Palacios

Universidade Federal da Bahia, Brasil.

Joan Roca

Minnesota State University, Mankato, USA.

Ramón Salaverría

Universidad de Navarra, Pamplona.

Robert Seal

Loyola Univ. Chicago, Evanston, Illinois, USA.

Ernesto Spinak

Consultor, Montevideo, Uruguay.

Jesús Tramullas

Universidad de Zaragoza, Zaragoza.

BASES DE DATOS

Academic search premier (Ebsco)

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

Compludoc (Universidad Complutense de Madrid)

<http://europa.sim.ucm.es/compludoc/>

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?article23>

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

http://wokinfo.com/products_tools/multidisciplinary/webofscience/ssci/
Impact Factor 2010 = 0,375

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 2010 = 0,030

<http://www.scopus.com>

CATÁLOGOS Y SERVICIOS BIBLIOTECARIOS

Argos-Bolsum (GVA, Generalitat Valenciana)

http://www.pre.gva.es/argos/es/contenido_general/recursos/bolsum/

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://ec3.ugr.es/in-recs/iii/Documentacion-fecha-2009.htm>

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://alj.c.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 FACEBOOK

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

EPI EN TWITTER

http://twitter.com/revista_EPI

PLATAFORMA DE PRODUCCIÓN OJS

Recyt, Repositorio español de ciencia y tecnología (Fecyt)

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

Sumario EPI, 2012, v. 21, n. 3

Tema central: Organización del conocimiento

OBSERVATORIO

- 225 **Knowledge organization: some trends in an emergent domain**
Richard P. Smiraglia

ARTÍCULOS

- 229 **Tesaurus: estándares y recomendaciones**
Sonia Sánchez-Cuadrado, María-Jesús Colmenero-Ruiz y José-Antonio Moreiro
- 236 **Linked data for open vocabularies and HIVE's global framework**
Eva Méndez and Jane Greenberg
- 245 **Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data**
Juan-Antonio Pastor-Sánchez, Francisco-Javier Martínez-Méndez y José-Vicente Rodríguez-Muñoz
- 254 **Evaluación de sitios web multilingües: metodología y herramienta heurística**
Mar Andreu-Vall y Mari-Carmen Marcos
- 261 **Criterios y hábitos de consumo de información periodística online en Colombia**
Liliana Gutiérrez-Coba, Javier-Andrés Gómez-Díaz, Andrea Salgado-Cardona, Juliana Estrada-Gutiérrez y María-José Ramírez-Méndez
- 268 **Carencias en los sistemas de gestión del conocimiento: una revisión bibliográfica**
Laura-Patricia Pinto-Prieto, Luis-Eduardo Becerra-Ardila y Luis-Carlos Gómez-Flórez
- 277 **Promoción de la lectura en las bibliotecas públicas de Cataluña: evaluación y resultados**
Mònica Baró, Teresa Mañà, Maite Barrios y Júlia Baena

ANÁLISIS

- 283 **Clasificación de iniciativas de crowdsourcing basada en tareas**
Enrique Estellés-Arolas y Fernando González-Ladrón-De-Guevara
- 292 **Linked data y linked open data: su implantación en una biblioteca digital. El caso de Europea**
Ana Ríos-Hilario, Diego Martín-Campo y Tránsito Ferreras-Fernández
- 298 **Herramientas web para la medición de la influencia digital: análisis de Klout y PeerIndex**
Javier Serrano-Puche
- 304 **Geobibline y su comparación con otras bases de datos geográficas**
Eva Novotná
- 312 **Mapa de procesos de un sistema de gestión de accesibilidad en un servicio web de la administración pública: el Ayuntamiento de Zaragoza**
José-Antonio Salvador-Oliván y María-Jesús Fernández-Ruiz

SOFTWARE DOCUMENTAL

- 319 **TemaTres: software para gestionar tesauros**
Audilio Gonzales-Aguilar, María Ramírez-Posada y Diego Ferreyra
- 326 **Foursquare: tu biblioteca en el momento y en el lugar adecuados**
Natalia Arroyo-Vázquez

NOVEDADES EDITORIALES

INFORMACIÓN PARA LOS AUTORES

Las mejores herramientas para la investigación en Humanidades.

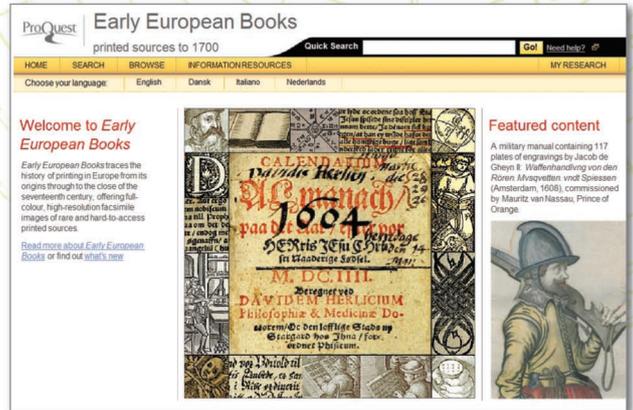
Publicaciones periódicas históricas.

Colecciones para la investigación en Humanidades.

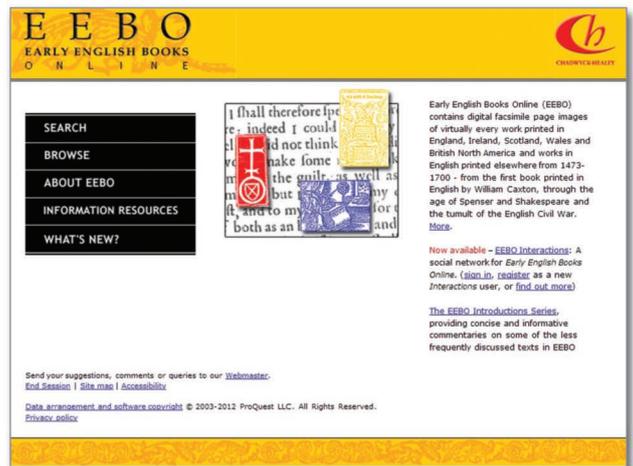
Los primeros libros impresos.



ProQuest está orgulloso de desempeñar un papel en la preservación de los primeros libros impresos y archivos de publicaciones periódicas, así como de servir de ayuda a las bibliotecas para que puedan ofrecerlos a sus usuarios en el área de Humanidades. A través de nuestras herramientas de investigación de renombre internacional como *Early English Books Online*, *Early European Books* y *Periodicals Archive Online*, los investigadores pueden acceder de forma sencilla a miles de libros y artículos en línea de vital importancia para la investigación en el campo de las Humanidades.



Early European Books home page



Early English Books Online home page



Periodicals Archive Online home page

www.proquest.com

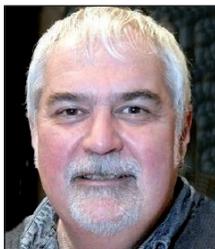
ProQuest®
Start here.



KNOWLEDGE ORGANIZATION: SOME TRENDS IN AN EMERGENT DOMAIN



Richard P. Smiraglia



Richard P. Smiraglia is visiting professor at the *School of Information Studies, University of Wisconsin-Milwaukee*. He is associate researcher at the *e-Humanities Group, Royal Netherlands Academy of the Arts and Sciences*, Amsterdam, Netherlands. He is editor-in-chief of the journal *Knowledge organization*, published by *Ergon-Verlag* of Würzburg.

*University of Wisconsin Milwaukee
School of Information Studies
Northwest Quadrant Building B, Room 3550
2025 E Newport, Milwaukee, WI 53211, USA*

Abstract

The organization knowledge concept is defined as the domain where scientific research interacts with its application to systems development. Disciplines that embraces, as information science; and its products, such as classification systems, are cited. Some recent trends and current activities are presented. The article concludes presenting briefly the ISKO society and its activities.

Keywords

Knowledge organization, Trends, Systems, Classifications, *UDC*, *ISKO*.

Título: Organización del conocimiento: algunas de las tendencias en un dominio emergente

Resumen

Se define el concepto organización del conocimiento como el dominio donde interacciona la investigación científica con su aplicación al desarrollo de sistemas. Se citan las disciplinas que acoge, como la ciencia de la información; y sus productos, como los sistemas de clasificación. Se presentan tendencias recientes y trabajos actuales. Se cita sociedad *ISKO* y se presentan brevemente sus actividades.

Palabras clave

Organización del conocimiento, Tendencias, Sistemas, Clasificaciones, *CDU*, *ISKO*.

Smiraglia, Richard P. "Knowledge organization: some trends in an emergent domain". *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 225-227.

<http://dx.doi.org/10.3145/epi.2012.may.01>

What is knowledge organization?

Knowledge organization (also well-known by its acronym KO) is the domain in which the order of knowledge is both the primary paradigm for scientific investigation and the primary application in the development of systems. These two threads are interwoven in the domain, and there is a fairly substantive discourse between the evolving theories of knowledge organization, on the one hand, and evolving systems for organizing knowledge (known as KOS) on the other. The applied products of the domain are KOSs such as

classifications, taxonomies, ontologies and thesauri, for example. The theoretical products are the rules for discovering the natural order of knowledge or for imposing a useful sequence on discovered knowledge. Both the science and its applications rely heavily on concept theory (**Dahlberg**, 2006; **Hjørland**, 2009), which some argue is the most basic or primal element in the knowledge universe (**Van den Heuvel**; **Smiraglia**, 2010; **Szostak**, 2011). Although the domain is closely associated with information science, and many of its practitioners are members of faculties of information science, the extension of the domain of KO is actually

Nota: Este artículo puede leerse traducido al español en:
<http://www.elprofesionaldelainformacion.com/contenidos/2012/mayo/richard-smiraglia.html>

somewhat broader, encompassing all disciplines in which the tools of KO are used. That is, KO is actually the domain that incorporates interdisciplinary approaches to the order of knowledge (Hjørland 2003, 2008).

The locus for much work in KO is the *International Society for Knowledge Organization (ISKO)* and its regional chapters. ISKO was founded by Dahlberg in 1986 (Dahlberg, 2006) to promote and coordinate research. The domain is fairly compact but highly active; its major venues are its journal *Knowledge organization*, the proceedings of its biennial international conferences in the series *Advances in knowledge organization* (both published by Ergon-Verlag of Würzburg, Germany), and the proceedings of the individual regional chapters.

Arguably, the primary approach to research in KO is called domain analysis, which is itself a tool-kit incorporating eleven methodological approaches articulated by Hjørland and Albrechtsen (1995). Most domain analyses are empirical, and many are bibliometric, but some other methodological approaches also are occasionally employed, including ethnography (Hartel, 2003).

Many applied products in KO are classifications, ranging from the classical bibliographical meta-classifications such as the *Universal decimal classification* to simple experimental taxonomies, sometimes referred to as naïve classifications (Beghtol, 2003). The means by which concepts are isolated, relationships and attributes recorded, and classifications are structured are articulated by what often is called “classification theory” (Beghtol, 2010). There has been a shift in the KO domain over the past quarter-century from the search for universally acceptable solutions, to domain-centric approaches. This shift has been the catalyst for an increasing emphasis on domain analysis, as well as for the emergence of methods such as cognitive work analysis (Mai, 2008, 2011).

Emergent trends in KO

I have used the tools of domain analysis to track the shifting intension and extension of KO since I became editor of *Knowledge organization*. These parameters tell us at any

given moment both the breadth of topics being treated in research in our domain, and the theoretical depth of the paradigms in operation. In a recent paper (Smiraglia, 2011), I have brought together several domain analytical snapshots of KO for meta-analysis. What we see is internal coherence in the domain around the poles of KO represented by concept theory, on the one hand, and KOS on the other. There has been a shift from the search for universal solutions that occupied early KO researchers, to a search for interoperability since the advent of the World Wide Web. It also is apparent that a dynamic epistemological tension exists in KO between empirical researchers working with traditional “scientific” methods, and theorists working with humanistic methods. This dichotomic epistemic stance helps provide dual dimensionality to the domain, keeping it always in a state of renewal as new topics enter the paradigmatic region or receive treatment from emerging methodological poles.

Emergent trends are often first identified by the regional chapters of ISKO. It has been the editorial policy of *Knowledge organization* in recent years to carry the top 3-5 papers from every regional chapter conference whenever possible. Topics that have emerged as critical for the future of KO are ontogeny, linked open data, people-centered properties, global agents, multimedia information retrieval, and especially faceted solutions.

Recent work

Finally, I would like to point to the three remarkable developments in the KO domain:

1) The first is represented by the reinvigoration of the *Universal decimal classification*, and in particular the proceedings of its 2011 *International UDC seminar* [Slavic; Civalero (eds.), 2011]:

<http://seminar.udcc.org/2011/index.htm>

All of the dimensions noted in the preceding section of this paper were apparent at the seminar, as semantic web technologies revealed their impact on theoretical and empirical work in KO. Web vocabularies, knowledge representation, elementary knowledge structures, interoperability, linked

data, ontologies, facets, and integration of new data models all made an appearance at this seminar.

2) My own *IOrg* (for *Information Organization*) research group based at the *University of Wisconsin-Milwaukee* has just sent to press an anthology of papers about epistemology and its critical role in KO. The book will be titled *Cultural frames of knowledge* (Smiraglia; Lee, 2012) and will be available at the 12th *Intl ISKO conf* in Mysore, India, in August 2012. The

ISKO International Society for Knowledge Organization	
home	This is the official website for ISKO – the premier international scholarly society devoted to the theory and practice of knowledge organization .
about ISKO	
join ISKO	
Knowledge Organization journal	
ISKO events	
ISKO chapters	
ISKO people	
ISKO publications	
KO literature	
	
	Recently published <ul style="list-style-type: none"> Knowledge organization. 39. Issue 3 mailed to ISKO members Knowledge organization literature. 2012. Issue 2 Ontology for big systems (Ontology summit) Epistemic foundation of bibliographic classification in early China (JDoc) Is classification necessary after Google? (JDoc) Conceptual syntagmatic associations in user tagging (JASIST) ADMS: Asset Description Metadata Schema Ontology-Lexica Community Group (W3C) draft of ISO 25964-2: Interoperability with other vocabularies Foundations of information science

<http://www.isko.org>

12th International ISKO Conference

Categories, Relations and Contexts in Knowledge Organization

Mysore, 6-9 August 2012



Home | Call for Papers | Guidelines for Authors | Important Dates | Scientific Committee | Registration | Tentative Programme | Accomodation | Contact

<http://www.isibang.ac.in/~isko>

volume contains eight literature reviews plus a foreword by **Hope Olson**, covering topics as diverse as discourse analysis, domain analysis, semiotics and genre, and focusing on cultural diversity through lenses of Chinese and Indian cultures, as well as a survey of feminist epistemologies.

3) Finally, the Mysore conference itself promises to push the boundaries of KO in new and exciting ways. Most likely, the conference will embrace the multi-dimensionality of the four poles represented by the empiricist-humanist concatenation and the concept theory-KOS continuum, while simultaneously bringing forth new emergent trends. In addition, the conference's location in India for the first time in twenty years also represents the first foray beyond North America-Western Europe. The cultural influences at the conference likely will also stretch the domain's intension. KO continues, therefore, to emerge and evolve as a dynamic area for research and discovery.

References

Beghtol, Clare. "Classification for information retrieval and classification for knowledge discovery: Relationships between 'professional' and 'naïve' classifications". *Knowledge organization* 2003, n. 30, pp. 64-73.

Beghtol, Clare. "Classification theory". *Encyclopedia of library and information sciences*, 3rd ed., 2010, v. 1, n. 1, pp. 1045-1060

Dahlberg, Ingetraut. "Knowledge organization: a new science?". *Knowledge organization*, 2006, n. 33, pp. 11-19.

Hartel, Jenna. 2003. "The serious leisure frontier in library and information science: hobby domains". *Knowledge organization*, 2003, n. 30, pp. 228-238.

Hjørland, Birger. "Some fundamentals of knowledge organization". *Knowledge organization*, 2003, n. 30, pp. 87-111.

Hjørland, Birger. "What is knowledge organization (KO)?". *Knowledge organization*, 2008, n. 35, pp. 86-101.

Hjørland, Birger. "Concept theory". *Journal of the American Society for Information Science and Technology*, 2009, n. 60, pp. 1519-36.

Hjørland, Birger; Albrechtsen, Hanne. "Toward a new horizon in information science: domain-analysis". *Journal of the American Society for Information Science*, 1995, n. 46, pp. 400-425.

Mai, Jens-Erik. "Design and construction of controlled vocabularies: analysis of actors, domain, and constraints". *Knowledge organization*, v. 35, n. 1, pp. 16-29.

Mai, Jens-Erik. "The modernity of classification". *Journal of documentation*, 2011, n. 67, pp. 710-730.

Slavic, Aida; Edgardo Civallero (eds.). *Procs of the UDC seminar classification and ontology, formal approaches and access to knowledge*, The Hague, 19-20 September 2011.

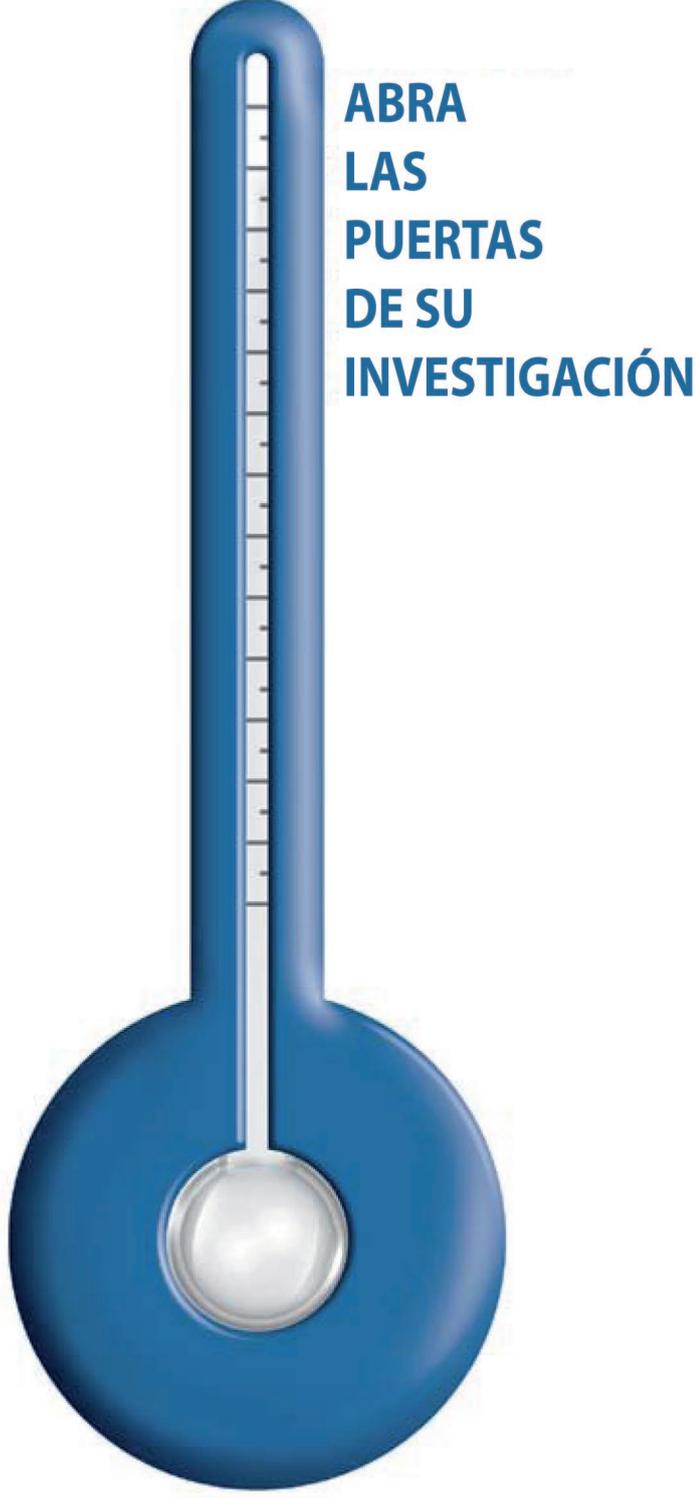
Smiraglia, Richard P. "Domain coherence within knowledge organization: people, interacting theoretically, across geopolitical and cultural boundaries". In: McKenzie, Pam; Johnson, Catherine; Stevenson, Sarah eds., *Exploring interactions of people, places and information: Procs of the 39th annual CAIS/ACSI conf*, University of New Brunswick, Fredericton, N.B. Canada, 2-4 June 2011.

http://www.cais-acsi.ca/proceedings/2011/73_Smiraglia.pdf

Smiraglia, Richard P.; Hur-Li, Lee (eds.) *Cultural frames of knowledge*. Würzburg: Ergon-Verlag, forthcoming, 2012.

Szostak, Rick. "Complex concepts into basic concepts". *Journal of the American Society for Information Science and Technology*, 2011, n. 62, pp. 2247-65.

Van den Heuvel, Charles; Smiraglia, Richard P. "Concepts as particles: metaphors for the universe of knowledge", In: Gnoli, C.; Mazzocchi, F. (eds.), *Paradigms and conceptual systems in knowledge organization: Procs of the 11th intl ISKO conf*, 23-26 Febr. 2010, Rome, Italy, Ergon-Verlag, Würzburg, pp. 50-56.



La llave para maximizar su potencial de investigación

SciVal es una pionera suite de herramientas de búsqueda desarrollada por Elsevier que le permite explotar todo su potencial de investigación. A través de sus extensas bases de datos, las soluciones de SciVal le proporcionan la visión estratégica que necesita para permanecer a la vanguardia de la investigación.

SciVal le ayuda a tomar decisiones seguras de inversión en materia de investigación en su lucha por destacar en un mercado tan competitivo, permitiéndole para que pueda analizar minuciosamente su rendimiento en investigación interdisciplinar, evaluar sus estrategias de investigación y hallar su talento natural.

Descubra las nuevas oportunidades que SciVal puede ofrecerle en www.scival.com/unlock
Se mire como se mire, tiene sentido.



TESAUROS: ESTÁNDARES Y RECOMENDACIONES



Sonia Sánchez-Cuadrado, María-Jesús Colmenero-Ruiz y José-Antonio Moreiro



Sonia Sánchez-Cuadrado es doctora en documentación por la *Universidad Carlos III de Madrid*. Desde 2002 es profesora en el *Departamento de Informática* de dicha *Universidad* con docencia en documentación e informática sobre temas relacionados con el modelado de tesauros y elaboración de sistemas de organización del conocimiento y recuperación de información.

Universidad Carlos III de Madrid
Avda. Universidad, 30. 28911 Leganés, Madrid
ssanche@ie.inf.uc3m.es



María-Jesús Colmenero-Ruiz es licenciada en ciencias biológicas por la *Universidad de Alcalá de Henares*. Diplomada, licenciada y doctora en documentación por la *Universidad Carlos III de Madrid*. Trabaja como profesora ayudante en el *Departamento de Biblioteconomía y Documentación*, y forma parte del grupo de investigación *Acrópolis*. Su investigación se centra en tesauros y vocabularios controlados, web semántica y *topic maps*.

Universidad Carlos III de Madrid
C/ Madrid, 126. 28903 Getafe, Madrid
mcolmene@bib.uc3m.es



José-Antonio Moreiro es catedrático de biblioteconomía y documentación de la *Universidad Carlos III de Madrid*. Es autor de trabajos sobre técnicas de análisis de contenido documental y cuestiones conceptuales de la documentación. Ha sido director del *Departamento de Biblioteconomía y Documentación* y decano de la *Facultad de Humanidades, Comunicación y Documentación* de la *Universidad Carlos III de Madrid*.

Universidad Carlos III de Madrid
C/ Madrid, 126. 28903 Getafe, Madrid
jamore@bib.uc3m.es

Resumen

Los tesauros se han adaptado al entorno tecnológico y constituyen una pieza esencial como medio de organización del conocimiento para la web semántica. Tras un largo período sin abordar la actualización de estándares y recomendaciones de los tesauros, en los últimos cinco años se han realizado revisiones en el seno de los organismos normalizadores. Un trabajo de consenso internacional ha culminado con la publicación de la primera parte de la norma *ISO* para tesauros. Aporta un modelo de datos y pone el acento en la interoperabilidad. Se recoge lo más destacado de las diferentes directrices para tesauros que se han sucedido desde los años ochenta hasta el estándar *ISO 25964*. Se reflexiona sobre cuáles han sido las características más destacadas de esta evolución y se exploran los usos más novedosos de los sistemas de organización del conocimiento en la Web.

Palabras clave

Tesauros, Estándares, Directrices, Recomendaciones, Pautas, Vocabularios controlados, Sistemas de organización del conocimiento.

Title: Thesauri: Standards and guidelines

Abstract

Thesauri and other knowledge organization systems have been adapted to the technological environment. Currently, this represents an essential aspect of knowledge organization on the semantic web. After a long period in which thesaurus guidelines and standards were not updated, some have been revised by the governing organisations within the past five years. Consensus among international experts has resulted in publication of the first part of an *ISO* standard for thesauri, which provides a data model and emphasizes the interoperability process. This paper compiles the most relevant guidelines for thesauri published between the '80s and the *ISO 25964* document. We reflect on the most relevant characteristics in this evolution and explore the most recent applications of knowledge organization systems on the Web.

Artículo recibido el 10-02-12
Aceptación definitiva: 16-03-12

Keywords

Thesauri, Standards, Guidelines, Recommendations, Controlled vocabulary, Knowledge organization systems.

Sánchez-Cuadrado, Sonia; Colmenero-Ruiz, María-Jesús; Moreiro, José-Antonio. "Tesauros: estándares y recomendaciones". *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 229-235.
<http://dx.doi.org/10.3145/epi.2012.may.02>

Introducción

En este estudio se presenta una revisión de las recomendaciones y estándares sobre sistemas de organización de conocimiento, también conocidos como *KOS* por la denominación inglesa *knowledge organization systems*, con atención preferente a los tesauros. Los *KOS* son esquemas para organizar información y gestionar conocimiento. Dentro del concepto se incluyen diversos tipos de listas de términos (diccionarios, glosarios, listas de autoridades y de nombres geográficos o *gazetters*), taxonomías, encabezamientos de materia, tesauros, redes semánticas y ontologías (Hodge, 2000; Zeng; Chan, 2004). La diferencia entre unos y otros reside en el tipo de estructuración y la información que pueden representar (Sánchez-Cuadrado *et al.*, 2007). Las tecnologías de la información, y en concreto internet, han motivado la aparición de diferentes iniciativas para adecuarlas a este nuevo medio digital (Pérez-Agüera, 2004). Inicialmente, las tendencias se orientaban a representar *KOS* sólo para indexar y recuperar información y después han evolucionado hacia la idea de la web semántica para compartir conocimiento.

Normas y recomendaciones: evolución

La evolución de las recomendaciones de los tesauros y los sistemas de organización del conocimiento en general gira en torno a algunas nociones fundamentales como: vocabulario controlado, término, concepto, descriptores, relaciones, interoperabilidad, *mapping*, etc. Cuando en las normas se habla de "términos" se refiere a palabras o frases cortas que representan un concepto unívoco, de tal manera que si se trata de una palabra homónima, entonces se deben utilizar mecanismos de desambiguación, es decir, que precisen el significado del término. Un "descriptor" designa un término aprobado para ser usado en indización. Asumiendo fundamentos comunes con la terminología, los conceptos de un tesauro son términos especializados de un campo determinado en los que se especifican las relaciones conceptuales y se asegura que un término representa inequívocamente un concepto (Amtz; Picht, 1995). Los conceptos se precisan como una unidad de pensamiento que puede ser definida o descrita. Sin duda la idea más novedosa desde que surgieron los tesauros es la de interoperabili-

dad tomada del campo de las TIC: Se denomina interoperabilidad a la propiedad de un sistema cuyas interfaces pueden trabajar con otros sistemas sin restricciones de acceso o implementación.

Aunque de forma reiterativa se plantea la cuestión de si los tesauros siguen teniendo vigencia (Rowley, 1994; Sveinonius, 1986; Sánchez-Cuadrado *et al.*, 2007), lo cierto es que continúan apareciendo ejemplos e iniciativas a favor de los vocabularios controlados. Recientemente *Google*, *Microsoft* y *Yahoo!* se unieron para ofrecer una colección de vocabularios compartidos: *Schema.org*. Los webmasters pueden usar el vocabulario definido en este esquema para definir y marcar sus páginas. El principal beneficio para los tres buscadores es compartir los mismos mecanismos de entendimiento.

<http://schema.org>

Lancaster (2002) planteó la evolución de los vocabularios controlados desde sus inicios hasta mediados de los ochenta, mostrando la influencia de unas recomendaciones en otras. Dextre-Clarke (2008) revisó los últimos 50 años de la organización del conocimiento desde su perspectiva personal. La llegada de los ordenadores personales, y más tarde la necesidad de interconectar diferentes vocabularios controlados marcó un punto de inflexión importante. Otros estudios comparan enfoques entre diferentes propuestas como el estándar *ISO 2788* y la recomendación *SKOS (W3C, 2008)*. Se puede consultar una visión de algunas normas y los principales protocolos o estándar de intercambio de datos en Dextre-Clarke (2010).

La figura 1 muestra la evolución de las normas y recomendaciones sobre tesauros. Se ha generalizado a las normas o recomendaciones de los sistemas de organización del cono-

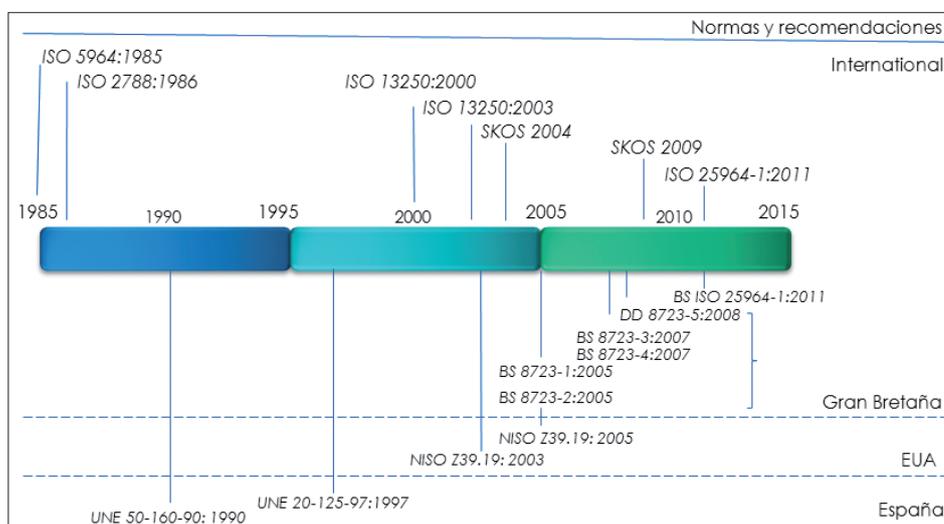


Figura 1. Evolución de normas y recomendaciones sobre tesauros y sistemas de organización de conocimiento

cimiento de instituciones de estándares de algunos países como Gran Bretaña o Estados Unidos por la innegable influencia y peso que han tenido en la última norma internacional publicada. El punto de inicio se ha marcado con la primera norma, que hasta la aparición de esta última estaba vigente y como punto de continuación para el esquema desarrollado por **Lancaster** (2002). En la parte superior se presentan las normas aparecidas a nivel internacional, mientras que en la parte inferior se han recogido algunas normas representativas aparecidas en los últimos años a nivel local.

ISO 5964:1985

Denominada *Guidelines for the establishment and development of multilingual thesauri*, la norma *ISO 5964* de 1985 insta los principales procedimientos para el tratamiento de varias lenguas en un mismo recurso de conocimiento, identificando los principales problemas en cuanto a la equivalencia semántica. Esta norma fue adoptada por el equivalente estándar nacional en diferentes países. En España se tradujo en la *UNE 50-160-90:1990*.

Sus directrices definen una lengua fuente y una serie de lenguas objetivo. Los conceptos de la lengua fuente se traducen a la lengua objetivo mediante reglas que dependen de la existencia de equivalencias exactas, inexactas, parciales, compuestas o no equivalencia.

ISO 2788:1986

La norma *ISO 2788* para tesauros monolingües (*Guidelines for the establishment and development of monolingual thesauri*), inicialmente publicada en 1974, fue revisada y publicada en 1986. Esta norma a nivel internacional se ha mantenido durante 25 años, hasta la aparición de la *ISO 25964-1* en 2011. Cada organización nacional publicó su versión traducida y adaptada a cada país. Estas normas locales han tenido posteriores revisiones al año 1986 como la británica y la americana.

ISO 13250:2000/2003

El estándar *ISO 13250* define los *topic maps (TM)*. *Topic maps* es un metamodelo para categorizar, fusionar y navegar mediante relaciones semánticas entre documentos y secciones de los mismos a partir de su contenido. Están especificados en lenguaje xml, con una sintaxis que se denomina xtm (**Pepper; Moore**, 2001). En el estándar xtm se declaran los elementos y los conceptos claves (*topic, association, occurrence, published subject* y *scope*). Un *TM* tiene como objetivo la indización de contenido en conjuntos de documentos, y permite realizar indizaciones múltiples para la misma colección de documentos web. El estándar *TM* también se suele incluir como un formato de intercambio electrónico de tesauros, aunque no está específicamente definido para ellos, siendo sólo una de sus aplicaciones.

SKOS 2004/2009

Simple knowledge organization systems (SKOS) es una recomendación del *W3C* que tiene su origen en la iniciativa *SWAD-Europe*. Propone un modelo de datos para definir e implementar estructuras básicas de contenido como organización del conocimiento semi-formal que permita un entorno compartido basado en un modelo simplificado y que haga posible su reutilización e interoperabilidad. La primera

publicación tuvo lugar en 2004 siendo actualizada y publicada en 2009. El lenguaje de formalización es rdf. Su potencia reside en que puede ser combinado y extendido con lenguajes formales de representación del conocimiento como *owl* y que admite la incorporación de otros vocabularios de metadatos (*Dublin core, FOAF*). Se dirige a conceptos que pueden ser representados por un uri (*universal resource identifier*), permitiendo representar *KOS* multilingües mediante la combinación de propiedades. La especialización del modelo *SKOS* con la extensión *XL (SKOS-XL)* permite el uso de propiedades de rdf (*rdfs:subPropertyOf* y *rdfs:subClassOf*) para crear subtipos de relaciones interpretables por un motor de inferencia adecuado para la web semántica. *SKOS* incluye la especificación básica del modelo de representación (*SKOS core*), recursos para el alineamiento e interoperabilidad de los conceptos (*SKOS mapping*), mecanismos de ampliación de propiedades de representación de conocimiento (*SKOS extensions*) y servicios para la consulta web de *KOS (SKOS API)*.

En la literatura se refleja una amplia aceptación para formalizar los vocabularios controlados con *SKOS (Agrovoc, Eurovoc, LCSH)*. Como modelo ligero de datos, *SKOS* ha demostrado su utilidad para los procesos de interoperabilidad a nivel terminológico (**Sah et al.**, 2007), y para el conocimiento colaborativo en la generación social de esquemas de clasificación (**Sah; Wade**, 2010).

ISO 25964-1:2011

La norma *Thesauri and interoperability with other vocabularies* está siendo actualizada. La primera parte denominada *Thesauri for information retrieval (ISO 25964-1:2011)* fue publicada en 2011. La segunda parte está prevista para 2012 y está orientada a guiar la interoperabilidad con otros vocabularios. Se basa en el estándar *BS8723 Structured vocabularies for information retrieval* que a su vez se inspiró en la propuesta *ANSI*. La nueva actualización de la norma *ISO 25964-1:2011* a diferencia de la británica está constituida en Ods partes.

SKOS propone un modelo de datos para definir e implementar estructuras básicas de contenido que permita un entorno compartido y que haga posible su reutilización e interoperabilidad

Es probablemente la renovación más profunda que ha experimentado el estándar *ISO* de los tesauros, ya demandada por los profesionales desde tiempo atrás. Fusiona los estándares anteriores para tesauros monolingües y multilingües. También propone un formato de intercambio basado en un esquema xml. Su objetivo es proporcionar recomendaciones para el desarrollo y mantenimiento de los tesauros. Se observa un aumento en el uso de facetas para la organización del esquema de conocimiento. Está dirigida a diferentes tipos de recursos de información, con independencia del medio utilizado (texto, multimedia, bases de conocimiento, etc.), que permitan su uso en aplicaciones de recuperación

de información. En esta versión *ISO* se incorporan técnicas de procesamiento textual que flexibilicen el tratamiento del vocabulario controlado y las técnicas de recuperación de información que ya estaban siendo aplicadas en software de tesauros y motores de búsqueda (Marrero *et al.*, 2010). Por ejemplo se tiene en cuenta la casuística de combinaciones de palabras posibles y en formatos diferentes teniendo en cuenta que una computadora debe recibir instrucciones para identificar variantes ortográficas de un mismo término.

ANSI/NISO Z39.19:2005

El *American National Standards Institute* realizó una profunda revisión del estándar de los tesauros bajo la denominación *Guidelines for the construction, format and management of monolingual controlled vocabularies*. Proponía una guía para la construcción de tesauros, incidiendo en los aspectos de formato y gestión de vocabularios monolingües. Las recomendaciones de la norma *NISO Z39.19:2005* incluían criterios para el mantenimiento de los sistemas de organización del conocimiento mediante gestores de tesauros automatizados. Esta norma estaba orientada a cubrir la representación de contenido de diferentes *KOS* como las listas de sinónimos, taxonomías, tesauros y otros tipos de vocabularios controlados.

La propuesta de esta *NISO (National Information Standards Organization)* entiende que los vocabularios controlados suelen ser utilizados para describir el contenido por asignación de términos para representar metadatos asociados al contenido de los objetos (*NISO Z39.19:2005:12*). El estándar *NISO Z39.19:2005* está pensado para ser aplicado a tesauros monolingües (*NISO Z39.19:2005*, apartado 2.4). Es decir, a diferencia de los estándares *ISO* aportados hasta ese momento, esta iniciativa incorporaba ya su aplicación en entornos informatizados, como demuestra el tratamiento de esquemas de metadatos como *Dublin Core*.

IFLA 2009

La *IFLA* publicó su informe sobre las directrices para tesauros multilingües en 2009 con el objetivo de completar la norma *ISO 5964:1985*. Complementaba también a otras directrices dirigidas a los encabezamientos de materias y al estándar *NISO Z39.19:2005* (*IFLA*, 2009). La principal aportación de este informe es sobre el tratamiento de construir tesauros de forma no-simétrica y la vinculación entre diferentes vocabularios controlados. Por tesauros no-simétricos entienden un tipo de tesauros multilingüe donde el número de descriptores de cada lengua no es necesariamente el mismo y tampoco el modo en que los descriptores se relacionan en las diferentes lenguas.

BS8723 (2005-2008) y BS ISO 25964-1:2011

El estándar *BS8723 (Structured vocabularies for information retrieval)* se publicó durante los años 2005-2008. Está constituido por cinco partes dónde la última, conocida como *DD8723-5*, está dedicada a protocolos y formatos de intercambios para la interoperabilidad. Está dirigido a vocabularios estructurados en general, caracterizando y comparando los diferentes tipos de *KOS* como esquemas de clasificación, taxonomías, esquemas de encabezamiento de materias, te-

sauros, listas de autoridades y ontologías. El software pasa a tener una función esencial en: la gestión del vocabulario, los recursos de formalización necesarios para los mapeos entre vocabularios, y los protocolos y los formatos de intercambios. El enfoque de organización de conocimiento se basa en la representación de conceptos.

Las distintas partes de esta norma británica han quedado anuladas tras la publicación de la *BS ISO 25964:2011* con el título *Information and documentation. Thesauri and interoperability with other vocabularies. Thesauri for information retrieval*, idéntica a la *ISO* del 2011 y que supone la traslación local de la norma internacional, con su plena adhesión.

Implicaciones de la evolución de la norma de los tesauros

Las revisiones más influyentes en la última *ISO* aprobada han sido la americana *NISO Z3919:2005* y la británica *BS5723-1-5:2005-2008*. Un aspecto a destacar es la modificación del nombre de la norma. La *NISO* generalizó la norma para los “vocabularios controlados”. En el caso del estándar británico se establecen diferencias entre tesauros y lo que no son tesauros como vocabularios controlados para la recuperación de información. Sin embargo, la *ISO* siempre ha mantenido en el título y en el objetivo a los tesauros.

“ Una de las diferencias más significativas la encontramos en el paso de estructuración terminológica a una dimensión de representación conceptual dónde las equivalencias se crean entre conceptos y no entre palabras ”

La norma *ISO 25964:2011* surge como revisión de las normas *ISO 2788* y *5964*, que después de 25 años ya no responden a las prácticas actuales de procesos de indización y recuperación de información, ni cubren la demanda de aplicaciones de red que quieren utilizar recursos heterogéneos. Por tanto, por una parte estamos hablando de un cambio por obsolescencia del formato (del papel a soporte electrónico), y por otro de un cambio o incremento de funcionalidad. Pero también hablamos de una adaptación a nuevos perfiles de usuarios. La unión de recursos distintos implica el desarrollo de mecanismos de *mapping* o alineamiento que permitan la interoperabilidad. Una de las diferencias más significativas la encontramos en el paso de estructuración terminológica a una dimensión de representación conceptual dónde las equivalencias se crean entre conceptos y no entre palabras. Si nos remitimos a las definiciones puede no tener sentido si consideramos que los términos representan conceptos. Los conceptos a su vez se definen con palabras y un término es una palabra de un lenguaje de especialidad. La explicación reside en si la finalidad de definir y representar se hace al modo de las bases de datos léxicas o las ontologías lingüísticas como *WordNet*, o bien desde un punto de vista formal más próximo al modelo matemático. Este hecho repercutirá en la definición de restricciones en los tipos de relaciones

como ocurre en otros esquemas de organización de conocimiento como las ontologías.

Los modelos de datos del estándar *ISO 25964-1:2011* y la recomendación de *SKOS* (2009) del *W3C* están basados en la representación de conceptos. A nivel conceptual debería ser fácil la conversión entre ellos, puesto que tratan de representar lo mismo. Ambos grupos han mostrado interés en sus trabajos recíprocos por la sensibilización existente en la profesión frente a las barreras que puedan crearse con el uso de distintos estándares, en lugar de producirse una suma de esfuerzos. Los trabajos del estándar *ISO* justifican su iniciativa con la inclusión de muchos países y la consideración de iniciativas paralelas. Según la *ISO* los formatos para la interoperabilidad existentes (*MARC*, *SKOS*, *Zthes*, esquema XML de la *DD8723-5* británica y la *ISO 16642*, el *Terminological markup framework (TMF)*) no son adecuados para expresar todos los aspectos de un tesauro según se definen en el estándar (*ISO 25964-1:2011*). **Dextre-Clarke** (2010) asegura que la comunidad *SKOS* está considerando hacer extensiones a la propia recomendación para mejorar el alineamiento con el estándar *ISO*.

Dextre-Clarke (2009) señala diferencias entre los dos modelos de intercambio. En primer lugar, *SKOS* no establece directrices para la construcción de un sistema de organización de conocimiento, solamente proporciona un lenguaje de formalización para su publicación en la Web. En segundo lugar, defiende que el modelo de datos de la norma *ISO* incluye todas las características de un tesauro, mientras que *SKOS* pretende servir a varios tipos de vocabularios diferentes, por lo que no prevé mecanismos para algunas de las características del tesauro. Sin embargo, una de las limitaciones que presenta la norma *ISO 25964-1:2011* es que no es aplicable a bases de datos o software que pueda ser utilizado directamente en las aplicaciones de búsqueda o indexación, limitándose a sugerir recomendaciones para la gestión de diccionarios de sinónimos.

La nueva norma de los tesauros incide en la importancia del *mapping* entre recursos asumiendo que es un proceso complejo. Una de las dificultades del proceso reside en la rareza de encontrar auténtica equivalencia exacta. A esto debemos sumar que muchos de los vocabularios están pobremente contruidos con objetivos distintos, con diferentes niveles de granularidad en la jerarquía, etc. Sin duda estas peculiaridades dificultan el mapeo e intensifican su labor, viéndose agudizado por los escasos recursos que se dedican normalmente. El metatesauro de medicina *UMLS (Unified medical language system)*, de la *US National Library of Medicine*, que integra 100 tesauros de medicina y cinco millones de conceptos, es uno de esos ejemplos donde se ha invertido gran esfuerzo para mapear estructuras de conocimiento diferentes en el entorno digital.

<http://www.nlm.nih.gov/research/umls>

Sistemas de organización de conocimiento en la Web

El *mapping* persigue entendimiento común, y cuando hablamos de intercambio de datos se convierte en interoperabilidad. El *mapping* de recursos diversos implica enlazar

datos o información. Una forma de crear tesauros también es posible mediante el enlace de conceptos de diferentes vocabularios controlados. Este método se caracterizaría por una gestión colaborativa del tesauro y la posibilidad de crear de forma flexible tesauros descentralizados. Esto significa que el sistema de organización del conocimiento puede ser creado y mantenido de forma externa, pero conectar con conceptos equivalentes. Esta idea está relacionada con los principios de *linked data*.

El grupo de *linked data W3C Library Linked Data Incubator Group* trabaja enlazando datos en rdf. Utiliza URIs para identificar los conceptos y ofrece información sobre los recursos usando rdf. Esto le permite incluir enlaces a otras URIs. De este modo consigue interconectar datos diferentes y a su vez reutilizar información en lugar de duplicarla.

Los aspectos básicos del diseño son los siguientes:

- que el dato esté en la Web;
- que sea un dato interpretable por las máquinas;
- que no esté en formato propietario;
- que se use el estándar rdf;
- que sean enlazados mediante rdf.

Estos principios apuntan directamente a la reutilización de conocimiento. El proyecto *DBpedia* es un ejemplo de este planteamiento. Trabaja para extraer datos de la *Wikipedia* y ponerlos a disposición web de forma estructurada, mostrando los resultados de búsqueda con *Zitgist*. Otros ejemplos de resultados de *linked data* significativos son *Geonames* y *The Data Hub*.

<http://zitgist.com>

El modo de unir diferentes vocabularios puede hacerse con mapeos uno a uno de los esquemas, pero también mediante una ontología de alto nivel o una ontología de referencia. Un ejemplo es el proyecto *Umbel (Upper mapping and binding exchange layer)* diseñado para ayudar a los contenidos a interactuar. Se establecen vínculos entre la ontología conceptual de referencia y los vocabularios controlados o de dominio y conjuntos de datos. Incluye vocabularios de metadatos y ontologías de amplia difusión como *SKOS*, *WordNet*, *RSS*, *FOAF*, *Sioc*, *PIM*, *GeoNames*, y *OpenCyc*, entre otras.

<http://rdfs.org/sioc/spec>

Ontologías y vocabularios de metadatos se vinculan mediante diferentes etiquetas para unir conceptos idénticos, ayudando así a la fusión y coexistencia de diferentes conjuntos de etiquetas. Es decir, el objetivo no es transformar los recursos, sino crear un mapa de su significado como en *TagCommons* o *Folcsonomías*. Otro proyecto relacionado es *Semse*, un sistema que permite la representación y recuperación conceptual de esquemas de metadatos mediante su mapeo contra una ontología de alto nivel (**Palcios**, 2010).

Los tesauros también han experimentado un fuerte impulso con el desarrollo de las ontologías. La ingeniería ontológica y las iniciativas para elaborar la web semántica se han centrado en los mismos fundamentos de estructuración y organización del conocimiento, pero con la aplicación de lenguajes de formalización como rdf y owl (**Daconta et al.**, 2003, p. 157).

2005 fue un año clave para un planteamiento interoperable. La *OMG (Organization Management Group)* publicó *Ontology definition metamodel*, que integra seis metamodelos entre los que se encuentra *UML*. Los tesauros, entendidos como un tipo de ontología ligera (**Daconta et al.**, 2003) o sistemas de representación del conocimiento (**Sanchez-Cuadrado et al.**, 2007), pueden ser definidos según el metamodelo de la *OMG*. *UML* no se creó pensando expresamente en los tesauros, pero pueden ser modelados (**Génova et al.**, 2011) como podemos observar en los modelos de tesauros publicados por las normas de 2011.

Conclusiones

El análisis de las normas de tesauros y *KOS* en general permite determinar que son un recurso caracterizado por su contenido, su estructura y su función. En su proceso evolutivo estos aspectos han cambiado. La nueva concepción de sistema de organización de conocimiento deja a un lado las diferenciaciones entre mono y multilingüismo. Requieren la representación formal del contenido, lo que implica un lenguaje de formalización más lógico. Este cambio de representación condiciona sin duda la estructura relacional. La función de los *KOS* se ha ampliado en la última década con las TIC, convirtiéndose la interoperabilidad, la vinculación entre *KOS* y la reutilización en factores esenciales.

Las normas y recomendaciones de la *ISO* se han centrado más en tesauros que en otros tipos de *KOS*, mientras que *SKOS*, planteando una formalización simple ha facilitado el uso y la eficiencia de los vocabularios controlados en la web semántica. Desde la eclosión de la Web, los tesauros han generado un vínculo inseparable con la informática, aunque esto no ha sido recogido por las recomendaciones hasta 2004 por el *W3C* y hasta 2005 por los organismos de estandarización. Este reconocimiento surgió de su necesidad de adaptación a los entornos tecnológicos actuales y pone de manifiesto la importancia y vigencia de los tesauros, así como otros *KOS* para la gestión del conocimiento. Su adaptación técnica supone un gran avance para estos sistemas de organización del conocimiento de gran tradición en nuestro campo.

Bibliografía

Aenor. *UNE 50106:1990. Directrices para el establecimiento y desarrollo de tesauros monolingües*. Madrid: Aenor, 1990.

Aenor. *UNE 50125:1997. Directrices para el establecimiento y desarrollo de tesauros multilingües*. Madrid: Aenor, 1997.

Amtz, Reiner; Picht, Heribert. *Introducción a la terminología*. Madrid: *Fundación Germán Sánchez Ruipérez*, 1995. ISBN: 84 8616 895 3

ANSI/NISO. *Z39.19: 2005 Guidelines for the construction, format and management of monolingual controlled vocabularies*. NISO Press: Bethesda, MD, 2005. ISBN: 1 880124 65 3

http://www.niso.org/apps/group_public/project/details.php?project_id=46

ISO. *BS 8723-1:2005 Structured vocabularies for information retrieval. Guide. Definitions, symbols and abbrevia-*

tions. British Standards Institution: London, 2005. ISBN: 0 580 46798 8

ISO. *BS 8723-2:2005 Structured vocabularies for information retrieval. Guide. Thesauri*. British Standards Institution: London, 2005, 60. ISBN: 0 580 46799 6

ISO. *BS 8723-3:2007 Structured vocabularies for information retrieval. Guide. Vocabularies other than thesauri*. British Standards Institution: London, 2007, 52. ISBN: 978 0 580 63072 9

ISO. *BS 8723-4:2007 Structured Vocabularies for information retrieval. Guide. Interoperability between vocabularies*. British Standards Institution: London, 2007, 62. ISBN: 978 0 580 63073 6

Daconta, Michael C.; Obrst, Leo J.; Smith, Kevin T. *The semantic web. A guide to the future of XML, web services, and knowledge management*. Indianapolis: Wiley, 2003. ISBN: 978 0471432579

DD *8723-5:2008 Structured vocabularies for information retrieval. Guide. Formats and protocols for data exchange*. British Standards Institution: London, 2008, 60. ISBN: 978 0 580 53864 3

Dextre-Clarke, Stella G. "International standard ISO 25964". En: *The 9th European networked knowledge organization systems (NKOS). Workshop at the 13 ECDL Conf, Corfu, Greece, Oct. 2009*.

Dextre-Clarke, Stella G. "Knowledge organization systems standards". *Encyclopedia of library and information sciences*, Third Edition. Edition by Taylor & Francis, 2010, pp. 3164-3175. ISBN: 0 8493 9711 1

Dextre-Clarke, Stella G. "The last 50 years of knowledge organization: a journey through my personal archives". *Journal of information science*, 2008, v. 34, n. 4, pp. 427-437. <http://dx.doi.org/10.1177/0165551508089225>

Dextre-Clarke, Stella G. "Thesaurus standards on a converging track". *Legal information management*, 2010, v. 10, n. 1, pp. 43-45. <http://dx.doi.org/10.1017/S1472669610000265>

Génova-Fuster, Gonzalo; Llorens-Morillo, Juan; Fuentes-Torres, José-Miguel; Morato-Lara, Jorge; Martínez-Fernández, Paloma. "Las jerarquías conceptuales en UML comparando la norma ISO 2788 con el metamodelo de UML". *Técnica administrativa*, 2011, v. 10, n. 45. <http://www.cyta.com.ar/ta1001/v10n1a1.htm>

Hodge, Gail. *Systems of knowledge organization for digital libraries. Beyond traditional authority files*. Washington, DC: The Council on Library and Information Resources, 2000. <http://www.clir.org/pubs/reports/pub91/contents.html>

IFLA. *Guidelines for multilingual thesauri*. (IFLA professional reports: 115). IFLA, 2009. ISBN: 978 90 77897 35 5 <http://www.ifap.ru/library/book411.pdf>

ISO. *ISO 25964-1:2011. Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval, 2011*.

ISO. *ISO 2788:1986. Guidelines for the establishment and*

development of monolingual thesauri. 2nd ed. Geneva: International Organization for Standardization, 1986.

ISO. *ISO 5964:1985. Guidelines for the establishment and development of multilingual thesauri*. Geneva: International Organization for Standardization, 1985.

ISO. *ISO/IEC 13250:2003. Information technology. SGML Applications. Topic maps*. Geneva: International Organization for Standardization, 2003.

Lancaster, Frederick W. *El control de vocabulario en la recuperación de la información*. Valencia: Universitat de València, 2002. ISBN: 978 84 3705 444 5

Marrero, Mónica; Sánchez-Cuadrado, Sonia; Urbano, Julián; Morato, Jorge; Moreira, José-Antonio. "Sistemas de recuperación de información adaptados al dominio biomédico". *El profesional de la información*, 2010, mayo-junio, v. 19, n. 3, pp. 246-254.
<http://dx.doi.org/10.3145/epi.2010.may.04>

Palacios-Madrid, Vicente. *Sistema de recuperación conceptual mediante niveles semánticos en la representación de esquemas de metadatos*. Tesis doctoral. Universidad Carlos III de Madrid, 2010.
<http://e-archivo.uc3m.es/handle/10016/9332>

Pepper, Steve; Moore, Graham. "XML topic maps (XTM) 1.0. Specification". *TopicMaps.Org*, 2001.
<http://www.topicmaps.org/xtm/1.0>

Pérez-Agüera, José-Ramón. "Automatización de tesauros y su utilización en la web semántica". *Textos universitaris de biblioteconomia i documentació*, 2004, dic., n. 13.
<http://www.ub.edu/bid/13perez2.htm>

Rowley, Jennifer. "The controlled versus natural languages debate revisited: a perspective on information retrieval practice and research". *Journal of information science*, 1994, April, v. 20, n. 2, pp. 108-118.

<http://dx.doi.org/10.1177/016555159402000204>

Sah, Melike; Hall, Wendy; Gibbins, Nicholas M.; De Roure, David C. "Sempport: a personalized semantic portal". En: *Procs of the 18th conf on hypertext and hypermedia*. ACM, New York, 2007, pp. 31-32.
<http://eprints.soton.ac.uk/264441>

Sah, Melike; Wade, Vicent. "Automatic metadata extraction from multilingual enterprise content". En: *Procs of the 19th ACM intl conf on information and knowledge management*. New York: ACM, 2010, p. 1665-1668.
<http://www.cngl.ie/drupal/sites/default/files/papers2/Automatic%20Metadata%20Extraction.pdf>

Sánchez-Cuadrado, Sonia; Morato-Lara, Jorge; Palacios-Madrid, Vicente; Llorens-Morillo, Juan; Moreira, José-Antonio. "De repente, ¿todos hablamos de ontologías?". *El profesional de la información*, 2007, nov.-dic., v. 16, n. 6, pp. 562-568.
<http://www.elprofesionaldelainformacion.com/contenidos/2007/noviembre/03.pdf>
<http://dx.doi.org/10.3145/epi.2007.nov.03>

Svenonius, Elaine. "Unanswered questions in the design of controlled vocabularies". *Journal of the American Society for Information Science*, 1986, Sept., v. 37, n. 5, pp. 331-340.
http://polaris.gseis.ucla.edu/gleazer/462_readings/Svenonius_1986.pdf

W3C. *SKOS Simple Knowledge Organization Systems Primer*. W3C Working Group, 2008.
<http://www.w3.org/TR/2008/WD-skos-primer-20080221>

Zeng, Marcia L.; Chan, Lois M. "Trends and issues in establishing interoperability among knowledge organization systems". *Journal of the American Society for Information Science and Technology*, 2004, v. 55, n. 5, pp. 377-395.
<http://dx.doi.org/10.1002/asi.10387>

Próximos temas centrales

Julio 2012	Comunicación digital
Septiembre 2012	Información y derecho
Noviembre 2012	Bibliotecas académicas
Enero 2013	Soportes digitales
Marzo 2013	Educación y biblioteca
Mayo 2013	Bibliotecas y documentación de museos
Julio 2013	Economía de la información
Septiembre 2013	Gestión de contenidos

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>



LINKED DATA FOR OPEN VOCABULARIES AND *HIVE'S* GLOBAL FRAMEWORK



Eva Méndez and Jane Greenberg



Eva Méndez is an associate professor at the *Library and Information Science Department* in the *University Carlos III of Madrid*, where she directs the *Master in Digital Libraries* program. She holds a PhD in information science and for the last fifteen years has been teaching and researching mainly in the following areas: metadata, semantic web, digital libraries, open access, information policies and social web. She is a member of the *DCMI Advisory Board* and the co-chair of the *DCMI Social Tagging Community*, and independent expert for the *European Commission* in the areas of digital libraries and open access to research data.

*Universidad Carlos III de Madrid
Facultad de Humanidades, Comunicación y Documentación
C/ Madrid, 128 (Dcho. 14.2.17) – 28903 Getafe (Madrid), España
emendez@bib.uc3m.es*



Jane Greenberg is currently chair of excellence in the *University Carlos III of Madrid* at the *Library and Information Science Department*. Dr. Greenberg is professor in the *School of Information and Library Science (SILS)* at the *University of North Carolina* at Chapel Hill, and director of the *SILS Metadata Research Center*. She earned her PhD from the *School of Information Sciences* at the *University of Pittsburgh* and received a MSc with a specialization in bibliographic control from the *School of Library Service* at *Columbia University*. She is the recipient of the *Margaret Mann Citation* for outstanding professional achievement in cataloging or classification.

*School of Information and Library Science
University of North Carolina at Chapel Hill
205 Manning Hall, CB # 3360 – Chapel Hill, NC 27599-3360
janeg@email.unc.edu*

Abstract

This paper summarizes new trends and advances in Knowledge Organization from the perspective of linked open data (LOD). Although this is particularly important to galleries, libraries, archives and museums, the so-called GLAM community, it is of more general relevance, and part of the value of LOD lies in its adoption beyond that community. LOD includes descriptive metadata and vocabulary encoding schemes that are being “skosified” (encoded in the SKOS format) or rendered in OWL (the web ontology language) and made available not only “on” the web, but “for” the semantic web. The paper highlights a few exemplary initiatives in the field. The paper also introduces the *HIVE (Helping Interdisciplinary Vocabularies Engineering)* framework and discusses the *HIVE-ES* (España) extension for Spanish language vocabularies, leading to a more global approach for linked open vocabularies (LOV).

Keywords

Linked open vocabularies (LOV), Linked data (LD), Linked open data (LOD), Knowledge organization systems (KOS), Vocabularies, Metadata, SKOS, Semantic web, *HIVE*; *HIVE-ES*.

Título: Datos enlazados para vocabularios abiertos: marco global de *HIVE*

Resumen

Se presentan brevemente las nuevas tendencias y avances en la organización del conocimiento desde la perspectiva de *linked open data* (LOD). Aunque esto es particularmente importante para galerías, bibliotecas, archivos y museos –la llamada comunidad GLAM–, es de importancia más general, y parte del valor de LOD se encuentra en su adopción más allá de esa comunidad. LOD incluye esquemas de metadatos descriptivos y de codificación de vocabularios que están siendo “skosificados” (codificados en el formato SKOS) o transformados en OWL (el lenguaje de ontologías web) y puestos a disposición no sólo “en” la Web, sino también “para” la web semántica. Se destacan algunas iniciativas ejemplares en este campo y se presenta el marco *HIVE (Ayuda a la Ingeniería de Vocabularios Interdisciplinarios)* y se analiza la extensión *HIVE-ES* (España) para los vocabularios en español, dando lugar a un enfoque más global a los vocabularios abiertos enlazados (LOV).

Nota: Este artículo puede leerse traducido al español en:
<http://www.elprofesionaldelainformacion.com/contenidos/2012/mayo/mendez-greenberg.html>

Artículo recibido el 19-05-12

Palabras clave

Vocabularios abiertos enlazados, Linked open vocabularies (LOV), Datos enlazados (LD), Linked open data (LOD), Sistemas de organización del conocimiento, Vocabularios, Metadatos, SKOS, Web semántica, HIVE, HIVE-ES.

Méndez, Eva; Greenberg, Jane. "Linked data for open vocabularies and HIVE's global framework". *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 236-244.

<http://dx.doi.org/10.3145/epi.2012.may.03>

1. Introduction

Over the last few years, increased attention has been directed toward linked data, linked open data, and sharing vocabularies in an open environment. Linked open data, or just LOD, has become a buzzword appearing in nearly every initiative focusing on digital information organization. LOD is an approach to fostering and advancing the semantic web and the web of data -an idea that initially gained prominence in the late 90's through the work of **Tim Berners-Lee**, the inventor of the World Wide Web. In many respects, linked data has evolved as the *de facto* means for the publication of structured data on the Web, involving a broad range of different data communities. The use of LOD is growing in terms of the number of disciplines committed to LD best practices and technologies for exposing and interlinking datasets for seamless access to and sharing, integration and reuse of those data. The progress has been tremendous, with close to 3,500 LOD datasets available according to *thedatahub.org* (May 2012).

Vocabularies, formalized as knowledge organization systems (KOS), help address problems with digital information overload and aid information discovery

Vocabularies are the base of linked data. Vocabularies, formalized as knowledge organization systems (KOS), based on the languages of a domain, discipline, or community, help address problems with digital information overload and aid information discovery. It is clear that vocabularies matter, not only in the traditional library world but also for many different digital information stakeholders. A recent survey by *Semantic Web Company* found that 85.4% of the 158 participants, coming from IT, science, public sector and education, use controlled vocabularies in their organizations. A very large proportion (88%) of these KOS users also agree that most organizations could benefit from linked data. 48.7% of respondents stated that standards like SKOS (simple knowledge organization system) are "very important" with a further 29.1% rating this approach as "relevant" (**Kondert; Schandl; Blumauer**, 2011).

Vocabulary users in the new LD landscape include different communities using complementary approaches to share and take advantage of vocabularies:

- Web developers embedding structured data describing web content into their html pages using encoding standards such as microformats, microdata and RDFa (instead of building up Sparql endpoints), described with a shared markup vocabulary such as *Schema.org*.

- The W3C semantic web community, who develop and use ontologies as formalized vocabularies. This includes using OWL to build vocabularies or ontologies, and SKOS for creating KOS in the process of data enrichment.
- Communities around metadata standards such as *Dublin Core* or particular metadata application profiles that include sets of descriptive elements and properties to constitute a vocabulary or metadata schema.
- Users of knowledge organization systems (KOS) that encompass value-space structures, such as thesauri, subject classifications and authority files, traditionally considered "controlled vocabularies".

A perceptive view of the LD landscape is found in a recent post on **Bernard Vatant's** blog (2012): He presents an inspiring list of existing and candidate "vocabulary publishers" including standard-developing organizations (*W3C, DCMI*); institutional heritage curators (*Library of Congress*) and the global organizations federating the work of those curators (e.g. *Ifla, European, Ocllc*); media groups and associations (*BBC, The NYT, The Guardian*); governments and institutional data providers (*data.gov, UN, World Bank*), research centers (*Deri, Inria*); specific funded research projects; small and medium enterprises in the field (*Talis, Mondeca*); large internet companies through initiatives like *Schema.org* (*Google, Bing, Yahoo!* and *Yandex*); and domain-specific and individual initiatives.

Linked data uses vocabularies in two ways: as schemas, providing a set of properties that an object might have; and as vocabulary encoding schemes describing the range of values a given property might take, in formal terms. Linked vocabularies aid knowledge acquisition by tightly controlling and contextualizing data (concepts, objects, etc.). This approach enables familiar metadata processes, while the ability to link both vocabularies and data records themselves provides an infrastructure that allows more effective information discovery and use.

This article first presents a brief history of ideas around linked data, from its semantic web foundations to the evolution of definitions of linked data, linked open data, and linked open vocabularies. We introduce the *HIVE* framework and discuss how the *HIVE-ES* extension for Spanish language vocabularies leads to a more global approach for LOV. The conclusion summarizes this work and offers several thoughts that have emerged from this inquiry.

2. Linked open data in context

Linked open data (LOD) or simply linked data (LD) has become a standard topic of calls for papers in the information and library science, web and computing science communities. Attention to linked data is also visible in a range of sub-

ject domains, such as health, government, and education. In fact, nearly every discipline producing data has a cluster of advocates promoting open and sharable data. Perhaps less obvious, but significant, is the goal to move toward linked open vocabularies (LOV), a step that aids data sharing and reuse as well as linking knowledge. This trend echoes earlier arguments put forth for moving toward the semantic web; linked data is integral for the development of a semantic web. In exploring LOV, it is useful to gain a sense of how this development fits into the history of the web.

2.1. A brief history of linked data

The idea of linked data did not emerge overnight. Rather, like many developments in the field of information and library science and computing, the foundational ideas and goals predate the Web. Common examples take us back to the notion of universal bibliographic control, or **Vannevar Bush's** (1945) presentation of the *Memex*, a hypothetical device using associative linking to sustain memory over time. As **Dan Brickley** (2012) stated recently, some current work in linked data and semantic web could benefit from understanding its place in a longer history going back to 1912 annual report from the Belgian *Institute of Bibliography*.

Tim Berners-Lee's original ideas from 1989 were not quite the Web we have today, but focused more toward linking concepts, ideas, facts –i.e. data– on a global scale. This initial idea of the Web has been promoted as the semantic web, “an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation” (**Berners-Lee**, 1999). **Berners-Lee** has been promoting, and perhaps defending, the same idea since at least 1989: a powerful structure of interconnected knowledge that linked information, documents

and data; he called it first the Web (1989), then, the semantic web (1998) and then linked data (2006). As he envisaged in *Weaving the Web* (1999), html allowed the hypertextual web of documents, while RDF and semantic web technologies (OWL, SKOS, Sparql) will allow the web of data through linked datasets defined as RDF triples.

Figure 1 shows historical landmarks toward the semantic web and linked data, using as a guideline the *World Wide Web Consortium (W3C)* standards and initiatives which made it possible. A thoughtful compilation of those semantic web standards has been completed in Spanish by **Pastor** (2011).

The semantic web allows machines to derive meaning from structured data which can be processed when published as linked data. Linked data is a key implementation step toward the semantic web, where representing information entities via URIs makes them machine-processable. Linked data are therefore essential ingredients of the semantic web. But to achieve a real global semantic web, just linked data or just open data is not sufficient. What is necessary is data that is both linked and open. That is why **Berners-Lee**, two decades after inventing the web, and after a decade focused on the development of the semantic web, encouraged people in his 2009 TED speech to open the “raw data, now!”, and express them as linked data, to enable getting them out of their silos.

2.2. Linked data, LOD dissected and defined

Explanations, discussion, and formal definitions for linked data generally reflect **Berners-Lee's** (2006) foundations and recognize LD as a set of best practices for publishing and connecting structured data on the Web. Linking data is about using the Web to connect related data that were not previ-

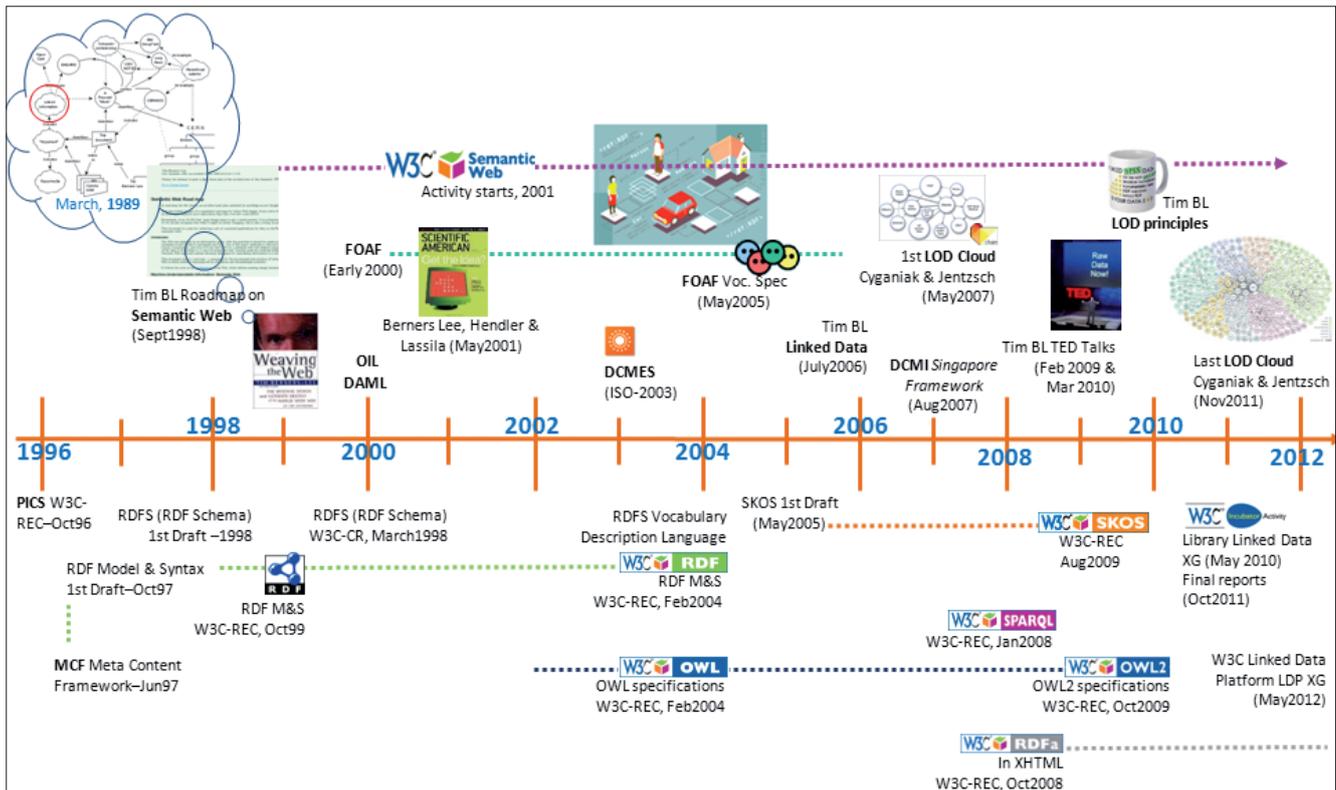


Figure 1. Evolution of the semantic web into linked data (image by E. Méndez)

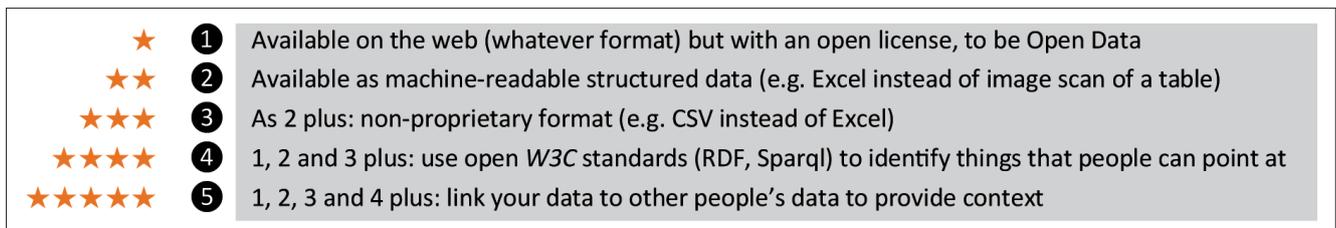


Figure 2. Openness principles added to the linked data paradigm (Berners-Lee, 2006 updated in 2010)

ously linked, or using the Web to lower the barriers to using data currently published using other methods (*Linkeddata.org*, 2012). Those foundations have basically four rules and five principles. The rules are pretty simple: 1) Use URIs as names for things (cool URIs for the semantic web are those that not change); 2) Use http URIs (dereferenciable URIs) so that people can look up those names; 3) When someone looks up a URI, provide useful information using the appropriate standards (RDF, Sparql); and 4) Include links to other URIs, so that people can discover more things. In 2010 he refined his conception of LD adding the openness philosophy in such a way that linked *open* data is linked data which is released under an open license, which does not impede its reuse for free, and he stated five levels or principles of LOD, like star ratings for hotels:

In general terms, linked data can be understood as an approach for encoding data at the most granular level. Data, in this context can be any atomic thing (including a concept), or an RDF statement or a set of statements, that has an identifier in the form of a URI. Data and objects that are part of linked data paradigm can come from standard vocabulary systems such as Dublin Core and from vocabulary encoding schemes in the form of controlled vocabularies, ontologies, taxonomies, name authority files, classificatory systems, and the like. Objects are selected from these vocabularies for property/value encoding, or refinements of this type of information. There is no limit to the types of vocabularies that can be transformed into linked data. One could publish the entire *Oxford English Dictionary* or the *Diccionario de la RAE* (Spanish Royal Academy Dictionary) as linked data.

Linked data uses uniform resource identifiers (URIs) as globally unique identifiers for any kind of resource analogously to how we use identifiers in librarianship for authority control. **Eric Miller** underlined the importance of the identifiers and pointed out that the persistence of these identifiers is a crucial part of integrity within systems. Using http URIs as a way of unifying local primary keys inside databases charts out a universal data space, not only for library organizations, but for any organization that wants to share information with others. "Traditionally we have kept these local identifiers inside systems. Now we are exposing the local identifiers so that external information can be hooked to them" (**Miller**, 2011), so every single data/thing in the LD world could be linkable.

A number of information professionals and researchers are realizing the significant added value that galleries, archives and museums can acquire by opening and linking their cultural data, notably **Byrne** and **Goddard** (2011) in Canada, **Oomen et al.** in the Netherlands (2012), **Saorín** (2012) and **Peset et al.** (2011) in Spain, and **Jon Voss** (LOD-LAM, 2011)

in the U.S. But the most important contribution, seen from the perspective of memory organizations and people working in digital humanities are the reports created by *W3C*'s library linked data incubator group (*W3C LLD XG*, 2011a & b; **Baker**, 2012). These define linked data as data published in accordance with principles designed to facilitate linkages among datasets, element sets, and value vocabularies. Those vocabulary encoding schemes (value vocabularies) are integral to the *HIVE* approach discussed here below.

3. From LOD to LOV: linked open vocabularies as a part of the new knowledge organization ecosystem

Research has shown that the "subject" or conceptual search based on a topic is the most common type of search in the context of the Web (**Yu; Young**, 2004; **Savolainen; Kari**, 2006). That is, people look for information on a topic such as a place of travel, a health condition, or a historical event, more frequently than they are looking for a specific organization, the official movie trailer for a new release, or a specific article. Even when a person is in search of a "known entity" such as a person or place, he may use conceptual terms. It is the *aboutness* (the subject) that we seek, and it is a semantic search that we need.

“It is the *aboutness* (the subject) what we seek, and it is a semantic search that we need”

For example a person interested in information about the history of the *Parque del Retiro* (Retiro park, in Madrid) may initiate a web query by searching with the concepts "Spain" and "parks," and perhaps the name "Retiro". The query in this case is not specific to the geographic coordinates, rather the searcher may be equally interested to learn of the park's development in 1632 as a retreat outside the city wall for King Philip IV and Royal family, or how it was used during the Spanish civil war. The *aboutness* or topical content can be represented by concepts from skosified vocabularies. With linked open vocabularies, there is a tremendous potential to bring this benefit to the larger global Web by linking information in an open environment.

3.1. Opening and linking vocabularies to build up a new KOS ecosystem

For vocabularies to be reused in this new ecosystem they need to be published as LOD. There are many vocabularies available on the web although their encoding may differ. Some time ago we foresaw the potential for XML/RDF to encode and express vocabularies at intranet or corporate level

(Méndez, 2000). The development of a standards-based semantic web (fig. 1) offers the possibility of global semantic interoperability through the web of LD. SKOS, a W3C recommendation since 2009, is a step forward to make thesauri more interoperable, making them sharable and accessible for project integration and linking content. This approach fosters the semantic web environment where vocabularies are explicitly created for the web. In this context, we present linked open vocabularies.

The technical implementations that link open vocabularies are based on semantic web technologies, standards and best practices, including RDF/SKOS, OWL, Sparql endpoints and triple stores. The use of semantic web technology offers the potential for better dissemination and integration of data into third-party applications and environments (Vatant, 2012). Using semantic web standards we not only put vocabularies “on” the web, for people to read about, but also publish them “for” the semantic web, enabling machines to use them directly. The *Datahub* is a community-run catalogue of datasets that also includes vocabulary encoding schemes, among a huge amount of “things” currently published as linked data. LOV might be understood as a clarified subset of the messy LOD cloud, those vocabularies that can be considered KOS (knowledge organization systems) or subject-oriented vocabularies. Bernard Vatant (2012) in his thorough provoking posts of *LOV stories* says that “there are as many data looking for good vocabularies as vocabularies looking for data”. He also thinks that the activities to create a new ecosystem for linked open vocabularies should be funded in the philosophy of the commons and *co-opetition*, as shared resources in which each stakeholder has an equal interest leading them to establish a cooperative competition.

Using semantic web standards we not only put vocabularies *on* the web, for people to read about, but also publish them *for* the semantic web, enabling machines to use them directly

3.2. Drawing the new knowledge organization system’s landscape. Linked open vocabularies initiatives

There are a number of collective registries providing access to the open vocabularies this way in a linked data/semantic web environment, reconfiguring a new landscape for network knowledge organization systems. Four approaches are noted here:

Linked Open Vocabularies project (LOV). The LOV registry created by Bernard Vatant and Pierre-Yves Vandenbussche and published by *Mondeca* labs as part of the *Datalift* project, applied in March 2012 to be hosted by the *Open Knowledge Foundation (OKFN)*, like the *Datahub*. As they stated in this application, LOV project aims to provide easy access to vocabularies, in particular the ways they link to each other, and by providing metrics on how they are used in the linked data cloud to help improve their comprehension, visibility, usability, and overall quality.

This is one of the most innovative projects in this area; and as already noted the ideas of LOV emphasize the ideas underlying this paper. The LOV project presents a “growing ecosystem of linked open vocabularies (RDFS or OWL ontologies) used in the LOD cloud”. The project catalog gives access (by search or navigation) to 262 vocabularies so far, rendered in RDF, OWL and SKOS that are classified by vocabulary spaces, interlinked using the dedicated vocabulary *VOAF (vocabulary of a friend)* to describe ontologies that are part of the linked data cloud. It works as a registry, but they have taken an extra step to classify the vocabularies in a broad approach.

Open Metadata Registry (OMR). The initiative “began its development as the *NSDL Registry*, attempting to address the big question: What should these registries do and how can they operate in an open services environment?” (Phipps; Hillman, 2011) and is today becoming the most crucial registry for the library community. The OMR extends well beyond the library, archival, and museum community, given the original seed in scientific education, and includes a range of vocabularies (element sets, ontologies and controlled vocabularies). An innovative aspect of this registry is the *sandbox* component where individuals can experiment, play, share, and learn to how to participate in the LOV environment. Work is ongoing to enable the OMR to support multilingual vocabularies and separate but associated language versions. One of the chief creators, Diane Hillman, is a founder and community leader of the *DCMI-Vocabulary Management Community (DCMI-VMC)*, a forum addressing best practices for vocabularies in the semantic web.

Amalgame (Amsterdam alignment generation metatool) is an interactive alignment server under development at the *VU University Amsterdam* in the context of the *PrestoPrime* and *EuropeanaConnect* projects. This tool realizes the second step of a specific workflow to skosify vocabularies and convert collection metadata to the *Europeana data model (EDM)*. *Amalgame* aims to find, evaluate and manage vocabulary alignment within the context of the *Ontology alignment evaluation initiative (OAEI)*, in which different alignment methods can be combined using a workflow setup. The main difference from the LOV project is that *Amalgame* includes a limited number of vocabularies, related mainly to cultural heritage, and looks for alignment among them. The main difference from OMR is that *Amalgame* only includes vocabulary encoding schemes, for an alignment among them based on axis vocabularies.

NCBO-BioPortal is a web-based application that provides access to 302 biomedical ontologies and vocabularies including thesauri in the field of biology and biomedicine. The ontology library specifically allows for browsing, searching, and downloading ontologies. There is also work enabling semantic mapping among vocabularies. In this case, ontologies registered are encoded in OWL; OBO, an ontology language for the bio domain; and increasingly SKOS renderings are being registered and provided. All these vocabularies or ontologies are used to support conceptual search over biomedical resources. It is more than a specialized vocabulary service. Once a search term is refined and aligned among different vocabularies, you can click and explore the re-

Figure 3. LOV Project, search about SIOC (Semantically-interlinked online communities) vocabulary

source in its original site. The main difference with this registry from those previously described is the domain specific approach.

“HIVE and HIVE-ES provide a solid foundation for linking and opening interdisciplinary vocabularies to the multilingual environment”

4. HIVE-ES framework: skosifying, opening, and linking vocabularies

HIVE (Helping Interdisciplinary Vocabulary Engineering), launched with support from the U.S. Institute of Museum and Library Studies, has been pursued as a demonstration project. HIVE presents a model using open, linked vocabularies for dynamically creating subject metadata at the time of indexing, drawing from multiple vocabularies. This process enables a selection of the best concepts for representing the content of the resource that is not limited to a single vocabulary. In the metaphor that **Greenberg** (principal investigator) and members of the HIVE team have used to explain it, a bee goes out to flowers seeking pollen, and brings nuggets back to the hive; in this case, relevant concepts. A similar idea, conveyed by **Bernard Vatant** (2012) through his “LOV gardens” metaphor where vocabularies like *Schema.org* are trees or flowers in the garden, complements this idea.

The HIVE-ES (España) initiative is building on this work to address known KOS challenges, and provide a means for simultaneous search and metadata generation drawing from multiple Spanish language vocabularies. Specifically, HIVE and HIVE-ES have been initiated to address vocabulary problems relating to cost, interoperability, and usability limitations (**Greenberg et al.**, 2011):

- Controlled vocabularies are expensive to create and maintain.
- Adherence to *ISO*, *Ansi/Niso*, *W3C*, *IETF*, and other standards does not confirm knowledge organization systems (KOS) interoperability.
- KOS design does not always support easy access and use.

The HIVE-ES initiative extends the initial HIVE activity to address these challenges in Spanish speaking countries. HIVE-ES allows for simultaneous search and metadata generation drawing from multiple Spanish language vocabularies. The project has been launched at the *Tecnodoc research group (Tecnologías Aplicadas a la Información y la Documentación)* of the *Department of Library and Information Science, Universidad Carlos III de Madrid*, the *National Library of Spain (BNE)*, and *Metadata Research Center, School of Information and Library Science, University of North Carolina at Chapel Hill (Sils-MRC)*.

HIVE-ES is making vocabularies and providing a demonstration project supporting interdisciplinary indexing. The emphasis here is on Spanish language vocabularies. Essentially, vocabularies integrated into the HIVE-ES are composed as linked data using SKOS encoding language, to represent their classes and properties.

HIVE-ES vocabulary server currently has three vocabularies: *The United Nations Food and Agriculture vocabulary (Agrovoc)* (the Spanish language section), *LEM (Lista de encabezamientos de materia)* and *EMBNE (Encabezamientos de materia-BNE, the Spanish National Library's equivalent of the Library of Congress Subject Headings)*. The integration of EMBNE in HIVE-ES was the most complex undertaking, given the need to convert Marc21 format for authority data into SKOS. The processing involved 1) conceptually mapping Marc21 fields (e.g., 1XX, 4XX and 5XX) to appropriate SKOS labels (e.g., ‘skos:prefLabel’ and ‘skos:altLabel’); 2) reading and parsing the EMBNE file in Marc21 format; and 3) re-labeling the individual contents of the EMBNE to the SKOS

Figure 4. HIVE-ES project wiki

equivalents. Conversion is not supported by a single tool, although the *Marc4J Java library*, which defines an application programming interface (API) for parsing large Marc files, was a tremendous help.

The framework and infrastructure of *HIVE-ES* follows that of the original *HIVE* and includes:

- *HIVE* Core for automatic metadata extraction, topic detection and concept retrieval. The extraction uses KEA (keyphrase extraction algorithm) and concept retrieval is supported via *Lucene*. These are the system's main functions. Additionally, the core includes RDF storage and management using *Elmo* to store objects and properties in an RDF repository supported by *Sesame*.
- *HIVE* vocabulary service includes *HIVE*'s web accessible interface so that users can browse and search through vocabularies in the demo system. This was developed using the *Google web toolkit*.
- *HIVE* REST (*representational state transfer*) aspect provides an API based on web services to facilitate integration for third party software.

HIVE-ES is in the early stages of development and implementation, and plans are under way to add more Spanish language vocabularies. Current *HIVE* renderings that have been documented are monolingual. The current *HIVE* code doesn't support the integration of multilingual vocabularies, but the open nature of this initiative and sharing of code opens the door for other developers to contribute to *HIVE* in this manner. The *HIVE-ES* project pushes the edge here highlighting the need to look at vocabularies in languages other than English, moving toward a global context.

5. Conclusion

This paper provides insight into the evolving trends advancing knowledge organization and sharing vocabularies from the perspective of linked data and linked open data. Among several firm conclusions to be offered are:

- The technological infrastructure to support LD/LOD has become quite powerful through the use of networked technologies and the development and adoption of W3C standards (RDF, SKOS, Sparql, etc.).
- Vocabularies in their many forms (thesauri, taxonomy, ontology, and discipline, domain and community languages) can be leveraged and made more powerful via RDF/SKOS.
- Many communities are embracing this new potential, skosifying their vocabularies and making them open and interoperable.
- Registries are progressing to a new level with developments such as *LOV project*, *OMR*, *Amalgame*, and the *NC Bioportal*, and providing a sustainable means for sharing vocabularies and supporting semantic web operations.
- Publishing open vocabularies is not just about vocabularies developed for the web. **Tim Berners-Lee's** "five stars" principles for LOD are generally applicable to enabling knowledge organization systems to augment the store of information available for effective use (open licensing).
- The community of people embracing LOD and LOV is growing swiftly and with enthusiasm, embracing both scientific domains and humanities-driven endeavors such as *Europeana*.

Proprietary thesauri and other knowledge organization systems have the potential to create a new landscape, and efforts are forging ahead beyond a limited display of textual or hypertextual content. These KOS are increasingly both "on" the web and being created "for" the semantic web. The examples provided in this paper, and additional progress with several aspects of the linked data infrastructure, are important steps in this new ecosystem. *HIVE* is also an exemplary approach in this new LOV landscape. The *HIVE-ES* extension of the original *HIVE* is for Spanish language vocabularies, leading to a more global approach for linked open vocabularies. These steps provide a solid foundation for linking and opening vocabularies to the multilingual environment and sharing value spaces vocabularies across languages.

As digital information initiatives aim to make their resources, or descriptions of their holdings, accessible via the global web, they strive to provide a venue for interdisciplinary and cross-disciplinary research. Linked open vocabularies are part of this growing trend; however, frameworks allowing for simultaneous searching and indexing via multiple vocabularies in this context require further work to reach a fully interoperable environment. The move to sharing in this way needs to be supported by a broader social and technological infrastructure that brings people, technologies, and vocabulary together. A sense of community and shared vision is integral to make this approach sustainable. The *DCMI Vocabulary Management Community* is gaining significant interest in this respect, with leadership, inspiration, and engagement from members of the *OMR* and *LOV* projects. It is the collection and coordination of developments and community that will allow this evolution to reach its full potential.

6. Acknowledgements

The authors would like to acknowledge the thoughtful review and feedback from colleagues **Dan Brickley** and **Charles McCathieNevile** to this paper, and members of the *HIVE* and *HIVE-ES* team for their contributions to the project.

The foundation of *HIVE* technology is supported by *IMLS* grant *LG-07-08-0120-08*.

7. References

- Baker, Thomas.** "Libraries, languages of description, and linked data: a Dublin Core perspective". *Library hi tech*, 2012, v. 30, n. 1, pp. 116-133.
<http://dx.doi.org/10.1108/07378831211213256>
- Berners-Lee, Tim.** Information management: a proposal, 1989.
<http://www.w3.org/History/1989/proposal.html>
- Berners-Lee, Tim.** "Semantic web road map". *Design issues*, 1998, September.
<http://www.w3.org/DesignIssues/Semantic.html>
- Berners-Lee, Tim.** *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. San Francisco: Harper, 1999.
- Berners-Lee, Tim.** "Linked data". *Design issues*, 2006-07-27.
<http://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, Tim.** "The next Web of open, linked data". *TED*, February, 2009
https://www.ted.com/talks/tim_bern timers_lee_on_the_next_web.html
- Brickley, Dan.** "Semantic web, part 5". In: *Libraries, media & the semantic web*, hosted by the *BBC*, 28 March 2012, video.
http://www.youtube.com/watch?v=-6mhdjE1XE&feature=player_embedded
- Bush, Vannevar.** "As we may think". *The Atlantic monthly*, 1945, July.
<http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/3881>
- Byrne, Gillian; Goddard, Lisa.** "The strongest link: Libraries and linked data". *DLib magazine*, 2010, v. 16, n. 11-12.
<http://www.dlib.org/dlib/november10/byrne/11byrne.html>
- Greenberg, Jane et al.** *HIVE: Helping Interdisciplinary Vocabulary Engineering*. *Bulletin of the American Society for Information Science and Technology*, 2011, v. 37, n. 4.
http://www.asis.org/Bulletin/Apr-11/AprMay11_Greenberg_etAl.html
- Kondert, Florian; Schandl, Thomas; Blumauer, Andreas.** *Do controlled vocabularies matter? Survey results*. Semantic Web Company, Viena, 2011.
http://issuu.com/andreas_blumauer/docs/survey_do_controlled_vocabularies_matter_2011_june#download
- Méndez, Eva.** "Metadatos y tesauros: aplicación de XML/RDF a los sistemas de organización del conocimiento en intranets". In: *VII Jornadas españolas de documentación*, Bilbao (Spain), 2000, October, 19-20, pp. 211-219.
<http://hdl.handle.net/10760/12698>
- Miller, Eric.** Linked data and libraries [recorded by Micheline Westfall]. *Serials librarian*, 2011, v. 60, n. 1-4, pp. 17-22.
<http://dx.doi.org/10.1080/0361526X.2011.556427>
- Oomen, Johan; Baltussen, Lotte-Belice; Van Erp, Marieke.** "Sharing cultural heritage the linked open data way: why you should sign up". In: *Museums and the Web 2012*. San Diego, April, 11-14.
http://www.museumsandtheweb.com/mw2012/papers/sharing_cultural_heritage_the_linked_open_data
- Pastor, Juan-Antonio.** *Tecnologías de la web semántica*. Colección EPI. Barcelona: UOC, 2011.
<http://www.elprofesionaldelainformacion.com/libros/tecnologias-web-semantica.html>
- Peset, Fernanda; Ferrer-Sapena, Antonia; Subirats-Coll, Imma.** "Open data y linked open data: su impacto en el área de bibliotecas y documentación". *El profesional de la información*, 2011, marzo-abril, v. 20, n. 2, pp. 165-173.
<http://dx.doi.org/10.3145/epi.2011.mar.06>
- Phipps, Jon; Hillman, Diane.** "The Open Metadata Registry: an update". *Bulletin of the American Society for Information Science and Technology*, 2011, v. 37, n. 4.
http://www.asis.org/Bulletin/Apr-11/AprMay11_Phipps_Hillmann.pdf
- Saorín, Tomás.** "Cómo linked open data impactará en las bibliotecas a través de la innovación abierta". *Anuario ThinkEPI*, 2012, v. 6, pp. 288-292.
<http://hdl.handle.net/10760/16913>
- Savolainen, Reijo; Kari, Jarkko.** "User-defined relevance criteria in web searching". *Journal of documentation*, 2006, v. 62, n. 6, pp. 685-707.
<http://dx.doi.org/10.1108/00220410610714921>
- Vatant, Bernard.** *LOV stories, Part 2: Gardeners and gatekeepers*. *The wheel and the hub: Tracks in the knowledge commons* [blog], 22-03-2012.
<http://blog.subjects.com/2012/03/lov-stories-part-2-gardeners-and.html>
- W3C LLD XG. *Library Linked Data Incubator Group Final Re-*

port, 25 October 2011. Tomas Baker *et al.*, eds., 2011a.
<http://www.w3.org/2005/Incubator/llid/XGR-llid-20111025>

W3C LLD XG. *Library Linked Data Incubator Group: Datasets, value vocabularies, and metadata element sets*, 25 October 2011, Antoine Isaac *et al.*, eds., 2011b.
<http://www.w3.org/2005/Incubator/llid/XGR-llid-vocabdataset-20111025>

Yu, Holly; Young, Margo. "The impact of web search engines on subject searching in opac". *Information technology and libraries*, 2004, v. 23, n. 4, pp. 168-180.
<http://dx.doi.org/10.1234/12345678>

8. Relevant cited websites

Amalgame (Amsterdam alignment generation metatool)
<http://semanticweb.cs.vu.nl/amalgame>

DC, DCMI (Dublin Core, Dublin Core metadata initiative)
<http://dublincore.org>

DCMI-VMC (DCMI Vocabulary Management Community)
<http://dublincore.org/groups/vocabulary-management>

DataHub
<http://thedatahub.org>

EDM (Europeana data model)
<http://pro.europeana.eu/edm-documentation>

FOAF (friend of a friend)
<http://www.foaf-project.org>

HIVE (Helping interdisciplinary vocabularies engineering)
<http://ils.unc.edu/mrc/hive>

HIVE Project Wiki
https://www.nescent.org/sites/hive/Main_Page

HIVE-ES (Helping interdisciplinary vocabularies engineering-España). Project wiki
<http://klingon.uc3m.es/hive-es/wiki>

HIVE-ES server
<http://klingon.uc3m.es:8080/home.html>

KEA (keyphrase extraction algorithm)
<http://www.nzdl.org/Kea>

Linked Data
<http://linkeddata.org>

LOV project (linked open vocabularies)
<http://labs.mondeca.com/dataset/lov>

Marc4J java library
<http://marc4j.tigris.org>

NCBO-BioPortal
<http://bioportal.bioontology.org>

OAEI (Ontology alignment evaluation initiative)
<http://oei.ontologymatching.org>

OBO (The open biological and biomedical ontologies)
<http://obofoundry.org>

OMR (Open metadata registry)
<http://metadataregistry.org>

OKFN (Open Knowledge Foundation)
<http://okfn.org>

Schema.org
<http://schema.org>
<http://blog.schema.org>

Semanticweb.com
<http://semanticweb.com>

SIOC (Semantically-interlinked online communities)
<http://sioc-project.org>

VOAF (Vocabulary of a friend)
<http://labs.mondeca.com/vocab/voaf>

W3C (World Wide Web Consortium)

- Linked data
<http://www.w3.org/standards/semanticweb/data>
- MCF (meta content framework)
<http://www.w3.org/TR/NOTE-MCF-XML-970606>
- Sparql
<http://www.w3.org/TR/rdf-Sparql-query>
- RDF (resource description framework)
<http://www.w3.org/standards/techs/rdf>
- RDFa (RDF attributes)
<http://www.w3.org/TR/xhtml-rdfa-primer>
<http://rdfa.info>
- RDFS (RDF schema)
<http://www.w3.org/TR/rdf-schema>
- SKOS (simple knowledge organization system)
<http://www.w3.org/standards/techs/skos>
- Semantic web
<http://www.w3.org/standards/semanticweb>
- Vocabularies
<http://www.w3.org/standards/semanticweb/ontology>



APLICACIÓN DE SKOS PARA LA INTEROPERABILIDAD DE VOCABULARIOS CONTROLADOS EN EL ENTORNO DE LINKED OPEN DATA



Juan-Antonio Pastor-Sánchez, Francisco-Javier Martínez-Méndez
y José-Vicente Rodríguez-Muñoz



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 construcción de servicios de información digital. Lleva a cabo su investigación en el ámbito de las tecnologías de la web semántica y la gestión de contenidos digitales.

Universidad de Murcia
Facultad de Comunicación y Documentación
Campus Universitario de Espinardo, 30100 Murcia
pastor@um.es

Francisco-Javier Martínez-Méndez es doctor en documentación por la *Universidad de Murcia* y profesor de tecnologías de la información en la *Facultad de Comunicación y Documentación*. Principalmente desarrolla su investigación en el área de la recuperación de la información en la Web.

javima@um.es

José-Vicente Rodríguez-Muñoz es doctor en informática y catedrático del *Área de Biblioteconomía y Documentación* de la *Universidad de Murcia*. Su actividad científica se ha centrado en los campos de la gestión de información, recuperación de información y evaluación de sistemas de búsqueda web.

jovi@um.es

Resumen

Se pretende determinar el alcance de la aplicación de SKOS en el ámbito de la web semántica. Para ello se han analizado los vocabularios controlados que aplican SKOS y que se encuentran registrados en *The data hub*, un catálogo de conjuntos de datos RDF. Además de establecer una tipología de vocabularios, se analizan los aspectos relacionados con el acceso abierto a los datos mediante su descarga directa, su consulta a través de un *Sparql endpoint* y la existencia de licencias adecuadas. Se examina el nivel de interoperabilidad existente a través de la definición de relaciones de mapeado entre vocabularios. Se concluye indicando que tesauros y clasificaciones están más integrados en el entorno de los datos abiertos, mientras que los encabezamientos de materia poseen un mayor grado de interoperabilidad.

Palabras clave

SKOS, Vocabularios controlados, Clasificaciones, Tesauros, Web semántica, *Linked open data*, Interoperabilidad semántica.

Title: SKOS application for interoperability of controlled vocabularies in the field of linked open data

Artículo recibido el 29-02-12
Aceptación definitiva: 23-05-12

Abstract

This paper aims to determine the scope of application of SKOS in the Semantic Web. Controlled vocabularies registered in the *The Data Hub* catalog of RDF datasets that use SKOS were studied. In addition to establishing a typology of vocabularies, authors analyzed the issues related to open access to data through its direct download, its query through a *Sparql endpoint* and the existence of appropriate licenses. The existing level of interoperability through the definition of mapping relationships between vocabularies are examined. The study concludes that thesauri and classifications are more integrated into the open data environment while subject headings lists have greater interoperability.

Keywords

SKOS, Controlled vocabularies, Classifications, Thesaurus, Semantic web, Linked open data, Semantic interoperability, *Sparql endpoint*.

Pastor-Sánchez, Juan-Antonio; Martínez-Méndez, Francisco-Javier; Rodríguez-Muñoz, José-Vicente. "Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de *linked open data*". *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 245-253.

<http://dx.doi.org/10.3145/epi.2012.may.04>

Introducción

SKOS (*Simple knowledge organization system*) es una de las ontologías que mayor éxito y aplicación ha alcanzado en el entorno de la web semántica. Ofrece un modelo para representar la estructura básica y el contenido de esquemas de conceptos tales como listas encabezamientos de materia, taxonomías, esquemas de clasificación, tesauros y cualquier otro tipo de vocabulario controlado¹. El desarrollo de SKOS comenzó en el seno del grupo de trabajo *SWAD-Europe* hacia el año 2002 y se difundió públicamente mediante un borrador en noviembre de 2005. En aquel momento la propuesta se dio a conocer como *SKOS Core*, etiqueta que aún perdura en muchos trabajos actuales. En agosto de 2009 SKOS alcanzó el estatus de recomendación del W3C (2009). En comparación con otras soluciones aportadas, tales como esquemas XML *ad-hoc*, ontologías quizá más complejas o *topic maps*, SKOS ofrece una alternativa cuya aplicación es sencilla y rápida (Pastor-Sánchez; Martínez-Méndez; Rodríguez-Muñoz, 2009).

En el contexto de *linked open data* el establecimiento de vínculos entre conjuntos de datos abiertos resulta esencial y los lenguajes documentales deben desempeñar un papel relevante

Actualmente son numerosos los sistemas de organización del conocimiento de todo tipo publicados mediante SKOS disponibles para su uso y consulta. Se han realizado breves estudios al respecto (Isaac *et al.*, 2011) que muestran una visión muy general de la implantación de SKOS. Sin embargo, creemos necesario determinar de forma más precisa su auténtico potencial, valorando su impacto, evolución y aplicación en los conjuntos de datos asociados a lenguajes documentales de todo tipo. De esta forma sabremos si esta iniciativa tiene visos de asentarse definitivamente en el ecosistema de la web semántica o si por el contrario es algo pasajero.

Algunos trabajos anteriores (Möller *et al.*, 2010) abordan enfoques más generales con conjuntos de datos RDF² sin llegar a abordar aspectos relacionados con la interoperabilidad. Únicamente analizan la explotación real de conjuntos de datos, volumen de transacciones o índices de análisis de consultas *Sparql*³ estableciendo métricas en este sentido. Algo más específico es el estudio de Francesconi *et al.* (2008) que analiza la definición de vínculos de mapeado entre tesauros aplicando técnicas de recuperación de información.

El presente estudio global de la presencia de SKOS en la publicación de vocabularios controlados tiene como objetivo identificar su tipología, volumen y relaciones existentes con otros vocabularios o conjuntos de datos. De este modo se podrá ponderar debidamente el alcance de esta tecnología, especialmente en el ámbito de *linked open data*, contexto donde el establecimiento de vínculos entre conjuntos de datos abiertos resulta esencial y donde los lenguajes documentales deben desempeñar un papel relevante.

SKOS

SKOS se define formalmente como una ontología *OWL-full*⁴ que permite representar cualquier tipo de sistema de organización del conocimiento mediante RDF. Su ámbito de aplicación se extiende a la práctica totalidad de vocabularios controlados: clasificaciones, tesauros, encabezamientos de materia, taxonomías, tesauros, glosarios, etc.

En SKOS los elementos de un vocabulario se representan mediante conceptos entre los que se establecen relaciones semánticas jerárquicas (simples o transitivas) y asociativas. A los conceptos se les asocian etiquetas en distintos idiomas:

- Preferentes: equivalentes a los descriptores en un tesauro. Un mismo concepto sólo puede tener una etiqueta preferente en cada idioma.
- Alternativas: similares a los no-descriptores. Permite enriquecer semánticamente un vocabulario definiendo varios puntos de acceso a un concepto.
- Ocultas: no son visibles directamente a los usuarios y se utilizan para su procesamiento por aplicaciones informáticas.

Mediante la extensión *SKOS-XL* es posible definir relaciones entre etiquetas, por ejemplo, cuando una etiqueta es un acrónimo o un préstamo lingüístico de otra.

Pueden definirse esquemas de conceptos y colecciones. Los esquemas agrupan conceptos normalmente asociados a un campo semántico o área de conocimiento determinada. *SKOS* ofrece dos propiedades que permiten relacionar un concepto con uno o varios esquemas y especificar si se sitúa como cabecera de una estructura jerárquica (*top concept*). Las colecciones permiten crear grupos de conceptos que complementan las estructuras de relaciones semánticas jerárquicas. Un mismo concepto puede formar parte de varias colecciones.

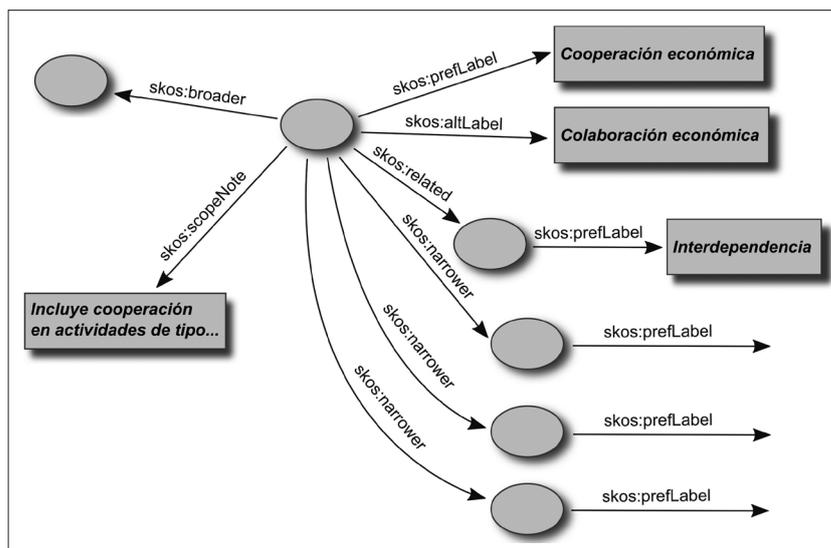


Figura 1. Representación de elementos de un vocabulario con *SKOS*
[http://www.w3.org/People/Ivan/CorePresentations/SW_Advanced/Slides.html#\(73\)](http://www.w3.org/People/Ivan/CorePresentations/SW_Advanced/Slides.html#(73))

SKOS ofrece una serie de relaciones semánticas para establecer vínculos de mapeado entre conceptos de diferentes esquemas. Esto permite indicar si un concepto de un esquema se considera idéntico a otro o cuándo tienen un significado cercano, genérico, específico o relacionado. La nueva norma de tesauros *ISO 29562* propone una función similar para definir relaciones entre diferentes lenguajes documentales con el objeto de poder utilizarse conjuntamente en operaciones de recuperación de información.

La relaciones de mapeado de *SKOS* son clave para la participación de los sistemas de organización del conocimiento en SRI basados en la web semántica

Por ejemplo, un repositorio digital utiliza un vocabulario controlado X para indizar documentos y un banco de imágenes hace lo propio con un vocabulario Y. Entre ambos vocabularios podrían definirse correspondencias entre los diferentes conceptos que lo componen. A través de una consulta en el banco de imágenes se recuperarían elementos del repositorio digital y viceversa. Estas relaciones son claves para la participación de los sistemas de organización de conocimiento en el escenario *linked open data*, puesto que el grado de interoperabilidad de un vocabulario controlado se determina por el establecimiento de relaciones de mapeado con otros vocabularios o conjuntos de datos.

The data hub como fuente de referencia del estudio

La fuente principal para la toma de datos de este estudio es el catálogo en línea *The data hub*⁵

(*TDH*). Se trata de un catálogo colaborativo con información descriptiva sobre todo tipo de conjuntos de datos disponibles en internet, mantenido por la *Open Knowledge Foundation* como parte de una iniciativa destinada a dar soporte a actividades para la difusión del conocimiento de carácter abierto.

La información que ofrece *TDH* se refiere a la identificación y localización de conjuntos de datos, descripción del tipo de contenido, ejemplos, información sobre la posibilidad de descarga total o parcial de dichos conjuntos de datos, existencia de *Sparql endpoints*⁶, número de tripletas RDF, vínculos con otros conjuntos de datos, espacios de nombres utilizados, etc. El servicio también permite asignar a cada conjunto de datos diferentes etiquetas que describen el contenido y formatos utilizados.

Se seleccionaron los conjuntos de datos que representan vocabularios controlados de todo tipo mediante *SKOS*. Debido al carácter abierto y cooperativo de este catálogo, las descripciones de los conjuntos de datos son heterogéneas y a menudo incompletas e incorrectas. Por ello, en algunas ocasiones se recuperaron resultados que, si bien en la descripción o etiquetas incluían el término “*SKOS*”, en realidad no se referían a ningún vocabulario controlado o realizaban un uso deficiente o incorrecto de *SKOS*. Se verificó si existían conjuntos de datos adicionales referidos a tesauros, taxono-

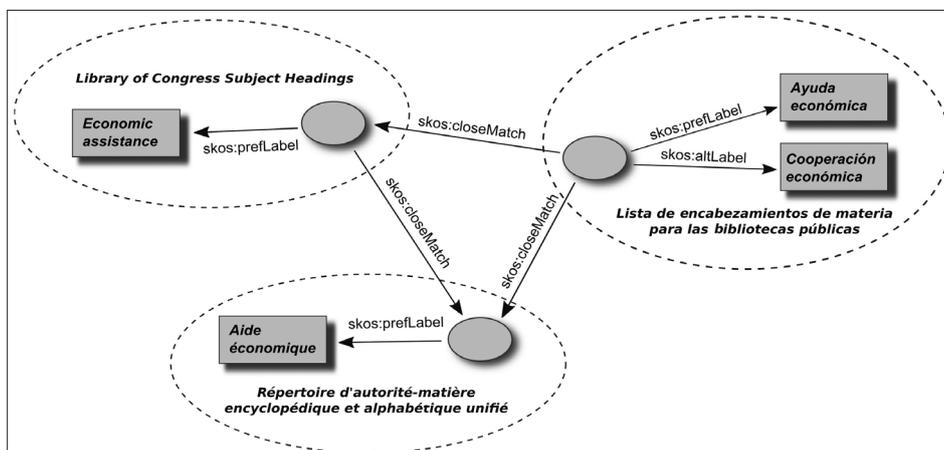


Figura 2. Interoperabilidad de vocabularios entre las LEM, LCSH y Rameau

mías, clasificaciones y glosarios de los que, pese a no aparecer descritos bajo la etiqueta “SKOS”, pudiera verificarse que hacen uso de dicha especificación. Esta verificación tuvo resultado negativo⁷.

⌋ Mientras que los tesauros son el tipo de vocabulario controlado más frecuente sobre el que se aplica SKOS, el mayor volumen de datos (más de un 73%) corresponde a los ficheros de control de autoridades

Se consultaron dos fuentes suplementarias: la página sobre *datasets* de SKOS del W3C⁸ y el informe del *Grupo Incubadora* (Isaac et al., 2011) sobre conjuntos de datos y vocabularios controlados en el ámbito de *library linked data*. Gracias a la primera fuente se localizaron 3 vocabularios adicionales: los del servicio *NASA Taxonomy*, el *UK Archival thesaurus (UKAT)* y la representación en SKOS de las distintas categorías de *Wikipedia* del proyecto *Dbpedia*. Otros vocabularios mencionados en dicha página no han podido ser analizados debido a la imposibilidad del acceso conjunto a los datos correspondientes o por hacer referencias a proyectos o trabajos inconclusos⁹.

De cada vocabulario se recopiló la siguiente información:

- denominación;
- volumen;
- relaciones con otros conjuntos de datos (incluidos los correspondientes a otros vocabularios);
- disponibilidad de servicios de consulta mediante *Sparql endpoint*;
- descarga de ficheros con datos RDF en cualquier serialización;
- existencia de licencias compatibles con el uso libre y abierto de los datos.

Esta información no pudo ser obtenida para todos los vocabularios a través de la búsqueda en TDH debido a las deficiencias del catálogo mencionadas anteriormente. Fue necesario recopilarla por otros medios, siempre y cuando fuera viable el acceso, descarga y análisis de los datos correspondientes. Se analizaron los ficheros fuente mediante el validador de documentos RDF del W3C¹⁰, obteniendo la información del número de tripletas y de vínculos con datos externos. La obtención de la información sobre la disponibilidad de los datos para su descarga o consulta mediante *Sparql endpoint* ha requerido la navegación por los sitios web donde están publicados los vocabularios. Algunos de ellos, como los encabezamientos de materia suecos o los vocabularios de la *Biblioteca Nacional de Hungría*, han quedado fuera de este estudio porque TDH no ofrece información estadística sobre el volumen de datos y relaciones y tampoco fue posible acceder y descargar los conjuntos de datos correspondientes.

Resultados obtenidos

Los resultados se han dispuesto en varios grupos de características para un análisis estructurado. Se han estudiado los

tipos de vocabularios identificados para una comprensión más adecuada del contexto de aplicación de SKOS en el ámbito de *linked open data*. También se ha analizado el nivel de apertura de los vocabularios para su consulta y descarga, teniendo en cuenta la presencia explícita de una licencia compatible con la reutilización abierta y libre de los datos. Finalmente se ha examinado el grado de interoperabilidad de los vocabularios a través de las relaciones de mapeado establecidas entre sí¹¹.

Tipos de vocabularios y volumen de datos

Se han examinado conjuntos de datos correspondientes a 55 vocabularios controlados agrupados en 6 tipos: tesauros, listas de encabezamientos de materia, ficheros de control de autoridades, clasificaciones, léxicos y vocabularios, taxonomías y ontologías. Como refleja la tabla 1, más de la mitad de los conjuntos de datos analizados se corresponden con tesauros (19) y clasificaciones (16). Desde el punto de vista del volumen de datos, de los más de 347 millones de tripletas que forman los conjuntos de datos analizados, un 73,6% (tabla 1) se corresponden con ficheros de control de autoridades y un 10,6% con listas de encabezamientos de materia. En consecuencia, la distribución porcentual del número de vocabularios en función de su tipo (% frecuencia) no se corresponde con la distribución porcentual de tripletas de cada tipo (% volumen de datos).

Los vocabularios de mayor volumen son el *Fichero de autoridades virtual internacional (VIAF)* y la *Aplicación facetada de terminología temática (FAST)* basada en la *Lista de encabezamientos de materia de la Biblioteca del Congreso*. Resulta interesante destacar que del volumen total de tripletas de tesauros, más del 70% se corresponde con el *Tesauro de medio ambiente multilingüe general (Gemet)*.

El “triángulo open data”

Hay que analizar si estos vocabularios cumplen 3 criterios que indiquen si se encuentran integrados en la iniciativa *linked open data*, es decir determinar si en cada conjunto de datos:

1. Es posible la descarga conjunta de todo el vocabulario en un formato compatible con RDF tal como *N3*, *Turtle* o *RDF/XML*. El acceso íntegro a los conjuntos de datos en alguno de estos formatos resulta esencial para su reutilización.

Tipo de vocabulario	Vocabularios analizados	Frecuencia (%)	Número de tripletas	Volumen de datos (%)
Tesauros	19	34,6	28.309.759	8,2
Clasificaciones	16	29,1	6.891.148	2,0
Encabezamientos de materia	7	12,7	36.724.495	10,6
Autoridades	4	7,3	255.528.450	73,6
Léxicos y glosarios	4	7,3	388.729	0,1
Ontologías	3	5,5	5.003.228	1,4
Taxonomías	2	3,6	14.542.505	4,2
Totales	55	100	347.388.314	100

Tabla 1. Distribuciones absolutas y porcentuales de las frecuencias y volumen de datos por tipo de vocabularios analizados

2. Existe la posibilidad de utilizar un *Sparql endpoint* para la consulta selectiva de los datos RDF correspondientes al vocabulario. De este modo es posible una reutilización más selectiva, integrando servicios web y reutilizando vocabularios.

3. Se ha establecido una licencia compatible con el uso abierto y libre de los datos.

Estos criterios se complementan entre sí: una licencia *open data* sirve de poco si no se suministran los datos o si no es posible su reutilización inmediata y selectiva a través de consultas *Sparql*. Los servicios de descarga y consulta tampoco resultan muy útiles si es imposible utilizar dichos datos debido a que carecen de una licencia adecuada (Alexander et al., 2009).

Resulta útil representar estos datos agrupándolos según el tipo de vocabulario. La tabla 2 muestra los datos absolutos en términos totales y porcentuales (columnas “abs n” y “abs %” respectivamente) del grado de cumplimiento de cada uno de los tres criterios indicados anteriormente por parte de cada tipo vocabulario. Las licencias *open data* más utilizadas en los diferentes conjuntos de datos de los vocabularios analizados son *Open knowledge definition*, *Open database licence* y diversas modalidades de *Creative commons*.

Clasificaciones y tesauros destacan por su alto grado de integración en el ecosistema de datos abiertos, con mecanismos y licencias adecuados para ello

Es necesario tener en cuenta el peso de cada vocabulario en función del volumen de datos que aporta para cada tipo de vocabulario. Un vocabulario con 1.000 tripletas no tiene el mismo peso que otro que aporte 10 millones. Por este motivo se ha calculado el grado de cumplimiento de cada uno de los criterios para cada tipo de vocabulario realizando una ponderación en función del número de tripletas que aporta.

Dicho cálculo se define de la siguiente forma: sea una familia *T* de conjuntos de tripletas de cada vocabulario de un determinado tipo $\{X_1, \dots, X_n\}$ donde el número de conjuntos

que forman parte de dicha familia viene definido por una función de cardinalidad¹² $card(T)$. Sea X_i un vocabulario tal que $X_i \in T$ cuyo número de tripletas viene dado por su cardinalidad $card(X_i)$ y $p(X_i, C)$ una función binaria que indica si dicho vocabulario cumple (1) o no (0) un determinado criterio *C* de participación en *open data*. El nivel de participación en términos porcentuales, $OD(T, C)$, de un tipo de vocabulario *T*, en un determinado criterio *C open data*, se calcula como:

$$OD(T, C) = \sum_{k=1}^{card(T)} \left(\frac{card(X_k) \cdot p(X_k, C)}{S(T)} \cdot 100 \right)$$

donde

$$S(T) = \sum_{j=1}^{card(T)} card(X_j)$$

Tras realizar los cálculos se han obtenido los datos porcentuales (columnas $OD(T, C_j)$ de la tabla 2) mediante la ponderación en función del volumen de los vocabularios de cada tipo. El mismo cálculo se ha realizado, entendiendo que todos los vocabularios forman parte de un único conjunto, lo que nos permite obtener una visión global.

Se ha escogido una gráfica radial (figura 3) para representar el grado porcentual de cumplimiento de cada uno de los criterios para tesauros, clasificaciones, listas de encabezamientos de materia y autoridades, los cuatro tipos de vocabularios que mayor volumen de datos aportan.

Cada uno de ellos conforma lo que hemos convenido denominar el “triángulo *open data*”. Destaca el alto grado de adecuación tanto de tesauros como de clasificaciones con los tres criterios, mientras que son las listas de encabezamientos de materia y los ficheros de control de autoridades los que más lejos están de cumplirlos, en especial la disponibilidad de *Sparql endpoint*. Estos resultados negativos son muy pronunciados en estos tipos, puesto que los dos vocabularios más numerosos del estudio, *VIAF* y *FAST*, carecen tanto de opciones de descarga como de *Sparql endpoint*, y en el caso de *VIAF* no se ha definido de forma explícita una licencia *open data*. El mismo cálculo se ha realizado de forma global (entendiendo las tripletas de todos los vocabula-

Tipo de vocabulario	Vocabularios analizados	Criterio 1: Descarga de datos			Criterio 2: Sparql endpoint			Criterio 3: Licencia open data		
		abs n	abs %	OD(T,C ₁) %	abs n	abs %	OD(T,C ₂) %	abs n	abs %	OD(T,C ₃) %
Autoridades	4	3	75	22	1	25	6	2	50	16
Clasificaciones	16	15	94	94	14	88	100	14	88	100
Encabezamientos de materia	7	5	71	18	3	43	1	5	71	95
Léxicos y glosarios	4	3	75	98	1	25	12	2	50	12
Ontologías	3	3	100	100	2	67	10	2	67	10
Taxonomías	2	2	100	100	1	50	98	2	100	100
Tesauros	19	17	89	97	11	58	92	15	79	98
Totales	55	48	87	33	33	60	18	42	76	36

Tabla 2. Resultados absolutos y ponderados de aplicación de criterios *open data* agrupados por tipos de vocabularios

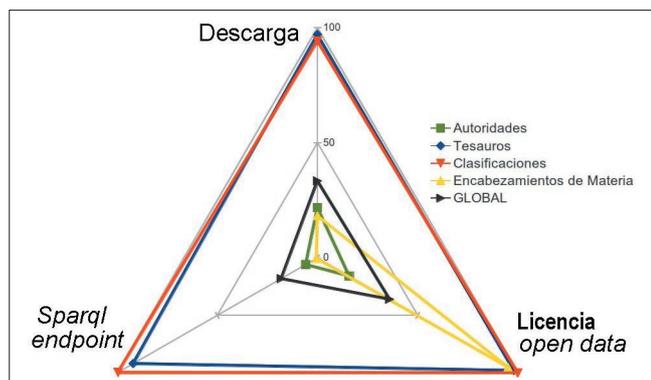


Figura 3. Triángulos *open data* de los cuatro tipos principales de vocabularios controlados

rios como parte de un único conjunto) pudiéndose observar cómo el nivel de integración en el ecosistema *open data* es muy bajo.

Interoperabilidad de vocabularios

Es otro de los aspectos que constituye un indicador de la calidad de un vocabulario (Mader, 2010). A mayor interconexión, mayor integración y cohesión entre conjuntos de datos. En consecuencia, se incrementan sus posibilidades de uso en sistemas de recuperación de información (SRI) basados en técnicas de la web semántica (Angjeli et al., 2009). Esta aplicación y reutilización inmediata de los vocabularios disponibles en SKOS contrasta enormemente con sus equivalentes ediciones en papel, que en ocasiones han tenido una aplicación limitada y obviamente poseen una interoperabilidad limitada o nula. En SKOS la interoperabilidad se centra en la declaración explícita de relaciones de mapeado entre elementos de diferentes vocabularios. Por lo tanto es un camino de ida y/o vuelta: hay que considerar las relaciones de mapeado que parten desde un vocabulario o que llegan a él. Por lo tanto, si en la sección anterior se abordó en qué medida los vocabularios son “open”, en esta se tratará hasta qué punto podemos hablar de “linked”.

Las listas de encabezamiento de materias son las que mayor visibilidad y nivel de interoperabilidad aportan en el conjunto de sistemas de organización del conocimiento representados mediante SKOS

Tras la recopilación de datos se han hallado 9,22 millones de relaciones: 7,75 millones (más de un 84%) corresponden a relaciones de mapeado entre vocabularios, y 1,47 millones son vínculos con otros conjuntos de datos. En esta investigación se han tratado únicamente las relaciones entre vocabularios para estudiar el nivel de interoperabilidad. Se han definido dos indicadores a los que hemos denominado *visibilidad* y *luminosidad*, utilizando la misma terminología de los estudios de la estructura de enlaces web (Madria et al., 1999). La visibilidad se asocia a las relaciones de mapeado que un vocabulario recibe por parte de otros. Por su parte

la luminosidad se refiere a las relaciones que dentro de un vocabulario se definen hacia otros vocabularios externos. El número de relaciones de mapeado se ha ponderado en función del volumen (número de tripletas) de conjunto de datos del vocabulario.

Sobre el conjunto de tripletas X_i que representa un vocabulario, y cuyo volumen se representa mediante su cardinalidad $card(X_i)$, se definen dos subconjuntos que representan las relaciones de mapeado: $R_e(X_i)$ que contiene las relaciones que el conjunto X_i recibe de otros vocabularios, y $R_s(X_i)$ con el número de relaciones de mapeado que el mismo conjunto establece con otros vocabularios externos. El número de relaciones de cada uno de los subconjuntos viene dado por las correspondientes cardinalidades $card(R_e(X_i))$ y $card(R_s(X_i))$. En consecuencia, se definen visibilidad y luminosidad de X_i , $VIS(X_i)$ y $LUM(X_i)$ respectivamente como medidas porcentuales:

$$VIS(X_i) = \frac{card(R_e(X_i))}{card(X_i)} \cdot 100$$

y

$$LUM(X_i) = \frac{card(R_s(X_i))}{card(X_i)} \cdot 100$$

También se ha definido una medida de agregación que permite estimar globalmente tanto la visibilidad como la luminosidad, aunando ambos indicadores. Esta medida a la que se ha denominado *Interoperabilidad combinada*, $IC(X_i)$, se define como:

$$IC(X_i) = \frac{card(R_e(X_i)) + card(R_s(X_i))}{card(X_i)} \cdot 100$$

equivalente a

$$IC(X_i) = VIS(X_i) + LUM(X_i)$$

Del estudio de interoperabilidad se han eliminado los vocabularios “aislados”: aquellos que ni reciben, ni emiten relaciones de mapeado y para los que $IC(X_i)$ es igual a cero. La tabla 3 muestra de forma detallada y agrupada (por tipo de vocabulario) los resultados de los cálculos de luminosidad, visibilidad e interoperabilidad combinada para 31 vocabularios de los 55 que han formado parte del estudio.

Los resultados indican que las listas de encabezamientos de materia poseen un alto grado de interoperabilidad en su conjunto, debido principalmente a la alta visibilidad que tiene la LCSH, seguidas a gran distancia por el fichero de control de autoridades GND de la Biblioteca Nacional de Alemania. Los tesauros cuentan con una escasa interoperabilidad global, puesto que los valores de luminosidad y visibilidad alcanzados por el tesoro con mayor volumen de datos (Gemet) son muy bajos. La figura 4 muestra los mismos resultados de un modo más comprimido, ordenados según la medida de interoperabilidad combinada de cada vocabulario individual y tipo de vocabulario.

Vocabulario	Tipo	Abreviatura	LUM(X)	VIS(X)	IC(X)
<i>TaxonConcept knowledge base</i>	Taxonomía	<i>Taxon</i>	0,0001	0,0000	0,0001
<i>Upper mapping and binding exchange layer</i>	Ontología	<i>Umbel</i>	0,0000	0,0019	0,0019
<i>National diet library of Japan authorities</i>	Autoridades	<i>NDLNA</i>	0,0481	0,0000	0,0481
<i>General multilingual enviromental thesaurus</i>	Tesauro	<i>Gemet</i>	0,0231	0,0284	0,0515
<i>Dewey decimal classification</i>	Clasificación	<i>DDC</i>	0,0000	0,1017	0,1017
<i>Common procurement vocabulary 2003</i>	Clasificación	<i>CPV2003</i>	0,0000	0,1831	0,1831
<i>EU's Multilingual thesaurus</i>	Tesauro	<i>Eurovoc</i>	0,0000	0,2358	0,2358
<i>Linked clean energy data thesaurus</i>	Tesauro	<i>Reegle</i>	0,3840	0,0000	0,3840
Global de los tesauros		G-T	0,2720	0,1538	0,4258
<i>Thesaurus of the FAO</i>	Tesauro	<i>Agrovoc</i>	0,4151	0,1189	0,5340
<i>Umwelt Thesaurus</i>	Tesauro	<i>Umthes</i>	1,1610	1,1610	2,3220
<i>The Virtual international authority file</i>	Autoridades	<i>VIAF</i>	2,0000	0,8948	2,8948
<i>Polythematic structured subject heading system</i>	Enc. materia	<i>Psshs</i>	3,0000	0,0000	3,0000
<i>National Agricultural Library thesaurus</i>	Tesauro	<i>NALT</i>	0,0000	3,6685	3,6685
Global de ficheros de control de autoridades		G-A	2,2964	2,4258	4,7222
<i>Environmental applications reference thesaurus</i>	Tesauro	<i>EARTh</i>	5,0000	0,0000	5,0000
<i>Combined nomenclature 2012</i>	Clasificación	<i>CN2012</i>	5,0035	0,0000	5,0035
<i>Faceted application of subject terminology</i>	Enc. materia	<i>FAST</i>	5,3333	0,0000	5,3333
Global de clasificaciones		G-C	2,4164	3,0215	5,4380
<i>TheSoz thesaurus for the social sciences</i>	Tesauro	<i>Gesis</i>	3,1584	2,8153	5,9737
<i>Common procurement vocabulary 2008</i>	Clasificación	<i>CPV2008</i>	0,1245	6,4135	6,5379
<i>Thesaurus W for local archives</i>	Tesauro	<i>TWLA</i>	9,5273	0,0000	9,5273
<i>Répertoire d'autorité matière encyclopédique et alphabétique unifié</i>	Enc. materia	<i>Rameau</i>	4,9817	5,7309	10,7126
<i>International standard industrial classification V4</i>	Clasificación	<i>Isicv4</i>	12,0299	0,0000	12,0299
<i>Central product classification 2008</i>	Clasificación	<i>CPC2008</i>	12,8706	0,0000	12,8706
<i>Listas de encabezamientos de materia de las bibliotecas públicas</i>	Enc. materia	<i>LEM</i>	13,7789	0,0000	13,7789
<i>Gemeinsame normdatei</i>	Autoridades	<i>GND</i>	4,6109	10,0458	14,6567
<i>Statistical classification of products by activity 2008</i>	Clasificación	<i>CPA2008</i>	17,3080	0,0000	17,3080
<i>North American industry classification system 2012</i>	Clasificación	<i>Naics2012</i>	18,3526	0,0000	18,3526
<i>Thesaurus for graphic materials</i>	Tesauro	<i>T4GM</i>	18,7379	0,0000	18,7379
<i>North American industry classification system 2017</i>	Clasificación	<i>Naics2007</i>	18,8278	0,0000	18,8278
<i>Standardthesaurus wirtschaft</i>	Tesauro	<i>STW</i>	14,9138	5,2267	20,1405
<i>MARC codes list</i>	Glosario	<i>MarcCL</i>	25,3176	0,0000	25,3176
Global de listas de encabezamientos de materia		G-E	4,8894	31,6983	36,5877
<i>Library of Congress subject headings</i>	Enc. materia	<i>LCSH</i>	1,3316	41,8306	43,1621

Tabla 3. Cálculos de visibilidad, luminosidad e interoperabilidad combinada de los vocabularios individuales y de los tipos globales

Se observa que en la mayoría de los casos la luminosidad es determinante en el cálculo de la interoperabilidad de los vocabularios: muchos de ellos con valores de cero en la visibilidad alcanzan valores de $IC(x)$ que superan el 10%. En este sentido cabe señalar el alto valor de luminosidad e interoperabilidad combinada de *MarcCL* únicamente con un volumen de 8.816 tripletas. Una minoría de vocabularios se sitúan en el caso contrario: su alta visibilidad les confiere un alto grado de interoperabilidad. Tal es el caso de *GND* y *LCSH*, en especial este último cuya visibilidad es de un 41,8%, lo que permite afirmar que actualmente es el vocabulario de referencia en el entorno de *linked open data*.

Conclusiones

El catálogo *TDH* precisa una exhaustiva revisión de los da-

“ *The data hub* resulta insuficiente para un control de los conjuntos de datos RDF: es esencial difundir el uso de ficheros *VOID* que permiten describir más detalladamente su contenido ”

tos que recoge sobre los diferentes vocabularios controlados que utilizan *SKOS*. Esta tarea podría realizarse por los propios editores de los vocabularios, conscientes de la necesidad de proporcionar una información descriptiva más homogénea y rigurosa. Los diferentes conjuntos de datos de vocabularios analizados no incorporan ficheros *VOID*¹³, lo que hubiera permitido analizar detalladamente su contenido. Los catálogos colaborativos deberían registrar y

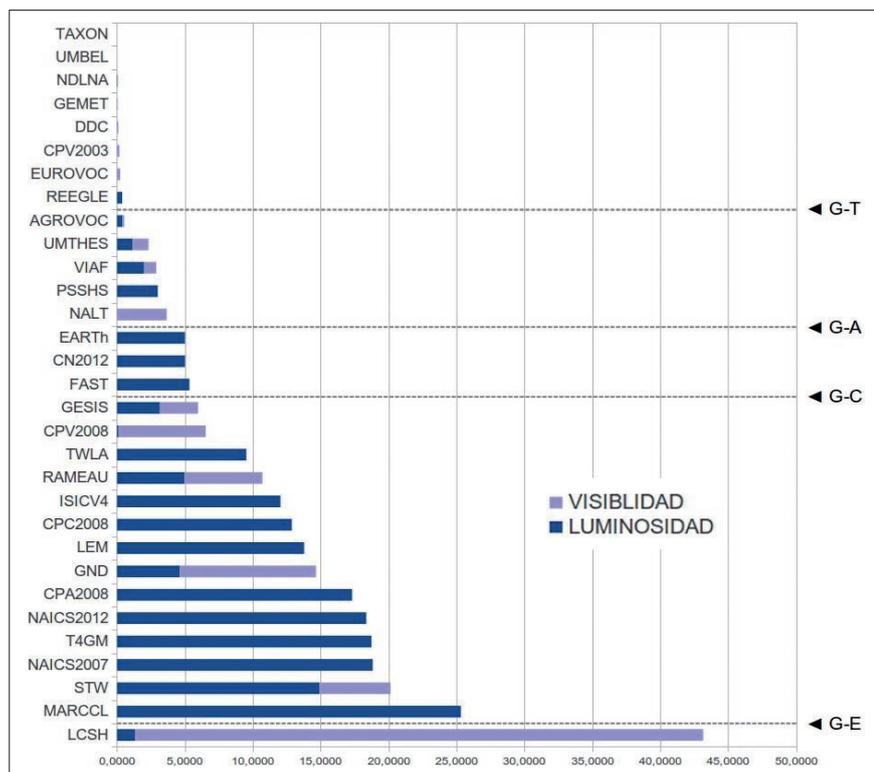


Figura 4. Interoperabilidad combinada de los vocabularios como acumulado de las medidas de visibilidad y luminosidad, y posición de la interoperabilidad global de los principales tipos de vocabularios

procesar ficheros *VOID*, sin que sea preciso registrar otros datos manualmente, puesto que la actualización de datos se realizaría inmediatamente tras un análisis automático de dichos ficheros.

La aplicación de *SKOS* en el ámbito de *linked open data* resulta muy heterogénea en función del criterio analizado y el tipo de vocabulario observado. El grado de interoperabilidad es relativamente alto en el campo de las listas de encabezamientos de materia, lo que hace pensar en un mayor nivel de penetración de *SKOS* en el sector de las bibliotecas y en los procesos de gestión y catalogación bibliográfica, que tienen como referente a los *Encabezamientos de materia* de la *Biblioteca del Congreso*. Destaca también el bajo grado de interconexión de los tesauros. El enriquecimiento de las propiedades de mapeado entre conceptos de tesauros como *Agrovoc*, *Eurovoc* o *Gemet* abriría nuevas posibilidades de reutilización de estos vocabularios y su aplicación podría conllevar un aumento de la eficiencia en SRI (no hay que olvidar que la nueva norma sobre tesauros *ISO-25964* apunta en este sentido).

Desde el punto de vista del acceso abierto a los datos, son los tesauros y las clasificaciones los que han obtenido mejores resultados, alcanzando un nivel prácticamente perfecto en todos los criterios establecidos en el estudio. Los encabezamientos de materia y ficheros de control de autoridades son muy deficientes en este aspecto, especialmente en lo referente a la existencia de *Sparql endpoints*, algo que resulta paradójico, especialmente en los encabezamientos de materia, que pese a su alto grado de interoperabilidad suelen carecer de mecanismos para su reutilización e integración. Esto puede limitar su uso, reduciendo su aplicación a las instituciones que editan dichos vocabularios.

Finalmente cabe indicar la existencia de numerosos conjuntos de datos no incorporados en este estudio, que hacen un uso incorrecto de *SKOS*. No es infrecuente encontrar conjuntos de datos que usan una única propiedad como *skos:prefLabel*, sin incorporar otro tipo de propiedades para el etiquetado de conceptos o el establecimiento de relaciones semánticas o de mapeado. Esto indica un exceso de “*skosificación*” de conjuntos de datos RDF, lo que resulta inapropiado y desaconsejable, más si cabe cuando existen otros esquemas RDF (incluso propiedades del propio RDF) que realizan esta función. La aplicación de *SKOS* debería reservarse a la publicación de sistemas de organización del conocimiento, poniendo especial énfasis en la definición de relaciones de mapeado entre diferentes vocabularios, algo que sin duda les aportará un gran valor añadido.

Notas

1. Puede encontrarse una amplia información y referencia sobre *SKOS* en: <http://www.w3.org/2004/02/skos/intro>
2. RDF: modelo marco para la descripción de recursos. Se basa en tripletes del tipo sujeto-objeto-predicado que conforman grafos complejos.
3. *Sparql* es un lenguaje de consulta para interrogar y recuperar datos RDF.
4. OWL: Lenguaje para la definición de ontologías, utilizado junto con RDF para la descripción de aspectos lógicos de las relaciones entre recursos.
5. <http://thedatahub.org>
6. Servicios web que permiten el uso de *Sparql* para realizar consultas sobre conjuntos de datos RDF y recuperarlos.
7. La búsqueda en el catálogo se realizó manejando directamente el volcado de datos de *The data hub* disponible en formato *JSON* en: <http://thedatahub.org/dump>
8. <http://www.w3.org/2001/sw/wiki/SKOS/Datasets>
9. Existe una versión de la *CDU* en *SKOS* disponible en: <http://www.udcc.org/udcsummary/exports.htm>
Sin embargo se trata de una versión muy incompleta por lo que no se ha incluido en este trabajo.
10. <http://www.w3.org/RDF/Validator>
11. Los datos obtenidos para la realización de este trabajo están disponibles para su descarga en: <http://skos.um.es/files/skos-lod-2012.ods>
12. La cardinalidad de un conjunto es el número de elementos que contiene.

13. *VoID: Vocabulary of intelinked datasets*. Esquema RDF que permite expresar diferentes tipos de metadatos sobre conjuntos de datos RDF.
<http://vocab.deri.ie/void>

Bibliografía

Alexander, Keith; Cyganiak, Richard; Hausenblas, Michael; Zhao, Jun. "Describing linked datasets: on the design and usage of *VoID*, the *vocabulary of interlinked datasets*. En: *Linked data on the web workshop (LDOW 09)*, 2009.
http://events.linkedata.org/ldow2009/papers/ldow2009_paper20.pdf

Angjeli, Anila; Isaac, Antoine; Cloarec, Thierry; Martin, Frédéric; Van-der-Meij, Lourens; Mattheizing, Henk; Schlobach, Stefan. "Semantic web and vocabulary interoperability: an experiment with illumination collections". En: *IFLA International cataloguing and bibliographic control*, April-June 2009, pp. 25-29.

Francesconi, Enrico; Faro, Sebastiano; Marinai, Elisabetta; Peruginelli, Ginevra. "A methodological framework for thesaurus semantic interoperability". En: *Fifth European semantic web conf.*, 2008, pp. 76-87.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.140.4905&rep=rep1&type=pdf>

Isaac, Antoine; Waites, William; Young, Jeff; Zeng, Marcia. *Library Linked Data Incubator Group: Datasets, value vocabularies, and metadata element sets*. W3C Incubator Group Report, 25 Oct. 2011.
<http://www.w3.org/2005/Incubator/ld/XGR-ld-vocabdata-set-20111025>
<http://skos.um.es/Incubator/ld/XGR-ld-vocabdataset>

ISO. *ISO 25964-1:2011. Thesauri and interoperability with*

other vocabularies. Part 1: Thesauri for information retrieval. ISO, 2011.

ISO. *ISO/DIS 25964-1:2011. Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies*. ISO, 2011.

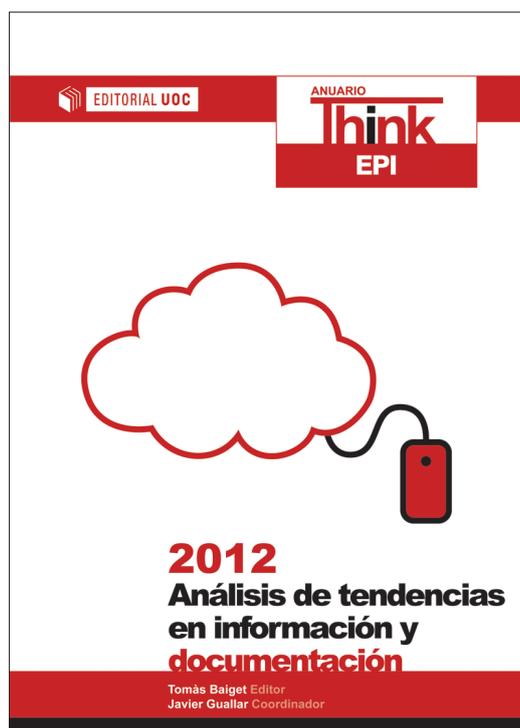
Mader, Christian. "Quality assurance in collaboratively created Web vocabularies". En: *European semantic web conf.*, 2010.
http://eprints.cs.univie.ac.at/3355/4/eswc_proposal_short.pdf

Madria, Sanjay; Bhowmick, Sourav S.; Ng, Wee-Keong; Lim, Ee-Peng. "Research issues in Web data mining". En: *Procs of the 1st Intl conf on data warehousing and knowledge discovery*, 1999, pp. 303-312.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.4020&rep=rep1&type=pdf>

Miles, Alistair; Bechhofer, Sean. *SKOS Simple knowledge organization system reference*. W3C Recommendation, 18 August 2009.
<http://www.w3.org/TR/skos-reference>

Möller, Knud; Hausenblas, Michael; Cyganiak, Richard; Grimnes, Gunnar-Astrand. "Learning from linked open data usage: patterns & metrics". En: *Procs of the WebSci10: Extending the frontiers of society on-line*, 2010.
<http://richard.cyganiak.de/2008/papers/lod-usage-web-sci2010.pdf>

Pastor-Sánchez, Juan-Antonio; Martínez-Méndez, Francisco-Javier; Rodríguez-Muñoz, José-Vicente. "Advantages of thesaurus representation using the *simple knowledge organization system (SKOS)* compared with proposed alternatives". *Information research*, 2009, v. 14, n. 4, paper 422.
<http://InformationR.net/ir/14-4/paper422.html>



Ya ha salido el nuevo

Anuario ThinkEPI 2012

Información y adquisición en:

<http://www.thinkepi.net/anuario-thinkepi/anuario-thinkepi-2012>

y

<http://www.editorialuoc.cat/anuariothinkepi2012-p-985.html>



EVALUACIÓN DE SITIOS WEB MULTILINGÜES: METODOLOGÍA Y HERRAMIENTA HEURÍSTICA



Mar Andreu-Vall y Mari-Carmen Marcos



Mar Andreu-Vall es licenciada en traducción e interpretación por la *Universidad Autónoma de Barcelona* (2004) y titulada en el *Máster en Gestión de Contenidos Digitales* por la *Universidad de Barcelona* (2011). Desde 2004 ha trabajado como traductora autónoma. Actualmente es lingüista y coordinadora de proyectos de traducción web en *AutomaticTrans*, empresa que ofrece servicios y soluciones para proyectos lingüísticos multilingües.

AutomaticTrans SL
Ronda Sant Pere, 13. 08010 Barcelona, España
mmandreu@automatictrans.es



Mari-Carmen Marcos es profesora de la *Universidad Pompeu Fabra* de Barcelona, subdirectora del *Máster Online en Documentación Digital* y codirectora del *Máster en Buscadores*. También es profesora del *Máster en Gestión de Contenidos UB-UPF*. Es doctora y licenciada en documentación. Su línea de investigación se centra en la evaluación de interfaces como medio para la interacción entre las personas y la tecnología.

Universitat Pompeu Fabra, Departamento de Comunicación
Roc Boronat, 138. 08018 Barcelona, España
<http://www.mcmarcos.com>
mcarmen.marcos@upf.edu

Resumen

Se propone una metodología de evaluación de sitios web dirigidos a un público multilingüe y multicultural. Para ello, se han revisado los estándares y recomendaciones de diversas fuentes validadas, se han creado unas pautas generales (heurísticas) y se han desarrollado indicadores detallados para ellas. El resultado es una *checklist* para la evaluación de sitios web con público internacional, tanto monolingües (con o sin internacionalización) como multilingües (con o sin localización).

Palabras clave

Usabilidad, Evaluación heurística, Localización, Internacionalización, Multilingüismo.

Title: Multilingual websites evaluation: methodology and heuristic tool

Abstract

A methodology for evaluating the suitability of websites to multilingual and multicultural audiences is proposed. Standards and validated recommendations have been checked in order to create a (heuristic) guideline. The result is a checklist for evaluating websites with international audiences, both monolingual (with or without globalization) and multilingual (with or without localization).

Keywords

Usability, Heuristic evaluation, Localization, Internationalization, Multilinguism.

Andreu-Vall, Mar; Marcos, Mari-Carmen. "Evaluación de sitios web multilingües: metodología y herramienta heurística". *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 254-260.

<http://dx.doi.org/10.3145/epi.2012.may.05>

1. Introducción

Las transformaciones relacionadas con el proceso de globalización de la sociedad y la economía han puesto de relieve la gran importancia estratégica que han adquirido las lenguas

en la cultura, la cohesión social, la economía y el desarrollo de las sociedades. En este sentido, la presencia de las diferentes lenguas en la Web va adquiriendo progresivamente más peso, y a pesar de la predominancia del inglés y el chino, el multilingüismo permite que más usuarios puedan

Nota: Una versión preliminar de este estudio fue presentada en el congreso *Interacción 2011*. Puede consultarse en: **Andreu, Mar; Marcos, Mari-Carmen.** "Multilingüismo en la web. Pautas para un diseño centrado en el usuario". *Interacción 2011: XII Congreso internacional de interacción persona-ordenador*, Lisboa, 2-5 sept. 2011. ISBN: 978 84 9281 234 9

Artículo recibido el 23-01-12

Aceptación definitiva: 07-05-12

acceder a contenidos web en su lengua. En este panorama hay dos conceptos que merece la pena aclarar: internacionalización y localización.

La internacionalización (i18n)¹ es un concepto que procede del área de la programación y que consiste en que el código esté preparado para poder desarrollar una aplicación para diversos idiomas. Aplicado al diseño web, se trata del desarrollo de un sitio o aplicación web que permita una fácil adaptación lingüística (localización) cuando va dirigido a un público con una lengua o cultura diferentes al idioma para el que se ha creado originalmente. Por otra parte, la localización (l10n) es el proceso de adaptación de un producto para adecuarlo a las necesidades, tanto lingüísticas como culturales, de un público destinatario concreto llamado público local. La localización engloba aspectos más complejos que la simple traducción del contenido, ya que incluye la adaptación de elementos como formatos numéricos, de fecha y de hora, etc.

En este artículo proponemos una metodología de evaluación de la adecuación de sitios web a un público multilingüe y multicultural. Para ello se han revisado los estándares y recomendaciones de diversas fuentes validadas, se han creado unas pautas generales (heurísticas) y se han desarrollado indicadores detallados para ellas. El resultado es un *checklist* para la evaluación de sitios web con público internacional, tanto monolingües (con o sin internacionalización) como multilingües (con o sin localización).

En la sección *head* del código debe incluirse una etiqueta *meta* de contenido en la que se indique que la codificación será de tipo UTF-8

2. Metodología: la evaluación heurística

Existen muchos criterios y puntos de vista desde los que analizar y evaluar un sitio web. En este estudio hemos utilizado una de las técnicas de evaluación centradas en el usuario denominada evaluación heurística. Consiste en que un grupo de expertos valore si los elementos de la interfaz del usuario de un sitio web o aplicación se corresponden con los principios de usabilidad establecidos a partir de una lista de principios o heurísticas. Esta técnica consigue detectar un gran número de problemas de usabilidad en sitios web con un menor coste que las técnicas que involucran usuarios (Nielsen, 1994).

Así como para la evaluación general de sitios web existen criterios muy establecidos y probados que proporcionan la seguridad de que se está evaluando todo lo que es necesario evaluar (como los de Nielsen y Molich, 1990, simplificados más tarde por Nielsen y Mack, 1994), estas pautas no cubren aspectos específicos, como el que ahora nos ocupa, que es el multilingüismo.

La evaluación que proponemos no pretende sustituir a la heurística tradicionalmente usada en sitios web, sino que es un complemento para aquellos sitios que sí son multilingües.

Para su elaboración se han tomado como punto de partida los principios heurísticos de Nielsen y la adaptación realizada posteriormente por Hassan y Martín (2003) y se han incorporado los estándares de localización y usabilidad de otros autores (Del Galdo, Nielsen, 1996; Nielsen, 2000; Hillier, 2002), además de las pautas del *World Wide Web Consortium (W3C)*.

El resultado es la guía de evaluación heurística para sitios web multilingües presentada en la sección 3.

No adecuar el formato de fecha puede provocar confusión a la hora de interpretar la información

3. Pautas para la evaluación de la localización en sitios web multilingües

Presentamos a continuación las 24 pautas definidas para la evaluación de un sitio web, estructuradas en 5 categorías. Para cada pauta se indica el impacto que tiene su falta de cumplimiento (1 = leve, es molesto pero no impide realizar la tarea que el usuario pretende; 2 = medio, dificulta la tarea; 3 = grave, impide realizar la tarea).

Grupo 1. Aspectos relacionados con el código fuente

Esta categoría evalúa los elementos del código html y su adecuación a las necesidades de la internacionalización. Se han tomado como fuente las siguientes pautas del W3C:

- *Guía breve de internacionalización*
- *An introduction to multilingual web addresses*
- *Etiquetas de idioma en html y xml*
- *El tamaño del texto en la traducción*
- *Internationalization best practices: Specifying language in xhtml & html content*

Pautas:

1.1. Codificación *unicode* en la web (*charset UTF-8*)

En la sección *head* del código debe incluirse una etiqueta *meta* de contenido en la que se indique que la codificación será de tipo UTF-8 (*8-bit Unicode transformation format*). Así el navegador interpretará los caracteres de la mayoría de los alfabetos. Un ejemplo:

```
<meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
```

Impacto: grave (3)

1.2. Coherencia de la lengua de las urls de las diferentes versiones con la lengua en la que está redactado el contenido

Las palabras que se usen como parte de la url deben estar en el mismo idioma de la página a la que esa url dirige. Por lo tanto, un sitio web que tiene una página en 3 idiomas deberá adaptar también los nombres de las categorías, subcategorías y páginas para que estén traducidas a los 3 idiomas (figura 1).

Impacto: medio (2)



Figura 1. Ejemplo de urls traducidas

1.3. Declaración del idioma del contenido en el encabezamiento de cada página

Al inicio del documento debe indicarse el idioma en el que está escrita la página. En caso de ser una variante del idioma para un país en concreto, se añade el país. El código sería así para español de España: `<html lang="es">`

Impacto: grave (3)

1.4. Cambios de idioma del contenido dentro de una misma página

En caso de que alguna palabra, frase o párrafo dentro de una página web esté en un idioma distinto al del resto de la página, deberá indicarse en el código. Por ejemplo, si es una página en español y comienza un párrafo en francés, se indicará con la etiqueta de párrafo seguida del atributo de lenguaje: `<p lang="fr">`

Impacto: grave (3)

1.5. Expansión del texto

Algunos idiomas se caracterizan por tener palabras muy largas. El maorí es conocido por ese rasgo, por ejemplo puede haber una palabra de 85 caracteres, pero sin ir más lejos, el alemán o el polaco son idiomas con palabras largas, lo que hace que necesiten más espacio en la página web. Por eso, al diseñar el sitio web hay que prever la posible expansión que tendrán las palabras al cambiar de lengua. Eso implica que el diseño debe permitir que los espacios sean flexibles, no cerrados, pues se puede encontrar que palabras muy cortas en inglés son mucho más largas en la mayoría de las lenguas (figura 2).

Impacto: leve (1)

1.6. Cambio de lengua del nodo de los enlaces

Es común incluir enlaces a páginas web que se encuentran en un idioma distinto al de la página en la que está el link. La buena práctica es indicar el cambio de idioma en el código en la misma etiqueta del enlace (`<a>`) con el atributo `hreflang`. Por ejemplo, en una página en español que contenga un enlace a una página en inglés, el código sería así:

```
<a href=http://www.otrapagina.com/ hreflang="en">Otra página </a>
```

Impacto: leve (1)

El diseño visual usado en la versión dirigida a un país puede que tenga que variar con respecto al que se use para público de otros países

Grupo 2. Aspectos relativos a la localización lingüística

Esta categoría tiene como objetivo valorar la adecuación de los diferentes elementos de la web a las lenguas y culturas de los usuarios finales. Se han tomado como base dos do-



Figura 2. Ejemplo de dos versiones lingüísticas de Wikipedia con distinta longitud de palabras en el menú



Figura 3. Ejemplo de cambio del formato de fecha en dos idiomas

cumentos del W3C para webs multilingües: *Formatos de fechas y Fechas y horarios*, y dos artículos (Marcus; Armitage; Frank, 1999) (Nielsen, 1996)

Pautas:

2.1. Formato de fecha

Debe considerarse cuál es la forma más común en el país al que se dirige la información, por ejemplo día-mes-año en España, pero año-mes-día en China. No adecuar el formato de fecha puede provocar confusión a la hora de interpretar la información. Un ejemplo sería la fecha de reserva de un vuelo, en la que el usuario quiera comprar un billete para el 1 de marzo y el sistema haga la reserva para el 3 de enero sin que el usuario se percate de que 01/03 no es lo que él cree que es (figura 3).

Impacto: grave (3)

2.2. Formato de hora

Como ocurre con la fecha, la hora también cambia entre países. En unas culturas se utiliza la hora en formato de reloj de 12 horas mientras que en otras está más aceptado usar el reloj de 24 horas, de manera que en unos lugares es más común encontrar las 5PM y en otros las 17:00. Siguiendo con el ejemplo del vuelo, un usuario acostumbrado al sistema de 24 horas podía confundirse entre las 9 de la noche y las 9 de la mañana si no ve claramente que se refiere a las 21:00.

Impacto: grave (3)

2.3. Formato de número

De forma similar a los dos casos anteriores, la forma de leer la puntuación en las cifras cambia en función de la cultura. En unas lenguas la marca de millares es un punto (en español), en otros una coma (en inglés). Y pasa lo contrario con la marca de decimales, a veces es una coma (en español) y a veces un punto (en inglés). No adaptar la presentación de las cifras a la cultura del usuario puede provocar confusiones.

Impacto: grave (3)

2.4. Sección de contacto

A la hora de diseñar la página de contacto, hay que tener en cuenta si van a llegar a ella usuarios que viven en otros países. En tal caso, hay que proporcionar la información necesaria para contactar desde el extranjero, indicando datos como código de país para el teléfono, o si hay que incluir alguna cifra intermedia entre el código de país y el número local.

Impacto: grave (3)

2.5. Actualización del contenido en las diferentes versiones lingüísticas

Es muy común que un sitio web en varios idiomas no sea tan completo en todos ellos. Puede que uno de los idiomas sea el que se actualiza con más frecuencia, y en cambio otras versiones lingüísticas no cuenten con tanta dedicación para mantenerse al día. Lo ideal, sin duda, es que todas las versiones se actualicen al mismo tiempo.

Impacto: medio (2)

2.6. Acceso desde cualquier página a la opción de elección de idioma

Esta pauta se refiere a un aspecto de usabilidad, que es poder cambiar de idioma en cualquier página sin tener que comenzar a navegar desde la página de inicio al clicar la opción de cambio de lengua. Disponer de esta función bien implementada es tan sencillo como poner el link de la opción de idioma directamente a la página equivalente, y no dejar que siempre apunte a la *home page*. En cambio esto no es posible si el sitio no se ha traducido en tu totalidad; en tal caso, dirigir al usuario a la página de inicio es la solución más adecuada.

Impacto: grave (3)

Es importante traducir la etiqueta meta *description* porque los buscadores la usan a menudo para crear el *snippet* en la lista de resultados

Grupo 3. Aspectos relacionados con los elementos visuales

Teniendo en cuenta que los usuarios de diferentes culturas no reaccionan igual ante ciertos estímulos visuales (Makki; Leppert, 2006), es importante considerar la localización de los elementos visuales que se incluyan en el sitio, especial-

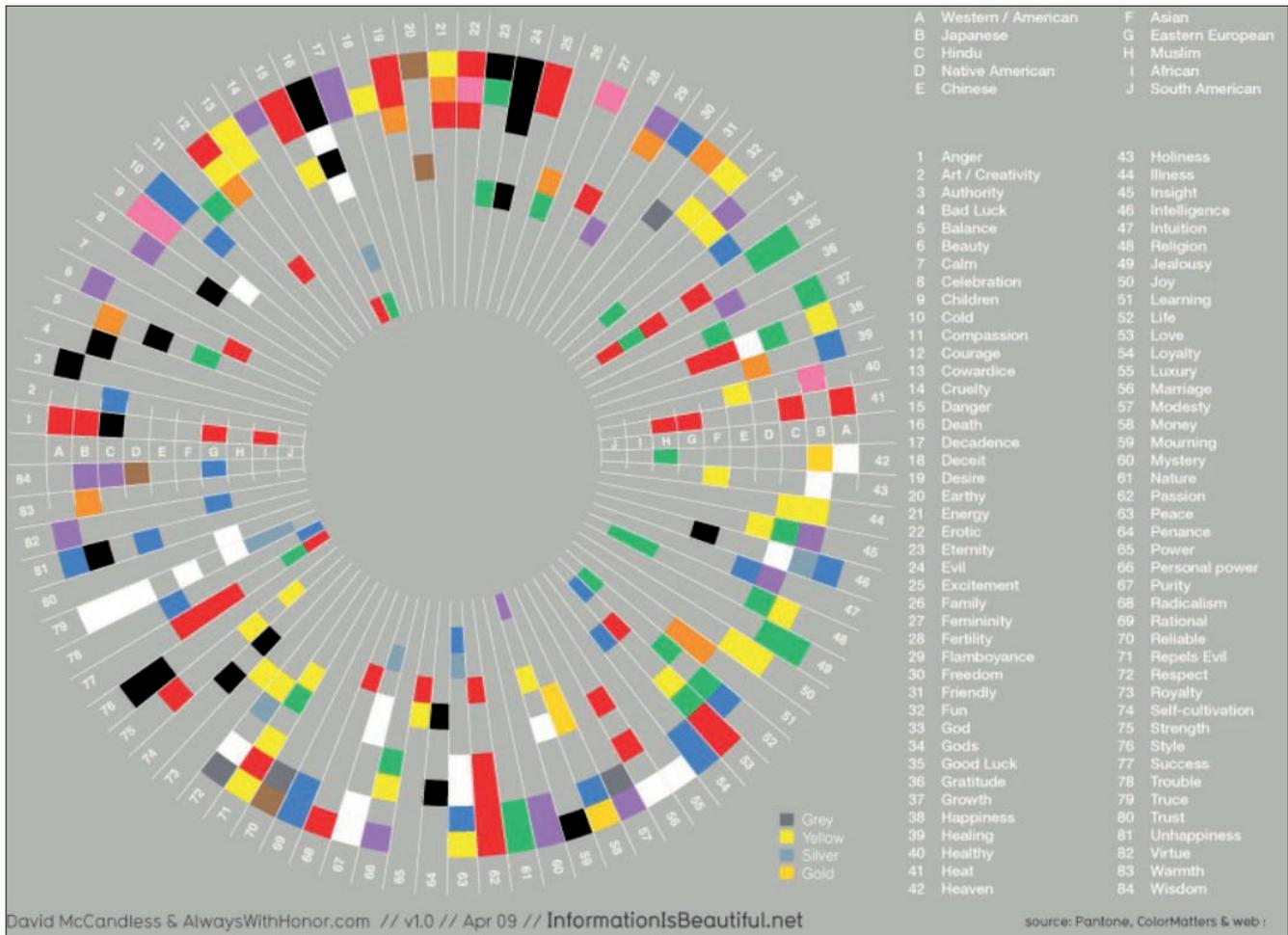


Figura 4. Distinto significado de los colores según las culturas, <http://www.informationisbeautiful.net/visualizations/colours-in-cultures>

mente imágenes, iconos y colores. Para los aspectos relacionados con la direccionalidad se ha tomado como fuente el documento del W3C sobre *Dirección del sistema de escritura e idiomas*.

Pautas:

3.1. Iconos

Los iconos deben adecuarse a cada cultura. Un ejemplo muy manido en usabilidad es el icono del buzón para recibir correo, tan distinto entre el español, normalmente cuadrado y plano, pegado a una pared o a una puerta, y el estadounidense, representado como un poste frente a la entrada de una casa con jardín, con un formato más cilíndrico.

Impacto: medio (2)

3.2. Simbolismo de los colores

Sobre el significado del color hay estudios muy interesantes que demuestran cómo colores que en unas culturas tienen un poso alegre en otros no lo tienen. La portada del libro *Information is beautiful* muestra un gráfico que ilustra muy bien esta idea:

<http://www.informationisbeautiful.net/visualizations/colours-in-cultures>

y nos hace pensar que el diseño visual usado en la versión dirigida a un país puede que tenga que variar con respecto al que se use para público de otros países.

Impacto: medio (2)

3.3. Imágenes

Aunque no suele ser un problema común, hay que tener en cuenta qué imágenes usar en un sitio web por si pueden tener connotaciones negativas u ofensivas para una cultura.

Impacto: medio (2)

3.4. Direccionalidad

La página web debe estar preparada para la lectura de izquierda a derecha y de derecha a izquierda, para los casos de traducciones a lenguas con diferente dirección, por ejemplo de inglés a árabe.

Impacto: grave (3)

“ El instrumento de evaluación heurística está disponible en <http://goo.gl/LJTR5> ”

Grupo 4. Aspectos que afectan a la búsqueda dentro del sitio web

En caso de sitios web grandes, contar con un buscador interno resulta muy útil para localizar información sin tener que navegar. Si el buscador no ha sido implementado considerando que el sitio tiene información en diversos idiomas y que los usuarios pueden hacer las consultas también en varias lenguas, podrán darse problemas en la comprensión

de los resultados obtenidos. En este punto se han seguido las indicaciones de **Jakob Nielsen** (1996).

Pautas:

4.1. Búsqueda en todas las versiones lingüísticas

El buscador debe recoger las páginas web de todo el sitio, no sólo las de un idioma en particular.

Impacto: medio (2)

4.2. Opciones avanzadas de búsqueda relativas al idioma

El usuario debe poder escoger si realiza una consulta en todos los idiomas o en alguno en particular.

Impacto: grave (3)

4.3. Idioma de los textos informativos de la página de resultados

Los textos que sirven de apoyo a los resultados de búsqueda deben aparecer en el idioma en el que el usuario esté consultando el sitio web. Esos textos son los que acompañan el listado y dan indicaciones de cómo mejorar la búsqueda, otras opciones para continuar navegando, etc.

Impacto: medio (2)

Grupo 5. Aspectos que afectan al posicionamiento del sitio web en los buscadores

Muchos sitios multilingües bien localizados no reciben el volumen de tráfico internacional que deberían porque no aparecen en los primeros resultados de las páginas de los buscadores. En esta categoría se evalúa si se han traducido los elementos relevantes para el posicionamiento web. Los indicadores desarrollados toman como fuente el artículo de **Iler** (2006). Es importante tener en cuenta que el posicionamiento de la página en cada uno de los idiomas es independiente, ya que los elementos relevantes de SEO (*search engine optimization*) son diferentes para cada versión lingüística.

Pautas:

5.1. Traducción de la etiqueta *title* (*head*)

Esta etiqueta corresponde al título de la página web, que es aquel que se lee en la parte superior del navegador y también el que identifica a la página web cuando aparece entre los resultados de búsqueda en los buscadores. Que esté traducido es importante para la identificación por parte de los usuarios, pero también para que los robots (*spiders*) de los buscadores puedan indexar ese título en el idioma en el que está escrita la página.

Impacto: grave (3)

5.2. Traducción de las etiquetas *description* y *keywords* (*head*)

Si bien estas etiquetas no son visibles desde el navegador, ya que se encuentran en la sección *head* del código fuente, los robots de los buscadores sí son capaces de leerlas e indexarlas. La etiqueta *description* tiene mucha importancia porque es la que en muchas ocasiones usan los buscadores para crear el resumen que dan en la lista de resultados

(*snippet*). La etiqueta *keywords* en cambio no la utilizan los buscadores, por lo tanto en SEO no es relevante. Aun así, la hemos incluido en las pautas porque puede ser de utilidad disponer de ella para la implementación de un buscador interno del sitio web.

Impacto: grave (3)

5.3. Traducción de las etiquetas *alt* y *title* de las imágenes

Las imágenes no pueden ser indexadas por los buscadores porque no son legibles como texto. En cambio sí pueden indexar los textos vinculados a ellas, como ocurre con las etiquetas *alt* y *title* de las imágenes. Estas etiquetas, que indirectamente benefician al SEO por poder indexar información, son determinantes para que las personas ciegas puedan identificar el contenido de las imágenes con los programas especiales de lectura de páginas web.

Impacto: grave (3)

5.4. Traducción de las etiquetas *title* de los enlaces

De forma similar al punto anterior, la etiqueta *title* que acompaña a los enlaces permite dotar de información semántica complementaria a un elemento de la página web, que en este caso es el enlace a otra página. Esta etiqueta es visible a los ojos de los usuarios al pasar el cursor sobre el enlace, y también es visible a los robots de los buscadores.

Impacto: grave (2)

4. Cómo utilizar esta *checklist* en la evaluación de la localización en sitios web multilingües

A partir de las pautas de la sección anterior, se ha creado un instrumento de evaluación heurística (ver <http://goo.gl/LJTR5>) en forma de tabla compuesto por:

- La lista de indicadores redactados en forma de pregunta.
- Una columna para indicar la frecuencia con la que se incurre en un error (0 = nunca, 1 = a veces, 2 = muchas veces o siempre).
- El impacto asociado a cada indicador (1 = leve, 2 = medio, 3 = alto).
- Una columna donde calcular la severidad del error multiplicando el valor de frecuencia por el valor del impacto del indicador. La severidad indica la prioridad con la que un error debe ser subsanado. Su valor puede ir desde 0 hasta 6. La tabla 1 muestra las posibles combinatorias de frecuencia e impacto y el resultado de severidad para cada combinación.
- La fuente de donde se ha tomado el indicador.

Una vez que se cuenta con la heurística y sus indicadores, el procedimiento para llevar a cabo la evaluación sigue los pasos siguientes:

- Determinar el ámbito de aplicación de la evaluación, si se analizará todo el sitio web o alguna de las secciones.
- Disponer de un equipo de 3 a 5 evaluadores. Cada uno realiza el análisis de forma individual y redacta un informe con los resultados obtenidos.
- Poner en común los informes individuales: se jerarquizan los problemas en base a su severidad y se redacta un informe.

Frecuencia	Impacto	Severidad
0	1	0 = No hay error
	2	0 = No hay error
	3	0 = No hay error
1	1	1 = Error muy leve
	2	2 = Error leve
	3	3 = Error de grado medio
2	1	2 = Error leve
	2	4 = Prioritario
	3	6 = Urgente

Tabla 1. Grados de severidad calculados multiplicando los valores de frecuencia e impacto

Este método puede llevarse a cabo en cualquier momento del proceso de desarrollo de un sitio web o aplicación; sin embargo, se recomienda aplicarlo durante las primeras etapas de la elaboración de un producto, de manera que no se arrastren errores de usabilidad a lo largo del diseño. En caso de que se trate de un sitio web monolingüe que está preparando versiones en otros idiomas, lo ideal será aplicar una evaluación heurística específica para aspectos de localización antes de dar por acabada la traducción y adaptación del sitio. Una vez realizada la evaluación heurística y subsanados los problemas más graves que se han detectado, es aconsejable realizar un test con usuarios para completar la evaluación del sitio web.

5. Nota

1. A veces las palabras “internacionalización” y “localización” se abrevian con los numerónimos “i18n” (primera letra “i”, última letra “n” y en medio 18 letras más) y “l10n”, respectivamente.

6. Referencias

Andreu-Vall, Mar; Marcos, Mari-Carmen. “Multilingüismo en la web. Pautas para un diseño centrado en el usuario”. En: *Interacción*, 2011. <http://hdl.handle.net/10760/16407>

Del-Galdo, Elisa; Nielsen, Jakob. *International user interfaces*. New York: John Wiley & Sons, 1996.

Griho (Grupo de Investigación en Interacción Persona Ordenador e Integración de Datos). “Métodos de evaluación de la usabilidad”. En: *Modelo de proceso de la ingeniería de la usabilidad y de la accesibilidad*. MPIu+a, 2010. <http://www.grihocitools.udl.cat/mpiu/evaluacionusabilidad.htm>

Hassan-Montero, Yusef; Martín-Fernández, Francisco J. “Guía de evaluación heurística de sitios web”, *No sólo usabilidad*, 2003. <http://www.nosolousabilidad.com/articulos/heuristica.htm>

Hillier, Mathew. “The role of cultural context in multilingual website usability”. *Journal of electronic commerce research and applications*, 2002, v. 2, n. 1, pp. 2-14.

[http://dx.doi.org/10.1016/S1567-4223\(03\)00005-X](http://dx.doi.org/10.1016/S1567-4223(03)00005-X)

Iler, Huiping. Maximizing visibility for multilingual web sites, 2006, wintranslation.com http://www.sempo.org/resource/resmgr/Docs/wintranslation_for_SEMPO.pdf

Internet WorldStats. *International world users by language. Top 10 languages*, 2010. <http://www.internetworldstats.com/stats7.htm>

Makki, Kami; Leppert, Greg. “Factors of usability design for multilingual and multicultural websites”. En: *IRI*, 2006, pp. 6-10. <http://dx.doi.org/10.1109/IRI.2006.252379>

Marcus, Aaron; Armitage, John; Frank, Volker. *Globalization of user-interface design for the web. Report of human factors & the web*, 1999. http://www.amanda.com/cms/uploads/media/AMA_GlobalizationUserInterfaceDesignWeb.pdf

Nielsen, Jakob. How to conduct a heuristic evaluation, 1994. http://www.useit.com/papers/heuristic/heuristic_evaluation.html

Nielsen, Jakob. International web usability. Alertbox, 1996. <http://www.useit.com/alertbox/9608.html>

Nielsen, Jakob. “Utilización internacional: atender a una audiencia global”. En: Nielsen, Jakob. *Usabilidad: diseño de sitios web*. Madrid: Pearson Education, 2000.

Nielsen, Jakob; Mack, Robert. (eds.). *Usability inspection methods*. New York: John Wiley & Sons, 1994.

Nielsen, Jakob; Molich, Rolf. “Heuristic evaluation of user interfaces”. En: *CHI*, 1990, pp. 249-256.

W3C. Dirección del sistema de escritura e idiomas. <http://www.w3.org/International/questions/qa-scripts>

W3C. El tamaño del texto en la traducción. <http://www.w3.org/International/articles/article-text-size>

W3C. Etiquetas de idioma en html y xml. <http://www.w3.org/International/articles/language-tags>

W3C. Fechas y horarios. <http://www.w3.org/International/O-time>

W3C. Formatos de fechas. <http://www.w3.org/International/questions/qa-date-format>

W3C. Guía breve de internacionalización. <http://www.w3c.es/divulgacion/guiasbreves/internacionalizacion>

W3C. Identifying the language of a target document. <http://www.w3.org/International/geo/html-tech/tech-lang.html#ri20050128.152033553>

W3C. Internationalization (I18n) activity. <http://www.w3.org/International>

W3C. Introducción a direcciones web plurilingües. <http://www.w3.org/International/articles/idn-and-iri>



CRITERIOS Y HÁBITOS DE CONSUMO DE INFORMACIÓN PERIODÍSTICA ONLINE EN COLOMBIA



Liliana Gutiérrez-Coba, Javier-Andrés Gómez-Díaz, Andrea Salgado-Cardona, Juliana Estrada-Gutiérrez y María-José Ramírez-Méndez



Liliana Gutiérrez-Coba es doctora en ciencias de la información por la *Universidad del País Vasco* (España), con énfasis en la investigación sobre temas relacionados con el periodismo y la opinión pública. Comunicadora social y periodista, egresada de la *Universidad de La Sabana* (Colombia), donde actualmente se desempeña como docente y directora del *Grupo de Investigación en Periodismo (GIP)*.

liliana.gutierrez1@unisabana.edu.co



Javier-Andrés Gómez-Díaz, psicólogo con maestría en investigación psicosocial y énfasis en comportamiento del consumidor, tiene más de ocho años de experiencia docente universitaria e investigativa. Ha sido asesor metodológico y estadístico de varios proyectos de investigación para el *Grupo de Investigación en Periodismo (GIP)*, y del *Observatorio de Medios de la Universidad de la Sabana* (Colombia).

psandresg@gmail.com



Andrea Salgado-Cardona es escritora, periodista y docente, y trabaja en la *Universidad de la Sabana*, en la *Universidad del Rosario* y en el *Taller de Escritores* de la *Universidad Central*. Ha sido reportera en la revista de música *Shock* y ha sido docente en la *Universidad de Texas* en El Paso. En año 2010 recibió mención en el concurso nacional de literatura *Ciudad de Bogotá* con su libro de relatos autobiográficos *La lectora de lágrimas*.

andrea.salgado@unisabana.edu.co



Juliana Estrada-Gutiérrez, comunicadora social y periodista de la *Universidad de La Sabana*, con énfasis en comunicación organizacional y relaciones públicas, es ejecutiva de cuenta en *G2*, agencia *BTL* y se ha desempeñado como coordinadora de talento en *RCN* y como evaluadora de servicio al cliente en *Fallabella*.

juliestrada87@hotmail.com



María-José Ramírez-Méndez es comunicadora social y periodista de la *Universidad de La Sabana*. Se dedica al diseño y ejecución de planes de comunicación como ejecutiva de cuenta en *Soul ADE Toro Mora* y de *Mediática Comunicaciones*.

majoramirez@hotmail.com

Resumen

Partiendo de una investigación financiada por la *Universidad de La Sabana* sobre calidad informativa y consumo de periódicos en internet, se exploran los hábitos y criterios de consumo de los lectores colombianos de noticias en internet. Con base en las teorías de difusión de la innovación (Rogers, 1995) y de usos y gratificaciones (Katz, 1959) se encontró que, aunque el 64,6% de los consumidores colombianos de diarios online dice preferir un medio que le permita personalizar la manera en que consulta la información, la importancia de elementos relacionados con la interactividad y la personalización no es todavía claramente percibida por los mismos.

Artículo recibido el 01-09-11

Aceptación definitiva: 31-12-11

Palabras clave

Hábitos de consumo periodístico, Páginas informativas, Consumidores de noticias en línea, Internet, Diseño web.

Title: The consumption criteria and habits of online news consumers in Colombia

Abstract

Based on a study funded by *Universidad de la Sabana*, this article describes the information quality and consumption of internet newspapers in Colombia. Applying the theories of the diffusion of innovations (Rogers, 1995) and usability and gratification (Katz, 1959) we found that, even though 64.6% of Colombian online news consumers say that they prefer media that allows them to personalize the way they search for information, the importance of a set of elements related to interactivity and personalization are not clearly perceived by those readers.

Keywords

Journalism consumption habits, News websites, Online news consumers, Internet, Web page design.

Gutiérrez-Coba, Liliana; Gómez-Díaz, Javier-Andrés; Salgado-Cardona, Andrea; Estrada-Gutiérrez, Juliana; Ramírez-Méndez, María-José. "Criterios y hábitos de consumo de información periodística online en Colombia". *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 261-267.

<http://dx.doi.org/10.3145/epi.2012.may.06>

1. Introducción

Gracias a las facilidades de acceso, la rapidez, la actualización constante y otras propiedades del entorno digital, el consumo de noticias por internet aumenta de manera acelerada en el mundo entero. **Rodríguez-Martínez, Codina y Pedraza-Jiménez** (2010) afirman que, entre los contenidos que pueden encontrarse en la Web, las noticias ocupan un lugar cada vez más destacado. El reciente estudio de la *Asociación para la Investigación de Medios de Comunicación* en España da cuenta de que un 32% de los internautas lee en el mismo día diarios en soporte papel e internet y esa cifra se eleva al 59% si se toman los últimos 30 días (AIMC, 2011). La encuesta realizada en Estados Unidos por el *Pew Research Center for the People and the Press* (2011), indica que cada vez más personas citan internet como su principal fuente de noticias, lo que refleja el crecimiento de la Red y el declive gradual de los lectores de periódicos: 34% en el 2007 y 31% en el 2010. El ejemplo del legendario periódico londinense *The guardian*, que disminuirá sus páginas impresas y dará prioridad a su edición digital, es un indicador de la importancia que ha adquirido la internet como opción informativa (**Oppenheimer**, 2011).

En América Latina ha sucedido algo similar. Para el 2010 la región superaba los 200 millones de usuarios de internet, de manera que en los últimos 5 años el crecimiento acumulado era de 143,8%, según señalaba el estudio *Usos de internet en Latinoamérica*, realizado en 15 países de la región (Tendencias digitales, 2010). El estudio encontró que uno de los usos más frecuentes que los latinos le dan a internet es la lectura de noticias: 60,1%.

Lo anterior da cuenta del desarrollo acelerado del ciberperiodismo, entendido como "la especialidad del periodismo que emplea el ciberespacio para investigar, producir y, sobre todo, difundir contenidos periodísticos" (**Salaverría**, 2005, p. 21).

El consumo de noticias online está supeditado a los hábitos del lector. Los hábitos son comportamientos simples o com-

plejos, que se aprenden a través de la repetición (**Martin**, 2008) y suelen estar asociados a las preferencias.

La mayoría de personas tienden a seguir patrones relativamente estables cuando consumen noticias online (**LaRose; Eastin**, 2004; **Diddi; LaRose**, 2006). El aspecto auto-reforzante de la conducta habitual hace que, casi siempre, los consumidores busquen en internet los mismos medios que consumen tradicionalmente en otros formatos (**Rosenstein; Grant**, 1997).

El hábito de leer noticias se forma aproximadamente en la adolescencia con la influencia de familiares y del entorno en el que la persona crece (**Stone; Wetherington**, 1979).

Surgen entonces las preguntas que guían esta investigación: ¿Qué hábitos de consumo de información periodística en internet tienen los colombianos? y ¿cuáles son los criterios que llevan al público al consumo de un determinado medio informativo en internet?

“ El hábito de leer noticias se forma en la adolescencia con la influencia de familiares y del entorno en el que la persona crece ”

2. Marco teórico

En el caso del consumo de medios digitales, para responder a este tipo de preguntas los investigadores suelen basarse en teorías como la de la difusión de innovación (**Rogers**, 1995), que habla de que un nuevo medio es compatible con las preferencias pasadas del usuario, siempre y cuando la innovación que introduzca sea fácilmente entendible, se pueda probar o ensayar cómo funciona el nuevo medio de una manera sencilla, y la percepción de importancia que tenga el medio para esa persona.

Por su parte, la teoría clásica de usos y gratificaciones (**Katz**, 1959), que relaciona la elección de un medio con su capacidad para dar respuesta a las necesidades del usuario, ha servido de marco a diversas investigaciones (**Beaudoin y Thorson**, 2004; **Kaye y Johnson**, 2002; **Papcharissi y Rubin**, 2000; **Rodgers y Thorson**, 2001), que tienen en cuenta factores sociales, psicológicos y diferencias individuales de los consumidores de medios digitales.

Con base en dicha teoría, **Thorson y Duffy** (2005) crearon el *Media choice model* para determinar cómo y por qué las personas consumen medios digitales. Allí se incluyen variables como la conectividad, la necesidad de información y de entretenimiento, la capacidad de pago; la interactividad (con autores y/o lectores); la inmediatez de la información y la personalización de consulta del medio.

En lo que se refiere específicamente al consumo de periódicos digitales, **Flavián y Gurrea** (2007, 2008) han creado el *Modelo estructural de análisis*, basado en cinco elementos: la usabilidad, la reputación, la credibilidad, la privacidad con que manejan los datos de sus lectores y la familiaridad de los consumidores con los sitios web.

Así pues, la presente investigación toma como referente estos dos modelos para indagar sobre los criterios y hábitos de consumo de información periodística en internet que tienen los colombianos.

3. Método

3.1. Sujetos y diseño

A partir del envío de 103.650 correos electrónicos durante el mes de septiembre de 2010, y de la publicación de un link de acceso al micrositio de la *Facultad de Comunicación* de la *Universidad de La Sabana* en el que estaba disponible el cuestionario, invitando a colaborar voluntariamente con el mejoramiento del periodismo colombiano en internet, se logró convocar un total de 344 hombres y mujeres (índice de respuesta = 0,003). En la tabla 1 se describen los diferentes datos socio-demográficos de los participantes en el estudio.

La mayoría de los participantes reportó pertenecer a los estratos socioeconómicos 3 (29,7%) y 4 (38,4%) de Colombia¹. La participación de hombres y mujeres fue balanceada y se obtuvo mayor participación de personas con pregrado universitario (45,3%) o con post-grado (39%) (ver tabla 1).

El presente estudio se concibe como un diseño no experimental descriptivo en el que se hace una exploración cuantitativa de las variables analizadas.

3.2. Instrumento

Para dar respuesta a las preguntas de investigación planteadas, se diseñó un cuestionario con 46 ítems, que permiten, por una parte, evaluar los criterios que tienen en cuenta las personas al momento de elegir las páginas web de carácter informativo o periodístico que consumen, y por otra parte, establecer los hábitos de consumo que tienen con respecto a este tipo de páginas.

La encuesta fue elaborada a partir de una exhaustiva revisión de la bibliografía existente sobre el tema, tomando

en consideración los conceptos que han sido identificados como los más relevantes para el estudio del consumo de páginas web, y se basaron en los modelos de preguntas de investigaciones precedentes como las de **Flavián y Gurrea** (2007, 2008) y **Hashim y Meloche** (2007), entre otras.

Los ítems inicialmente diseñados fueron sometidos a una revisión de expertos y luego fueron piloteados con 40 estudiantes de comunicación de la *Universidad de La Sabana* para determinar los índices de confiabilidad iniciales y para mejorar el comportamiento de las escalas.

El éxito de los medios periodísticos digitales está muy ligado al conocimiento y uso adecuado de las herramientas y servicios que ponen a disposición del usuario

Se optó por diseñar las preguntas con una escala de calificación tipo Likert de seis puntos. Para los hábitos de consumo se construyeron 6 ítems con un escalamiento que iba desde totalmente en desacuerdo = 1, hasta totalmente de acuerdo = 6, cuyo índice de confiabilidad fue de 0,76 (alpha de Cronbach).

Para los criterios de consumo de medios informativos online se tuvieron en cuenta 6 variables contempladas en la bibliografía revisada y definidas así:

1. Usabilidad

Grado de esfuerzo que, de acuerdo con el usuario, es necesario para utilizar el sitio web (**Rivera**, 2002; **Baeza-Yates; Rivera; Velasco**, 2004).

VARIABLES SOCIO-DEMGRÁFICAS	Categorías	Frec.	%
Género	Femenino	165	48,0
	Masculino	178	48,6
	sub-total	333	96,7
Rangos de edad	18 a 27 años	156	45,2
	28 a 37 años	71	20,6
	38 a 47 años	66	19,2
	años o más	40	11,6
	sub-total	333	96,7
Grado de escolaridad	Postgrado	134	39,0
	Pregrado universitario	156	45,4
	Secundaria	43	12,2
	sub-total	333	96,7
Estrato socioeconómico	Bajo	23	6,7
	Medio	234	65,0
	Alto	86	25,0
	sub-total	333	96,7
Datos no reportados		11	3,3
Total		344	100

Tabla 1: Frecuencias y porcentajes correspondientes a las variables socio-demográficas de los participantes.

2. Accesibilidad

Facilidad con que se accede a la página web del medio periodístico (Faber; Lee; Nan, 2004).

3. Convergencia

Integración de herramientas, espacios, métodos de trabajo y lenguajes distribuidos en múltiples plataformas digitales (Garrison, 2003; Salaverría; García-Avilés; Masip, 2010);

4. Interactividad

Posibilidad de dar opiniones en un foro, obtener respuesta del autor de la nota periodística y/o compartir con otros lectores información a través de la Red o de otros medios (Deuze, 1999; Hall, 2001; Kawamoto, 2003).

5. Personalización de las noticias

Posibilidad de consumir noticias de acuerdo con los temas de interés de los lectores (Merialdo; Lee; Luparello; Roudaire, 1999), y

6. Posicionamiento extendido

Relación con el posicionamiento del medio en su versión tradicional (Rosenstein; Grant, 1997).

Cada variable estaba compuesta por una serie de ítems, que fueron medidos en escala tipo Likert, que iba desde Indiferente = 1, hasta Muy importante = 6, con el fin de identificar el grado de importancia otorgado por los participantes a cada una de las afirmaciones. La confiabilidad, calculada con base en el alpha de Cronbach, y el número de ítems usado en cada escalas fueron: usabilidad = 0,76 (9 ítems); accesibilidad = 0,71 (8 ítems); convergencia = 0,78 (4 ítems); interactividad = 0,87 (10 ítems); personalización de las noticias = 0,81 (9 ítems); posicionamiento extendido = 0,66 (3 ítems).

4. Resultados

4.1. Hábitos de consumo de información periodística por internet

Si bien el 89,9% de los participantes consulta internet diariamente, el 55,8% consulta noticias todos los días. Más del 75% de los encuestados consulta las noticias en webs de medios tradicionales como periódicos o revistas que en principio tienen una versión impresa (57,3%) o pertenecen a algún noticiero de televisión (20,6%). Un 32,6% visita el mismo sitio informativo online por costumbre, mientras que sólo un 5,5% consulta diferentes sitios informativos.

Adicionalmente, el 45,8% de los encuestados tiene el hábito de buscar noticias de su interés, es decir, que utilizan internet principalmente para profundizar sobre temas que han llamado su atención en otro medio. Además, el 64,6% de los consumidores de páginas informativas online asegura que es importante que el medio les permita personalizar la manera en que consultan la información de acuerdo con sus intereses, posea un servicio de alertas sobre aquellos temas o secciones de su interés (54,8%), cuente con un espacio donde se puede acceder a las noticias más leídas diariamente (65,4%), y clasifique las noticias según el orden en que han sido publicadas (70%). En la tabla 2 se presentan algunos hábitos adicionales encontrados en la investigación.

Items	%
Consulta diariamente internet	89,9
Consulta diariamente noticias online	55,8
Consulta noticias en páginas de medios tradicionales	75,1
Visita el mismo sitio informativo	32,6
Consulta diferentes sitios informativos	5,5
Busca noticias de su interés	45,8
Consulta noticias por ordenador	56,1
Accede a noticias por el teléfono móvil	6,4

Tabla 2: Hábitos de consumo de noticias en internet

El 47,1% no está suscrito a sistemas de mensajería RSS y tan sólo el 6,4% prefiere acceder a las noticias mediante un teléfono móvil, frente a un 56,1% que recurre al ordenador. Sin embargo, el 54,1% considera importante que el sitio web permita acceder a la información a través de dispositivos diferentes del ordenador (teléfono móvil, iPad, Blackberry, etc.), lo que hace pensar que un desarrollo más amigable de estos dispositivos podría potenciar la consulta de webs informativas.

El auto-refuerzo de los hábitos de comportamiento hace que el lector busque la versión digital del medio que ha consultado anteriormente en formatos tradicionales

4.2. Criterios de consumo de información periodística por internet

Aunque sólo el 8,4% de las personas encuestadas dice prestar atención al diseño del sitio web, el hecho de que las diferentes páginas del sitio mantengan coherencia visual (colores, letras, distribución) resulta un criterio importante para el 70% de los consultados.

Por otro lado, el 62,7% considera bastante importante que una misma pieza periodística esté construida por varios tipos de elementos, como imágenes, vídeo o audio, al punto que el 77,3% resalta la importancia de poder encontrar seguimiento a una nota periodística, en un mismo día, en diferentes formatos (texto, audio, vídeo, etc.) de fuentes propias del medio o de otras fuentes informativas. Para el 59% de los encuestados también es importante que desde la página de consulta habitual se pueda acceder a otros medios informativos directamente (por ej.: emisoras de radio o canales de televisión online).

Así mismo, los encuestados mostraron que los hábitos de consumo a la hora de buscar noticias en internet están encaminados a escoger sitios informativos online que contengan los siguientes atributos: organizado (71%), visualmente atractivo (70,4%), legible (74,4%), gratuito (63%), que permita que la modificación del tamaño de letra (61%), que esté siempre disponible (71,2%), que sea fácil de usar (62,8%), que se actualice con frecuencia (68,3%), que sus vínculos carguen rápido (68,3%), que tenga buscador propio (52%),

y que tenga páginas de ayuda que faciliten la navegación (66,3%).

El 68,7% de los encuestados considera relevante que el sitio web informativo les permita discutir las noticias publicadas, mientras que el 67% le da importancia a poder escribirle un correo al autor de la nota periodística, y otro 68% considera importante obtener respuestas del mismo.

Llama la atención que, aunque el 64,6% de los consumidores de diarios online dice preferir un medio que le permita personalizar la manera en que consulta la información, la importancia de una serie de elementos relacionados con la interactividad y la personalización no es todavía claramente percibida por los mismos. Por ejemplo, que el sitio web permita modificar el contenido publicado (corregir información, agregar datos, vínculos, etc.), es indiferente para el 39% y medianamente importante para el 36%; que les permita organizar la presentación de la página a su gusto es medianamente importante para el 30% e indiferente para el 34,8%; no obstante, el que tenga un espacio para publicar noticias propias es medianamente importante para el 40% de los consultados y muy importante para el 36%.

Finalmente, el 40,7% de los consultados está de acuerdo con ignorar la publicidad cuando lee noticias online, y el 43,9% suele participar en encuestas, juegos o concursos que aparecen en los diarios digitales. En la tabla 3 están recopilados los principales criterios que promueven el consumo de medios informativos online.

5. Conclusiones

De acuerdo con los hábitos encontrados, se confirma la propuesta de **Rosenstein y Grant (1997)** sobre el auto-refuerzo de los hábitos de comportamiento, que hacen que el lector busque la versión digital del medio que ha consultado anteriormente en formatos tradicionales. Con respecto a los criterios de consumo, se destaca que el diseño de la página, la tipografía y la implementación de herramientas de comunicación visual sean considerados como elementos que reafirman la calidad de una página web informativa y adicionalmente trazan un estándar para una comunicación efectiva y eficiente en internet.

Los consumidores colombianos de noticias en internet tienen claras las características que no pueden faltar a la hora de escoger una página informativa online, particularmente las asociadas con la usabilidad y la accesibilidad del sitio web. Por intuición, se considera que el grupo objetivo de las páginas web informativas tiene los conocimientos necesarios para consultarlas, pero no todos los lectores tienen la misma habilidad para entender los sistemas de navegación de cada página o establecer en qué parte pueden encontrar el contenido que necesitan. Por esta razón, sería conveniente que los medios de comunicación colombianos, de carácter periodístico, analicen si sus sitios web cumplen con los atributos destacados como importantes por los encuestados en esta investigación para realizar procesos de mejora, como por ejemplo, aquellos que se refieren a la disponibilidad de diversos formatos digitales para el cubrimiento de una misma noticia o la posibilidad de profundizar la información a través de hipervínculos o en otros medios igualmente digitales.

Complementariamente, la interactividad de este particular medio permite y, hasta cierto punto exige, poder retroalimentar o consultar directamente al autor de la nota periodística para obtener información adicional sobre la misma, e incluso poder participar como 'reportero' digital y publicar su propia información o modificar la ya existente. Igualmente, valdría la pena abordar una investigación sobre las nuevas formas en que circula la información noticiosa, ya que muchas veces las redes sociales como *Facebook* y *Twitter* cumplen con la tarea de informar sobre las noticias más relevantes o las que más interesan a un grupo de personas, de manera que desplazan en cierto modo a las páginas de los periódicos o medios noticiosos virtuales. Sin embargo, esto genera preguntas de orden ético en la publicación de la información ya que criterios de calidad como la consulta y publicación de diversas fuentes, de puntos de vista y de información del contexto noticioso, entre otros (**Alessandri; Edwards; Pellegrini; Puente; Rozas; Saavedra; Porath, 2001**) no estarían bajo el control del medio sino de quien publica, por lo que los intereses inherentes a la divulgación de la información generarían importantes sesgos en los consumidores de noticias online.

Items	%	Items	%
Noticias construidas por varios tipos de elementos: imágenes, video, audio	62,7	Que el medio permita personalizar la consulta de información	64,6
Seguimiento a una misma noticia, en un mismo día, en diferentes formatos (texto, audio, video, etc.),	77,3	Que el sitio posea un servicio de alertas sobre temas de su interés	54,8
Acceso a otros medios informativos directamente (programas de radio o TV)	59,0	Que el sitio cuente con un espacio de acceso a las noticias más leídas	65,4
Que el sitio sea organizado	71,0	Que el medio clasifique las noticias según el orden de publicación	70,0
Que el sitio sea visualmente atractivo	70,4	Que sea legible	74,4
Que sea de consulta gratuita	63,0	Que se actualice con frecuencia	68,3
Que permita la modificación del tamaño de letra	61,0	Que sus vínculos carguen rápido	68,3
Que esté siempre disponible	71,2	Que permita discutir las noticias publicadas	68,7
Que sea fácil de usar	62,8	Considera importante tener acceso a la información por dispositivos diferentes al computador	54,1

Tabla 3: Criterios que promueven el consumo de noticias en internet

Es evidente la relación entre el posicionamiento del medio en su versión tradicional y el consumo de su versión online. Sin embargo, también es claro que el éxito de los medios periodísticos digitales está muy ligado al conocimiento y uso adecuado de las herramientas y servicios que pone a disposición del usuario esta plataforma, de tal manera que los medios deben mostrar a sus usuarios, de forma clara y sencilla, qué ventajas pueden recibir tanto en comunicación, como en ahorro de tiempo, personalización de la información, entre otras características que permite internet, de manera que se mantenga fiel al medio.

6. Nota

1. Los estratos socioeconómicos son una forma de estratificación social basada en las remuneraciones que reciben las personas o los impuestos que pagan. En Colombia son empleados para clasificar los inmuebles residenciales. Legalmente existen 7 estratos socioeconómicos: 0 = Inferior, 1 = Muy bajo, 2 = Bajo, 3 = Medio-bajo, 4 = Medio, 5 = Medio-alto, 6 = Alto, 7 = Superior o de alto poder económico.

http://es.wikipedia.org/wiki/Estratificación_socioeconómica_en_Colombia

7. Bibliografía

Alessandri, Francisca; Edwards, Cristóbal; Pellegrini, Silvia; Puente, Soledad; Rozas, Eliana; Saavedra, Gonzalo; Porath, William. “VAP: un sistema métrico de la calidad periodística”. *Cuadernos de la información*, 2001, v. 14, pp. 112-120.

Asociación para la Investigación de Medios de Comunicación (AIMC). “La prensa: digital vs papel”, 2011.

http://www.aimc.es/spip.php?action=acceder_documento&arg=1817&cle=301c9291dbedaa2e29aa35d8f6ef681&file=pdf%2Fimm_prensa2011.pdf

Baeza-Yates, Ricardo; Rivera, Cuahtémoc; Velasco, Javier. “Arquitectura de la información y usabilidad en la Web”. *El profesional de la información*, 2004, mayo-junio, v. 13, n. 3, pp. 168-178.

Beaudoin, Christopher; Thorson, Esther. “Testing the cognitive mediation model: The roles of news reliance and three gratifications sought”. *Communication research*, 2004, v. 31, n. 4, pp. 446-471.

Deuze, Mark. “Journalism and the Web: an analysis of skills and standards in an online environment”. *International communication gazette*, 1999, v. 61, n. 5, pp. 373-390.

Diddi, Arvind; LaRose, Robert. “Getting hooked on news: Uses and gratifications and the formation of news habits among college students in an internet environment”. *Journal of broadcasting & electronic media*, 2006, v. 50, n. 2, pp. 193-210.

Faber, Ronald; Lee, Mira; Nan, Xiaoli. “Advertising and the consumer information environment online”. *American behavioral scientist*, 2004, v. 48, n. 4, pp. 447-466.

Flavián, Carlos; Gurrea, Raquel. “Perceived substitutability between digital and physical channels: the case of newspapers”. *Online information review*, 2007, v. 31, n. 6, pp. 793-813.

Flavián, Carlos; Gurrea, Raquel. “Reading newspapers on the internet: the influence of web sites’ attributes”. *Internet research*, 2008, v. 18, n. 1, pp. 26-45.

Garrison, Bruce. “The perceived credibility of electronic mail in newspaper newsgathering”. Documento presentado a la Newspaper Division, de la Association for Education in Journalism and Mass Communication, 2003, Aug., Kansas City, Missouri.

Hall, Jim. *Online journalism: a critical primer*. London: Pluto Press, 2001.

Hashim, Nor-Hazlina; Meloche, Joseph. “Australian online newspaper: an exploratory study on internet savvy users, using Q-methodology”. *Journal of human subjectivity*, 2007, v. 5, n. 2., pp. 5-68.

Katz, Elihu. “Mass communications research and the study of popular culture: an editorial note on a possible future for this journal”. *Studies in public communication*, 1959, n. 2, pp. 1-6.

Kawamoto, Kevin (ed.). *Digital journalism: emerging media and the changing horizons of journalism*. Lanham [etc.]: Rowman and Littlefield, 2003.

Kaye, Barbara; Johnson, Thomas. “Online and in the know: Uses and gratifications of the Web for political information”. *Journal of broadcasting & electronic media*, 2002, March, v. 46, n. 1, pp. 54-71.

LaRose, Robert; Eastin, Matthew. “Is online buying out of control? Electronic commerce and consumer self-regulation”. *Journal of broadcasting & electronic media*, 2002, v. 46, n. 4, pp. 549-564.

Martin, Neale. *Habit: The 95% of behavior marketers ignore*. New Jersey: Pearson Education, 2008.

Meso-Ayerdi, Koldobika; Díaz-Noci, Javier; Larrondo-Ureta, Ainara. “Presencia y uso de internet en las redacciones de los diarios vascos y navarros”. *Mediatika*, 2010, v. 12, pp. 301-319.
http://dspace.unav.es/dspace/bitstream/10171/17955/1/Presencia_y_uso_de_internet_en_las_redacciones_de_los_diarios_vascos_y_navarros__Mediatika_12.pdf

Masip, Pere; Díaz-Noci, Javier; Domingo, David; Micó-Sanz, Josep-Lluís; Salaverría, Ramón. “Investigación internacional sobre ciberperiodismo: hipertexto, interactividad, multimedia y convergencia”. *El profesional de la información*, 2010, v. 19, n. 6, pp. 568-576.

Merialdo, Bernard; Lee, Kyung-Tak; Luparello, Dario; Roudaire, Jeremie. “Automatic construction of personalized tv news programs”. En: *ACM multimedia conf*, 1999, pp. 323-331.

Oppenheimer, Walter. “The Guardian da prioridad a la edición digital frente a la de papel”. *El país*, 16 junio 2011.
http://www.elpais.com/articulo/internacional/The/Guardian/da/prioridad/edicion/digital/frente/papel/elpepuint/20110616elpepuint_12/Tes

Papcharissi, Zizi; Rubin, Alan. “Predictors of internet use”. *Journal of broadcasting and electronic media*, 2000, v. 44, n. 2, pp. 175-196.

Rivera, Cuauhtémoc. "Usabilidad en la Web". Tutorial del Centro de investigación en la Web, 2002.

<http://www.dcc.uchile.cl/~crivera/CCBOL-martes.pdf>

Rodgers, Shelly; Thorson, Esther. "The interactive advertising model: How users perceive and process online ads". *Journal of interactive advertising*, 2000, v. 1, n. 1, pp. 42-61.

<http://www.jiad.org/vol1/no1/rodders/index.htm>

Rodríguez-Martínez, Ruth; Codina, Lluís; Pedraza-Jiménez, Rafael. "Cibermedios y web 2.0: modelo de análisis y resultados de aplicación". *El profesional de la información*, 2010, enero-febrero, v. 19, n. 1, pp. 35-44.

Rosenstein, Aviva; Grant, August. "Reconceptualizing the role of habit: A new model of television audience activity". *Journal of broadcasting & electronic media*, 1997, v. 41, n. 3, pp. 324-344.

Rogers, Everett. *Diffusion of innovations*. New York: Free Press, 1995.

Salaverría, Ramón. *Redacción periodística en internet*. Pamplona: Eunsa, 2005.

Stone, Gerald; Wetherington, Roger. "Confirming the newspaper reading habit". *Journalism quarterly*, 1979, v. 56, n. 3, pp. 554-561.

Tendencias Digitales. *Usos de internet en Latinoamérica*. Caracas, 28 octubre 2010.

<http://www.tendenciasdigitales.com/1010/492-de-los-lati-noamericanos-prefiere-internet-para-comunicarse-con-las-marcas>

The Pew Research Center for the People and the Press. "Internet gains on television as public's main news source". 4 January 2011.

<http://people-press.org/2011/01/04/internet-gains-on-television-as-publics-main-news-source>

Thorson, Esther; Duffy, Margaret. "Newspapers in the new media environment". En *NAA Marketing conference*, Orlando, March 22, 2005.



Construye tu futuro



Máster universitario en Gestión de Contenidos Digitales

En colaboración con la Universitat Pompeu Fabra

Máster universitario en Bibliotecas y Colecciones Patrimoniales

Máster propio en Gestión Documental e Información en las Empresas

Más información: <http://www.ub.edu/biblio/masters/masteres.html>

Facultat de Biblioteconomia i Documentació

Edifici UB-Sants C/Melcior de Palau, 140 08014 Barcelona

Tel.: 93 403 57 70 Fax: 93 403 57 72 bibdoc@ub.edu <http://www.ub.edu/biblio>



CARENCIAS EN LOS SISTEMAS DE GESTIÓN DEL CONOCIMIENTO: UNA REVISIÓN BIBLIOGRÁFICA



Laura-Patricia Pinto-Prieto, Luis-Eduardo Becerra-Ardila y Luis-Carlos Gómez-Flórez



Laura-Patricia Pinto-Prieto es ingeniera de sistemas egresada con distinción cum laude de la *Univ. Industrial de Santander (UIS)*, Colombia, en 2009. Es becaria del programa *Jóvenes Investigadores del Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología Francisco José de Caldas (Colciencias)*. Estudiante de maestría en ingeniería industrial vinculada a los grupos de investigación *STI (Sistemas y Tecnologías de Información)* e *Innotec (Centro para la gestión y la innovación tecnológica)* de la *UIS*. Investiga en gestión del conocimiento, tecnologías de información, sistemas de información, e-learning, pensamiento sistémico y gestión tecnológica.

*Universidad Industrial de Santander
Facultad de Ingenierías Fisicomecánicas, Escuela de Ingeniería de Sistemas
Grupo de Investigación en STI. Oficina LP 325
Carrera 27, Calle 9, Apartado aéreo 678 – Bucaramanga, Colombia
lauris2687@hotmail.com*



Luis-Eduardo Becerra-Ardila es ingeniero industrial egresado de la *Universidad Industrial de Santander (UIS)*, Colombia, magister en administración del *Instituto Tecnológico de Estudios Superiores de Monterrey* y doctorando en ingeniería en la *UIS*. Es profesor asociado y jefe de división financiera de la misma universidad, y coordinador del proyecto *SUMA (Towards Sustainable Financial Management of Universities in Latin America)* financiado por la *Comisión Europea*, mediante el programa *ALFA III*; investigador del grupo de investigación *Innotec (Centro para la Gestión y la Innovación Tecnológica)*; investiga en áreas de la gestión tecnológica y la administración de empresas.

*Universidad Industrial de Santander
Facultad de Ingenierías Fisicomecánicas, Escuela de Ingeniería de Sistemas
Grupo de investigación Innotec
Carrera 27, Calle 9, Apartado aéreo 678 – Bucaramanga, Colombia
lbecerra@uis.edu.co*



Luis-Carlos Gómez-Flórez es ingeniero de sistemas y magister en informática de la *Universidad Industrial de Santander (Colombia)*. Profesor titular y director de investigación y extensión de la *Facultad de Ingenierías Fisicomecánicas* de la misma universidad. Director del grupo de investigación *STI (Sistemas y Tecnologías de la Información)*. Investigación sobre administración de la información y gestión del conocimiento en las organizaciones, auditoría y control en sistemas y tecnologías de información, gestión de procesos y calidad en ingeniería del software, pensamiento y metodología de los sistemas blandos, sistemas y tecnologías de la información en las organizaciones.

*Universidad Industrial de Santander
Facultad de Ingenierías Fisicomecánicas, Escuela de Ingeniería de Sistemas
Grupo de Investigación en STI. Oficina LP 325
Carrera 27, Calle 9, Apartado aéreo 678 – Bucaramanga, Colombia
lcgomezf@uis.edu.co*

Resumen

La importancia del conocimiento como ventaja competitiva sostenible y el desarrollo de las tecnologías de información, han dado lugar a los sistemas de gestión del conocimiento (SGC), facilitando e impulsando los procesos de generación, codificación y transferencia del conocimiento en las organizaciones. El interés de profesionales, académicos e investigadores es creciente desde hace varios años, y la gestión de conocimiento es un tema en auge. Sin embargo, al examinar su estado en las organizaciones, se encuentran grandes dificultades en su incorporación. Mediante una revisión de la bibliografía se analizan los fallos usuales producidos en la implementación de los SGC. Los resultados reflejan la visión reduccionista que se centra sólo en las tecnologías de la información, revelándose la necesidad de una visión holística para poder tener éxito.

Palabras clave

Gestión del conocimiento, Conocimiento, Sistemas y tecnologías de información, Sistemas de gestión de conocimiento, Procesos de gestión de conocimiento, Fallos, Carencias.

Artículo recibido el 10-12-11
Aceptación definitiva: 29-02-12

Title: Gaps in knowledge management systems: a bibliographical review**Abstract**

The importance of knowledge as a sustainable competitive advantage and the increasing development of information technology (IT) have led to knowledge management systems (KMS), facilitating and encouraging the processes of generation, codification and transfer of knowledge in organizations. The interest of professionals, academics and researchers has been growing for several years, to the point that it can be stated that knowledge management is a topic on the rise. Nonetheless, if we review their status in organizations, there are great difficulties in their implementation, difficulties that cover various aspects. Through a review of the literature, this paper explores the present gaps in the implementation of KMS. The results reflect the dominance of a reductionist approach in this area, with a focus on IT, revealing the need for a holistic approach, guided by what makes sense in the organization.

Keywords

Knowledge management, Knowledge, Information systems and technologies, Knowledge management systems, Knowledge management processes, Gaps in knowledge management systems.

Pinto-Prieto, Laura-Patricia; Becerra-Ardila, Luis-Eduardo; Gómez-Flórez. "Carencias en los sistemas de gestión del conocimiento: una revisión bibliográfica". *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 268-276. <http://dx.doi.org/10.3145/epi.2012.may.07>

Introducción

Estamos en un momento de fuertes cambios económicos y sociales, dentro de la llamada sociedad del conocimiento, que con el avance de las tecnologías de la información (TI) permite a las organizaciones gestionar el conocimiento (GC) para apoyar procesos de innovación que generan nuevos productos y servicios, y por ende obtienen mejores ventajas competitivas.

Algunos autores (Tseng, 2008; Lindner; Wald, 2010; Kruger; Johnson, 2010) consideran las TI como facilitadoras de la GC, al hacer posible que el conocimiento fluya en la organización; sin embargo, revisada la bibliografía se hallaron varios *gaps* o lagunas para su correcta aplicación. Según Lin y Tseng (2005) citado por Li y Chang, (2009), la mayoría de fallos se presentan entre la planificación y la implementación de proyectos de sistemas de GC, y en la diferencia entre los conocimientos necesarios según la percepción de los directivos y la de los demás empleados.

En este artículo se destaca la importancia de tratar la situación desde una visión holística o sistémica, específicamente desde el paradigma perspectivista de Checkland (1993), permitiendo identificar los conocimientos relevantes y la tecnología apropiada para la GC en la organización y de esta manera contribuir a reducir los *gaps* mencionados.

“ Sin las TI sería imposible realizar los cambios necesarios en las organizaciones para implementar la gestión del conocimiento ”

Gestión del conocimiento y tecnologías de información

A lo largo del tiempo han existido diversos enfoques que explican la obtención de ventajas competitivas de una orga-

nización. La evolución de dichos enfoques ha dado lugar a la *Visión basada en las competencias*, que se ha convertido en uno de los paradigmas dominantes en la dirección estratégica (Camisión, 2002; Foss, 1998).

Dentro de esta visión se encuentran los enfoques basados en los recursos, capacidades dinámicas y conocimiento.

En relación con el conocimiento, varios autores lo consideran la fuente de ventaja competitiva sostenible de la organización: Nonaka (1994), Nonaka y Takeuchi (1995), Spender (1996), Grant (1996), Kogut y Zander (1992), Garud (1997), (Forés, 2010; Alavi; Leidner, 2001) especialmente importante por su carácter intangible e inimitable (Prahalad; Hamel, 1990; Grant, 1991; Stalk; Evans; Shulman, 1992; Knudsen, 1996; Teece; Pisano; Shuen, 1997).

Alavi y Leidner (2001) mencionan que dependiendo de la perspectiva bajo la cual se vea el conocimiento (transformación de datos a información, estado de la mente, objetos, procesos, facilidad de acceso a la información, capacidad) la GC y las TI cumplen un rol determinado. Si el conocimiento es visto como un proceso de aplicación de la experiencia (Carlsson *et al.*, 1996; McQueen, 1998; Zack, 1998), o como un objeto que puede ser almacenado y manipulado (Zack, 1998), la GC es un proceso que comprende crear, aplicar y transferir conocimientos en la empresa, para obtener un beneficio (Macintosh, 1996; Wiig, 1997; Andreu; Sieber, 1999; Alavi; Leidner, 2001).

En cuanto a las actividades y/o procesos que componen la GC no se ha llegado a un consenso. Muchos autores han creado modelos para explicar cuáles son, entre ellos Forés (2010), quien cita los modelos:

- crecimiento de conocimiento (Kogut; Zander, 1992);
- espiral de creación de conocimiento (Nonaka; Takeuchi, 1995);
- transferencia y transformación de conocimiento (Hedlund; Nonaka, 1993; Hedlund, 1994);
- aprendizaje (Leonard-Barton, 1995);

- las 4-i del aprendizaje organizativo (Crossan; Lane; White, 1999);
- cinco ciclos de aprendizaje (Sánchez, 2001);
- ciclo evolutivo del conocimiento (Zollo; Winter, 2002).

Analizando a varios autores, Jaime, Gardoni y Frank (2005) listan las siguientes actividades de GC: identificación, adquisición, selección, estructuración, creación, transferencia, uso, codificación y medición. Por su lado, Alavi y Leidner (2001) consideran cuatro procesos de GC: creación del conocimiento, almacenamiento, transferencia y aplicación.

Las TI son un facilitador de la GC y permiten una mejor comunicación y flujo del conocimiento en la organización. Algunos autores como Khandelwal y Gottschalk (2003), Spiegler (2003) y Sher y Lee (2004), destacan la importancia de la aplicación de las TI para dar soporte en los procesos de creación y transferencia de conocimiento. Schindler y Eppler (2003); Karlsen y Gottschalk (2004); Adenfelt y Lagerström (2006); Hanisch et al. (2009); y Lindner y Wald, (2010) indican que las TI se encuentran entre los factores de éxito más relevantes en la implementación de proyectos de carácter temporal de GC en organizaciones.

Dos conceptos ampliamente utilizados para relacionar las TI con la GC son los sistemas basados en el conocimiento (SBC) y los sistemas de gestión de conocimiento (SGC).

- Los primeros pretenden entender e imitar los conocimientos humanos haciendo uso de técnicas de inteligencia artificial (Wiig, 1997; Hendriks; Vriens, 1999; Huang, 2009; Chang, 2011). Sus principales componentes son: una base de conocimiento, un motor de inferencia, una herramienta de ingeniería de conocimiento y una interfaz de usuario (Wiig, 1997; Hendriks; Vriens, 1999).
- Los SGC son softwares diseñados para soportar los procesos de GC, y son los que principalmente nos referimos aquí (Von-Krogh, 1998; Baroni-de-Carvalho; Tavares, 2001; Croasdell et al., 2003; Benbya; Passiante; Aissa, 2004; Nevo; Chan, 2007; Richardson; Courtney; Haynes, 2006).

Sin embargo, pese a las bondades de las TI para la GC, se han identificado algunos *gaps* que dificultan la obtención de resultados satisfactorios en la implementación y uso de los SGC en las organizaciones.

Fallos en los SGC. Necesidad de una visión holística

Diversos estudios analizan las causas del fracaso de los proyectos de GC, sus limitaciones, carencias y enfoque dominante. Por ejemplo Alavi y Leidner (2001) indican que mientras la mayor parte de los estudios teóricos sobre la GC se centra en la estrategia de la organización, existen pocas investigaciones empíricas sobre el apoyo de las TI a la GC.

Varios investigadores se han referido al excesivo énfasis puesto en las tecnologías. Según Nonaka y Takeuchi (1995), Anand, Manz y Glick (1998),

Davenport y Prusak (1998), referenciados por Chua (2004), el éxito de la aplicación de la GC no se puede basar en el despliegue de una solución tecnológica por sí sola. Las investigaciones han revelado que la mayor dificultad en la GC es cambiar el comportamiento de la gente, y el mayor impedimento para la transferencia de conocimiento actual es la cultura organizativa. Superar las limitaciones tecnológicas, por el contrario, suele ser un asunto trivial (Ruggles, 1998). La tecnología existente no puede ser un sustituto completo del contacto cara a cara, que es crucial para la construcción de una cultura de intercambio de conocimientos (Roberts, 2000). Tampoco puede sustituir la interacción social humana, la interactividad entre los individuos necesarios para la creación de conocimiento (Fahey; Prusak, 1998).

El énfasis sobre los SGC se debe a que prevalece una visión reduccionista, centrada en las TI, en lugar de una visión holística. Esta situación también está presente en la elaboración de sistemas de información y en el uso de TI para la mejora de procesos organizativos (Davenport; Prusak, 1998; Barceló; Baglietto, 2001; Olave; Gómez, 2002; 2007).

Chua (2004) afirma que “la tecnología es capaz de superar las barreras del tiempo y el espacio que de otra manera serían limitantes en las actividades de gestión. También sirve como repositorio en el que el conocimiento puede ser contrastado, almacenado y recuperado de manera eficiente”. Este autor se centra en cómo reducir la distancia entre los consultores y los tecnólogos, y plantea un modelo de arquitectura para plasmar cómo pueden ser usadas las TI para soportar los procesos de GC. En la arquitectura identifica tres partes:

- infraestructura (almacenamiento y comunicación);
- conocimiento (creación, intercambio y reutilización); y
- presentación (personalización y visualización).

Lin y Tseng (2005) identificaron cinco *gaps* o carencias en la GC e ilustran la relación entre las actividades de gestión y el desempeño corporativo. A partir de esto, Tseng (2008) propone un marco conceptual que explora el rol y los efectos de las TI en los SGC, y cómo mejorar la eficacia y la eficiencia. La figura 1 presenta esos 5 *gaps* y cómo las TI pueden ayudar a reducirlos.

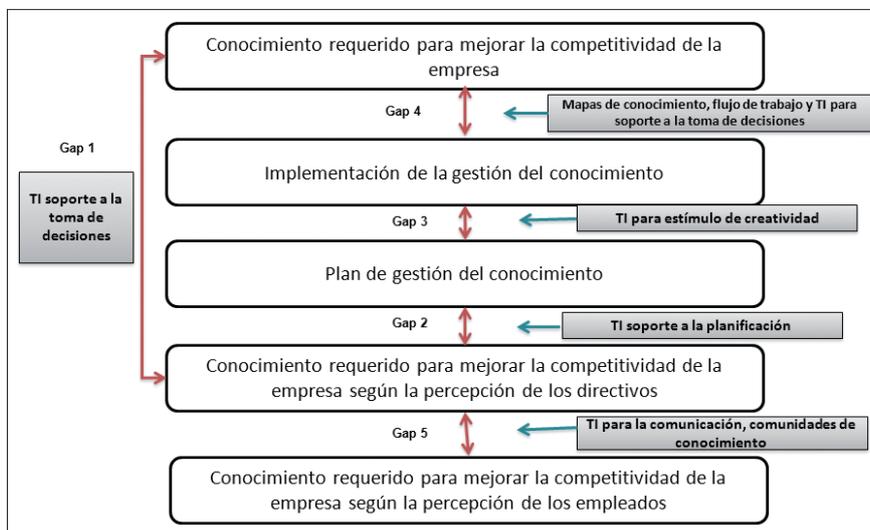


Figura 1. Gaps o fallos en la gestión del conocimiento e influencia de las TI

Los *gaps* se encuentran entre los conocimientos requeridos por la organización según la percepción de los directivos, y según la de los demás empleados. Existen inconsistencias entre la planificación e implementación de los SGC, que no son diseñados de tal manera que estén alineados con los objetivos y estrategias de la organización, y por lo tanto los conocimientos gestionados no son los requeridos para lograr los objetivos. La tabla 1 presenta los principales resultados obtenidos por **Tseng** (2008) al analizar el efecto de las TI sobre los *gaps* mencionados anteriormente.

Las TI para la GC pueden contribuir a la reducción de las lagunas y carencias, pero para ello cuando se compra o elabora un software para GC se tiene que tener en cuenta el dominio de la aplicación, el tipo de conocimiento y los procesos que van a ser soportados. En la tabla 2, tomada de **Rodríguez et al.** (2008), se listan varios autores y el enfoque de sus artículos relacionados con las TI para la GC. Sin embargo, todavía son pocos los estudios en los que se analiza la organización sistémicamente, teniendo en cuenta sus objetivos y estrategias. En el desarrollo de los SGC prevalece una visión excesivamente simplista.

Como se puede ver en la tabla 2, la mayoría de los trabajos se han centrado en analizar cómo las TI soportan las actividades de GC, siendo vistas como simples herramientas. Con este panorama si una organización desea diseñar un SGC es muy probable que se fije en las tecnologías, dejando a un lado otros factores mucho más relevantes que contribuirían a determinar si realmente es factible la GC bajo las condiciones actuales de la organización.

Por desgracia la definición de las tecnologías por parte de los técnicos y de la estrategia de GC por parte de la alta dirección se realizan independientemente, sin tener una visión holística del problema, y cada uno piensa que el SGC ya se adaptará. Esto se constata con los planteamientos de

Rubenstein et al. (2001), quienes realizaron un análisis de marcos de trabajo (*frameworks*) para la GC. Los clasifican en:

- descriptivos (proporcionan orientación sobre los tipos de procedimientos de GC, sin dar detalles específicos de cómo pueden ó deben llevarse a cabo), y
- prescriptivos (identifican atributos de la GC importantes por su influencia en el éxito o el fracaso de las iniciativas).

Los anteriores autores mencionan que en la literatura la mayor parte de los marcos de trabajo son prescriptivos y por lo tanto están más orientados hacia tareas.

Parte de las causas de los gap en los SGC, está en la forma en cómo se han visto los SGC, en la mayoría de los casos como: “las herramientas, tecnologías o software diseñados para soportar los procesos de generación, codificación y transferencia de conocimiento.”

Metodología de sistemas blandos: una alternativa para reducir la distancia entre gestores y técnicos

Para entender por qué es necesario ver la GC desde un enfoque sistémico es necesario tener claro qué es el pensamiento sistémico, dentro del cual existen diversos enfoques o perspectivas, que **Jackson** describió en 2003. En lo relacionado con la GC, la perspectiva que más sobresale es la llamada metodología de sistemas suaves o blandos (MSB) –*soft systems methodology*–, que busca definir la transformación de una organización mediante el debate y el acuerdo entre los actores involucrados. es un enfoque sistémico para abordar situaciones problemáticas. Proporciona un marco conceptual para personas que deben enfrentarse con situaciones problemáticas y no tienen una definición formal del problema.

La MSB proporciona herramientas y principios que ayudan a identificar los conocimientos relevantes y necesarios y la forma de gestionarlos. De esta forma los criterios no son impuestos desde la alta dirección sino que son definidos en conjunto con todos los empleados, garantizando que la implementación esté acorde con la planificación, dado que todos los actores han aceptado sus acuerdos y desacuerdos.

Otros autores (**Kingston; Macintosh, 2000; Rubenstein et al., 2001; Skok; Kalmanovitch, 2005; Butler et al., 2008; Kotiadis; Robinson, 2008; Georgiou, 2008**) destacan así mismo la utilidad del método de sistemas blandos. Los trabajos de **Yim et al.** (2004), **Georgiou** (2008), **Powell y Swart** (2010) aplican el pensamiento sistémico a los procesos de toma de decisiones.

Como se ha visto, existen muchas oportunidades de investigación en torno a los SGC, principalmente cuando se

Gap	Influencia de las TI
Gap 1	<ul style="list-style-type: none"> - El principal beneficio de las TI es facilitar a los directivos la identificación de un conocimiento valioso. - Las TI son sólo un medio para fomentar el conocimiento.
Gap 2	<ul style="list-style-type: none"> - Los repositorios de GC facilitan la captura, difusión y almacenamiento de conocimiento. - La necesidad de cambios incesantes en el conocimiento y el sistema estimula la creación de conocimiento y la innovación. - Los repositorios de conocimiento ayudan a los gerentes a planificar la GC, intensificar el aprendizaje organizativo, y mejorar la planificación y toma de decisiones.
Gap 3	<ul style="list-style-type: none"> - Los SGC estimulan la creatividad y la innovación en la organización post-industrial y permiten monitorear y controlar la ejecución de los planes de GC.
Gap 4	<ul style="list-style-type: none"> - Es necesario un sistema de medición completa que permita a la organización evaluar si la empresa mejora su competitividad después de la implementación del SGC. Esto se puede realizar con herramientas como el <i>Balanced scorecard</i> y los mapas estratégicos.
Gap 5	<ul style="list-style-type: none"> - Software de trabajo en grupo, de soporte de decisiones grupales, de flujo de trabajo (<i>work flow</i>), videoconferencias e intranets pueden facilitar el intercambio de información interna, la discusión y la comunicación en las organizaciones.

Tabla 1. Uso de las TI y posibles fallos o *gaps* de los SGC (basado en Tseng, 2008)

Autor	Año	Propósito	Aspectos cubiertos
Borghoff y Pareschi	1997	Clasificar las tecnologías para la GC	Actividades de GC Estructura del conocimiento (tácito o explícito)
Newman y Conrad	2000	Organizar y clasificar los métodos, practicas y tecnologías de la GC	Actividades de GC Estructura del conocimiento (tácito o explícito) Dominio de aplicación.
Alavi y Leidner	2001	Analizar el rol de los SGC en la GC organizativa	Actividades de GC
Binney	2001	Clasificar aplicaciones mencionadas en la literatura	Uso del conocimiento Actividades de GC
Marwick	2001	Describir características de las tecnologías de GC	Actividades de GC Estructura del conocimiento (tácito o explícito)
Rus <i>et al.</i>	2001	Clasificar las tecnologías para la GC usadas en ingeniería del software	Uso del conocimiento Dominio del conocimiento
Tiwana y Ramesh	2001	Clasificar las tecnologías de GC y describir algunos SGC	Dominio de aplicación (utilización de los conocimientos) Actividades de GC Cuestiones técnicas (específicos para sistemas basados en la web)
Valente y Housel	2002	Analizar y comparar herramientas de GC	Actividades de GC Estructura de conocimiento Dominio de aplicación
Woitsch y Karagiannis	2002	Definir funcionalidades requeridas de un SGC empresarial	Actividades de GC Estructura de conocimiento Cuestiones técnicas
Kankanhalli <i>et al.</i>	2003	Analizar el rol de las TI en la GC en un grupo de organizaciones	Uso del conocimiento
Van-Elst <i>et al.</i>	2003	Clasificar y analizar enfoques de GC basados en agentes	Actividades de GC Dominio de la aplicación
Qureshi <i>et al.</i>	2004	Analizar herramientas de groupware como tecnologías de GC	Actividades de GC Dominio de aplicación (utilización de los conocimientos) Cuestiones técnicas (relacionadas con groupware herramientas)
Chua	2004	Desarrollar una arquitectura de sistemas de GC que busca reducir la brecha entre los consultores y tecnólogos.	Actividades de GC Cuestiones técnicas
Banerjee	2005	Analizar cómo las herramientas de GC son usadas en la organización	Actividades de GC Dominio de aplicación (uso de conocimiento) Cuestiones técnicas (distribución de fuentes de conocimiento, temporalidad del acceso al conocimiento)
Rao	2005	Clasifica y describe características de las tecnologías de GC	Actividades de GC Dominio de aplicación (uso de conocimiento) Cuestiones técnicas (características de conocimientos y de tipos específicos de herramientas)
Vizcaino <i>et al.</i>	2005	Evaluar si los sistemas colaborativos soportan las actividades de GC	Actividades de GC
Rodríguez <i>et al.</i>	2008	Presenta un marco que puede utilizarse para analizar los sistemas de información como facilitadores del flujo de conocimientos en los procesos de la organización	Dominio de la aplicación Actividades de la GC Estructura de conocimiento Cuestiones técnicas
Gottschalk y Dean	2010	Crear un modelo de escenario para los sistemas de gestión del conocimiento en la vigilancia de la delincuencia financiera	Dominio de la aplicación Cuestiones técnicas
Zhao	2010	Desarrollar un marco de trabajo para facilitar el aprendizaje de los docentes	Dominio de la aplicación Actividades de GC Cuestiones técnicas

Tabla 2. Autores y enfoque de sus artículos sobre técnicas para SGC (adaptado de Rodríguez *et al.*, 2008)

busca relacionar el pensamiento sistémico con la GC. Pese a los intentos realizados por reducir los *gaps* mencionados no se ha conseguido un gran avance, dado que todavía prevalece la posición mecanocéntrica (centrada en la máquina) en lugar de la centrada en el ser humano (Olave; Gómez, 2007).

Conclusiones

La GC se incorpora cada vez más a las organizaciones para mejorar su rendimiento, pero existen carencias que dificultan su completa implementación de acuerdo con los objetivos y estrategias de la organización, y se queda sólo en gestión de la información. Esto se debe a un enfoque centrado

únicamente en los aspectos técnicos, dejando a un lado factores necesarios para la GC como la cultura organizativa, las políticas y objetivos de la organización y sobre todo el capital humano; en otras palabras, dejando de lado el verdadero sistema.

Una de las posibles causas de este problema radica en la forma reduccionista cómo se presentan los SGC, como si la simple tecnología fuera suficiente para implementar un sistema de gestión del conocimiento.

Bibliografía

Adenfelt, Maria; Lagerström, Katarina. "Enabling knowledge creation and sharing in transnational projects". *Intl journal of project management*, 2006, v. 24, n. 3, pp. 191-198.

<http://dx.doi.org/10.1016/j.ijproman.2005.09.003>

Alavi, Maryam; Leidner, Dorothy. "Review: knowledge management and knowledge management systems: conceptual foundations and research issues". *MIS quarterly*, 2001, v. 25, n. 1, pp. 107-136.

<http://dx.doi.org/10.2307/3250961>

Anand, Vikas; Manz, Charles; Glick, William. "An organizational memory approach to information management". *Academy of management review*, 1998, v. 23, n. 4, pp. 796-800.

<http://dx.doi.org/10.5465/AMR.1998.1255639>

Andreu, Rafael; Sieber, Sandra. "La gestión integral del conocimiento y del aprendizaje". *Economía industrial*, 1999, v. 326, pp. 63-72.

Banerjee, Ritendra. "A fool with a tool is still a fool". En: Rao, Madanmohan (ed.). *Knowledge management tools and techniques*. Amsterdam: Elsevier, 2005, pp. 283-292.

Barceló, María; Baglietto, Alicia. *Hacia una economía del conocimiento*. Price Waterhouse Coopers, ESIC Editorial, 2001. ISBN: 84 7356 272 0

Baroni-de-Carvalho, Rodrigo; Tavares, Marta. "Using information technology to support knowledge conversion processes". *Information research*, 2001, v. 7, n. 1.

<http://InformationR.net/ir/7-1/paper118.html>

Benbya, Hind; Passiante, Giuseppin; Aissa, Nassi. "Corporate portal: a tool for knowledge management synchronization". *Intl journal of information management*, 2004, June, v. 24, n. 3, pp. 201-220.

<http://choo.fis.utoronto.ca/fis/courses/lis2102/Readings/benbya.pdf>

<http://dx.doi.org/10.1016/j.ijinfomgt.2003.12.012>

Binney, Derek. "The knowledge management spectrum – understanding the KM landscape". *Journal of knowledge management*, 2001, v. 5, pp. 33-42.

<http://dx.doi.org/10.1108/13673270110384383>

Borghoff, Uwe; Pareschi, Remo. "Information technology for knowledge management". *Journal of universal computer science*, 1997, v. 3, n. 8, pp. 835-842.

http://www.jucs.org/jucs_3_8/information_technology_for_knowledge/Borghoff_U_M.pdf

Butler, Tom; Feller, Joseph; Pope, Andrew; Emerson, Bill; Murphy, Ciaran. "Designing a core IT artefact for knowledge management systems using participatory action research in a government and a non-government organisation". *Journal of strategic information systems*, 2008, v. 17, n. 4, pp. 249-267.

<http://dx.doi.org/10.1016/j.jsis.2007.10.002>

Carlsson, Sven; ElSawy, Omar; Eriksson, Innger; Raven, Arjan. "Gaining competitive advantage through shared knowledge creation: in search of a new design theory for strategic information systems". En: *Procs of the 4th European conf on information systems*, Lisbon, 1996.

Camisón, César. "Las competencias distintivas basadas en activos intangibles". En: Morcillo, Patricio; Fernández-Agüedo, Javier. *Nuevas claves para la dirección estratégica*. Barcelona: Ariel, 2002, pp. 117-151. ISBN: 84 344 2191 7

Chang, Wei-Lun. "iValue: a knowledge-based system for estimating customer prospect value". *Knowledge-based systems*, 2011, v. 24, n. 8, pp. 1181-1186.

<http://dx.doi.org/10.1016/j.knosys.2011.05.004>

Checkland, Peter. *Pensamiento de sistemas, práctica de sistemas*. México: Limusa Noriega, 1993. ISBN: 978 9681845254

Chua, Alton. "Knowledge management system architecture: a bridge between KM consultants and technologists". *Intl journal of information management*, 2004, v. 24, n. 1, pp. 87-98.

<http://dx.doi.org/10.1016/j.ijinfomgt.2003.10.003>

Croasdell, David; Jennex, Murray; Zhihong, Yu; Christianson, Tony; Chakradeo, Meenal; Makdum, Wagas. "A meta-analysis of methodologies for research in knowledge management, organizational learning and organizational memory: five years at HICSS". System Sciences. *Procs of the 36th Hawaii intl conf on system sciences*, 2003.

<http://www.computer.org/csdl/proceedings/hicss/2003/1874/04/187440110a-abs.html>

<http://doi.ieeecomputersociety.org/10.1109/HICSS.2003.1174253>

Crossan, Mary; Lane, Henry; White, Roderick. "An organizational learning framework: from intuition to institution". *Academy of management review*, 1999, v. 24, n. 3, pp. 522-537.

Davenport, Tom; Prusak, Laurence. *Working knowledge: how organizations manage what they know*. Boston: Harvard Business School Press, 1998.

Fahey, Liam; Prusak, Laurence. "The eleven deadliest sins of knowledge management". *California management review*, 1998, v. 40, n. 3, pp. 265-276.

<http://www.comp.dit.ie/dgordon/Courses/ResearchMethods/Countdown/11DeadliestSins.pdf>

Forés, Beatriz. *Prácticas organizativas, capacidades dinámicas y desempeño económico*. Tesis doctoral. Universitat Jaume I, 2010.

Foss, Nicolai. "Introduction: the emerging competence perspective". En: Foss, Nicolai J.; Knudsen, Christian (eds.). *To-*

wards a competence theory of the firm. Londres: Routledge, 1998, pp. 1-12. ISBN: 978 0415144728

Garud, Raghu. "On the distinction between know-how, know-why and know-what in technological systems". *Advances in strategic management*, 1997, v. 14, pp. 81-101. http://www.personal.psu.edu/users/r/u/rug14/48.On_the_Distinction_between_know-how__kno.pdf

Georgiou, Ion. "Making decisions in the absence of clear facts". *European journal of operational research*, 2008, v. 185, pp. 299-321. <http://dx.doi.org/10.1016/j.ejor.2006.12.038>

Gottschalk, Petter; Dean, Geoff. "Stages of knowledge management systems in policing financial crime". *Intl journal of law, crime and justice*, 2010, v. 38, n. 3, pp. 94-108. <http://dx.doi.org/10.1016/j.ijlcrj.2010.09.001>

Grant, Robert. "The resource-based theory of competitive advantages: implications for strategy formulation". *California management review*, 1991, v. 33, n. 3, pp. 114-135. http://www.skynet.ie/~karen/Articles/Grant1_NB.pdf

Grant, Robert. "Toward a knowledge-based theory of the firm". *Strategic management journal*, 1996, v. 17, n. 7, pp. 109-122. http://www.wedb.net/download/valoracao/aula_7/grant_1996.pdf

Hanisch, Bastian; Müller, Ana; Lindner, Frank; Wald, Andreas. "Knowledge management in temporary project environments". *Journal of knowledge management*, 2009, v. 13, n. 4, pp. 148-160. <http://dx.doi.org/10.1108/13673270910971897>

Hedlund, Gunnar. "A model of knowledge management and the n-form corporation". *Strategic management journal*, 1994, v. 15, pp. 73-90. <http://dx.doi.org/10.1002/smj.4250151006>

Hedlund, Gunnar; Nonaka, Ikujiro. "Models of knowledge management in the West and Japan". En: Lorange, Peter; Chakravarthy, Bala; Roos, Johan. *Implementing strategic processes*. Blackwell, 1993, pp. 117-144. ISBN: 978 0631185659

Hendriks, Paul; Vriens, Dirk. "Knowledge-based systems and knowledge management: friends or foes?" *Information & management*, 1999, v. 35, n. 2, pp. 113-125. [http://dx.doi.org/10.1016/S0378-7206\(98\)00080-9](http://dx.doi.org/10.1016/S0378-7206(98)00080-9)

Huang, Hao-Chen. "Designing a knowledge-based system for strategic planning: a balanced scorecard perspective". *Expert systems with applications*, 2009, v. 36, n. 1, pp. 209-218. <http://dx.doi.org/10.1016/j.eswa.2007.09.046>

Jaime, Astrid; Gardoni, Mickael; Frank, Christian. "Communications tools in research projects to support semi and non structured information". *Journal of systemics, cybernetics and informatics*, 2005, v. 3, n. 3, pp. 85-93. [http://www.iiisci.org/journal/CV\\$/sci/pdfs/P250685.pdf](http://www.iiisci.org/journal/CV$/sci/pdfs/P250685.pdf)

Jackson, Michael. *Systems thinking: creative holism for managers*. John Wiley, 2003. ISBN: 978 0470845226

Kankanhalli, Atreyi; Tanudidjaja, Fransiska; Sutanto, Juli-

ana; Tan, Bernard. "The role of IT in successful knowledge management initiatives". *Communications of the ACM*, 2003, v. 46, n. 9, pp. 69-73. <http://dx.doi.org/10.1145/903893.903896>

Karlson, Jan; Gottschalk, Petter. "Factors affecting knowledge transfer in IT projects". *Engineering management journal*, 2004, v. 16, n. 1, pp. 3-10.

Khandelwal, Vijay; Gottschalk, Petter. "Information technology support for interorganizational knowledge transfer: an empirical study of law firms in Norway and Australia". *Information resources management journal*, 2003, v. 16, n. 1, pp. 14-23. <http://dx.doi.org/10.4018/irmj.2003010102>

Kingston, John; Macintosh, Ann. "Knowledge management through multi-perspective modelling: representing and distributing organizational memory". *Knowledge-based systems*, 2000, v. 13, n. 2-3, pp. 121-131. [http://dx.doi.org/10.1016/S0950-7051\(00\)00053-8](http://dx.doi.org/10.1016/S0950-7051(00)00053-8)

Knudsen, Christian. "The competence perspective: a historical view". En: Foss, Nicolai J.; Knudsen, Christian (eds.). *Towards a competence theory of the firm*, 1996, pp. 13-37. ISBN: 978 0415144728

Kogut, Bruce; Zander, Udo. "Knowledge of the firm, combinative capabilities, and the replication of technology". *Organization science*, 1992, v. 3, n. 3, pp. 383-397. <http://zonecours.hec.ca/documents/H2008-1-1548530.seance10knowledgeofthefirm.pdf>

Kotiadis, Kathy; Robinson, Stewart. "Conceptual modelling: knowledge acquisition and model abstraction". *Procs of the 2008 Winter simulation*, 2008, pp. 951-958. <http://www.informs-sim.org/wsc08papers/114.pdf>

Kruger, Neels; Johnson, Roy. "Information management as an enabler of knowledge management maturity: A South African perspective". *Intl journal of information management*, 2010, v. 30, pp. 57-67. [http://137.215.9.22/bitstream/handle/2263/11613/Kruger_Information\(2009\).pdf?sequence=1](http://137.215.9.22/bitstream/handle/2263/11613/Kruger_Information(2009).pdf?sequence=1) <http://dx.doi.org/10.1016/j.ijinfomgt.2009.06.007>

Leonard-Barton, Dorothy. *Wellsprings of knowledge: Building and sustaining the sources of innovation*. Boston: Harvard Business School Press, 1995. ISBN: 978 0875848594

Li, Sheng-Tun; Chang, Won-Chen. "Exploiting and transferring presentational knowledge assets in R&D organizations". *Expert systems with applications*, 2009, v. 36, n. 1, pp. 766-777. <http://dx.doi.org/10.1016/j.eswa.2007.10.024>

Lin, Chinho; Tseng, Shu-Mei. "Bridging the implementation gaps in the knowledge management system for enhancing corporate performance". *Expert systems with applications*, 2005, v. 29, n. 1, pp. 163-173. <http://dx.doi.org/10.1016/j.eswa.2005.01.015>

Lindner, Frank; Wald, Andreas. "Success factors of knowledge management in temporary organizations". *Intl journal of project management*, 2010, v. 29, pp. 877-888. <http://dx.doi.org/10.1016/j.ijproman.2010.09.003>

- Macintosh, Ann.** *Position paper on knowledge assets management*. Artificial Intelligence Applications Institute, University of Edinburgh, Scotland, May 1996.
- Marwick, Alan.** "Knowledge management technology". *IBM systems journal*, 2001, v. 40, n. 4, pp. 814-830.
<http://dx.doi.org/10.1147/sj.404.0814>
- McQueen, Robert.** "Four views of knowledge and knowledge management". En: *Procs of the 4th Americas conf on information systems*, Aug 1998, pp. 609-611.
- Nevo, Dorit; Chan, Yolande.** "A Delphi study of knowledge management systems". *Information & management*, 2007, v. 44, n. 6, pp. 583-597.
<http://dx.doi.org/10.1016/j.im.2007.06.001>
- Newman, Brian; Conrad, Kurt.** "A framework for characterizing knowledge management methods, practice, and technologies". En: *Procs of the Practical aspects of knowledge management (PAKM 2000)*, The Knowledge Management Forum, Basel, Switzerland, 2000.
- Nonaka, Ikujiro.** "A dynamic theory of organizational knowledge creation". *Organization science*, 1994, v. 5, n. 1, pp. 14-37.
- Nonaka, Ikujiro; Takeuchi, Hirotaka.** *The knowledge-creating company. How Japanese companies create the dynamics of innovations*. Oxford University Press, Nueva York, 1995. ISBN: 978 0195092691
- Olave, Yesid; Gómez, Luis.** "IndiSIO: Instrumento de indagación sobre los sistemas de información en las organizaciones". *UIS ingenierías*, 2002, v. 1, n. 2, pp. 51-70.
- Olave, Yesid; Gómez, Luis.** "Una reflexión sistémica sobre los fundamentos conceptuales para sistemas de información". *Revista colombiana de computación*, 2007, v. 8, n. 1, pp. 71-92.
- Prahalad, Coimbatore; Hamel, Gary.** "The core competence of the corporation". *Harvard business review*, 1990, v. 68, n. 3, pp. 79-91.
http://faculty.fuqua.duke.edu/~charlesw/s591/willstuff/oldstuff/PhD_2007-2008/Papers/C08/Prahalad_Hamel_1990.pdf
- Powell, John; Swart, Juani.** "Mapping the values in B2B relationships: a systemic, knowledge-based perspective". *Industrial marketing management*, 2010, junio, v. 39, n. 3, pp. 437-449.
<http://dx.doi.org/10.1016/j.indmarman.2008.11.011>
- Qureshi, Sajda; Hlupic, Vlatka; Briggs, Robert.** "On the convergence of knowledge management and groupware". En: *Procs of the 10th intl workshop on groupware (Criwg'2004)*, Springer, San Carlos, Costa Rica, 2004, pp. 25-33.
- Rao, Madanmohan.** "Overview: the social life of KM tools". En: Madanmohan, Rao (ed.). *Knowledge management tools and techniques*, Amsterdam: Elsevier, 2005, pp. 1-73. ISBN: 978 0750678186
- Richardson, Sandra; Courtney, James; Haynes, John.** "Theoretical principles for knowledge management system design: application to pediatric bipolar disorder". *Decision support systems*, 2006, v. 42, n. 3, pp. 1321-1337.
<http://dx.doi.org/10.1016/j.dss.2005.11.001>
- Roberts, Joanne.** "Knowledge systems and global advertising services". *Creativity and innovation management*, 2000, sept., v. 9, n. 3, pp. 163-170.
<http://dx.doi.org/10.1111/1467-8691.00169>
- Rodríguez, Óscar; Martínez, Ana; Vizcaíno, Aurora; Favela, Jesús; Piattini, Mario.** "A framework to analyze information systems as knowledge flow facilitators". *Information and software technology*, 2008, v. 50, n. 6, pp. 481-498.
<http://dx.doi.org/10.1016/j.infsof.2007.07.002>
- Rubenstein, Bonnie; Liebowitz, Jay; Buchwalter, Judah; McCaw, Douglas; Newman, Butler; Rebeck, Kenneth.** "A systems thinking framework for knowledge management". *Decision support systems*, 2001, v. 31, n. 1, pp. 5-16.
[http://dx.doi.org/10.1016/S0167-9236\(00\)00116-0](http://dx.doi.org/10.1016/S0167-9236(00)00116-0)
- Ruggles, Rudy.** "The state of the notion: knowledge management in practice". *California management review*, 1998, v. 40, n. 3, pp. 80-89.
http://www.nipc.net/km/article/km_group/Ruggles-1998-State_of_the_Notion.pdf
- Rus, Ioana; Lindvall, Mikael; Sinha, Sachin.** "Knowledge management in software engineering: a state of the art report". *Data & analysis center for software*, 2001, 53 pp.
- Sánchez, Ron.** "Managing knowledge into competence: the five learning cycles of the competent organization". En: Sánchez, Ron (ed.). *Knowledge management and organizational competence*. Nueva York: Oxford University Press, 2001, pp. 3-37. ISBN: 978 0199259281
<http://dx.doi.org/10.1093/acprof:oso/9780199259281.003.0001>
- Sher, Peter; Lee, Vivid.** "Information technology as a facilitator for enhancing dynamic capabilities through knowledge management". *Information & management*, 2004, v. 41, n. 8, pp. 933-945.
<http://dx.doi.org/10.1016/j.im.2003.06.004>
- Schindler, Martin; Eppler, Martin.** "Harvesting project knowledge: a review of project learning methods and success factors". *Intl journal of project management*, 2003, v. 21, n. 3, pp. 219-228.
http://www.uncg.edu/bae/people/amoako/ISM654/reading_%2323.pdf
- Skok, Walter; Kalmanovitch, Caroline.** "Evaluating the role and effectiveness of an intranet in facilitating knowledge management: a case study at Surrey County Council". *Information & management*, 2005, v. 42, pp. 731-744.
<http://www2.warwick.ac.uk/fac/soc/wbs/conf/olkc/archives/oklc3/papers/id261.pdf>
- Spender, John.** "Making knowledge the basis of a dynamic theory of the firm". *Strategic management journal*, 1996, v. 17, (special issue: Knowledge and the firm), pp. 45-62.
<http://www.jcspender.com/uploads/Spender-SMJ96SI.pdf>
- Spiegler, Israel.** "Technology and knowledge: bridging a 'generating' gap". *Information & management*, 2003, v. 40, n. 6, pp. 533-539.
[http://dx.doi.org/10.1016/S0378-7206\(02\)00069-1](http://dx.doi.org/10.1016/S0378-7206(02)00069-1)

Stalk, Georg; Evans, Philip; Shulman, Lawrence. "Competing on capabilities: the new rules of corporate strategy". *Harvard business review*, 1992, v. 70, n. 2, pp. 57-69.

Teece, David; Pisano, Gary; Shuen, Amy. "Dynamic capabilities and strategic management". *Strategic management journal*, 1997, v. 18, n. 7, pp. 509-533.
<http://faculty.fuqua.duke.edu/~charlesw/s591/Bocconi-Duke/Papers/C10/TeecePisanoShuenSMJ.pdf>

Tiwana, Amrit; Ramesh, Balasubramaniam. "Integrating knowledge on the web". *IEEE internet computing*, 2001, v. 5, pp. 32-39.
<http://www.inf.ufes.br/~falbo/download/aulas/tengsoft/2007-2/Texto8.pdf>

Tseng, Shu-Mei. "The effects of information technology on knowledge management systems". *Expert systems with applications*, 2008, v. 35, pp. 150-160.
<http://dx.doi.org/10.1016/j.eswa.2007.06.011>

Valente, Andre; Housel, Thomas. "A framework to analyze and compare knowledge management tools". En: *Procs of the Knowledge-based intelligent information engineering and allied technologies (KES'2002)*, Crema, Italy, 2002, pp. 291-295.

Van-Elst, Ludger; Dignum, Virginia; Abecker, Andreas. "Towards agent-mediated knowledge management". En: *Procs of the intl symposium AMKM 2003*, Stanford, CA, USA, 2003, pp. 1-30. ISBN: 978 3 540 20868 6
<http://www.springerlink.com/content/7w83u24a12wv5fc0/fulltext.pdf>

Vizcaíno, Aurora; Piattini, Mario; Martínez, Manuel; Aranda, Gabriela. "Evaluating collaborative applications from a knowledge management approach". En: *Procs of the Evaluation of collaborative information systems and support for virtual enterprises (ECE)*, Linkoping, Sweden, 2005. ISBN: 0 7695 2362 5

<http://doi.ieeecomputersociety.org/10.1109/WETICE.2005.36>

Von-Krogh, Georg. "Care in knowledge creation". *California management review*, 1998, v. 40, n. 3, pp. 133-153.

Wiig, Karl. "Integrating intellectual capital and knowledge management". *Long rang planning*, 1997, v. 30, n. 3, pp. 399-405.
[http://dx.doi.org/10.1016/S0024-6301\(97\)90256-9](http://dx.doi.org/10.1016/S0024-6301(97)90256-9)

Woitsch, Robert; Karagiannis, Dimitris. "Process-oriented knowledge management systems based on KM-services: the Promote approach". *Intl journal of intelligent systems in accounting, finance & management*, 2002, v. 11, n. 4, pp. 253-267.
<http://dx.doi.org/10.1002/isaf.235>

Yim, Nam-Hong; Kim, Soung-Hie; Kim, Hee-Woong; Kwahk, Kee-Young. "Knowledge based decision making on higher level strategic concerns: system dynamics approach". *Expert systems with applications*, 2004, v. 27, n. 1, pp. 143-158.
<http://dx.doi.org/10.1016/j.eswa.2003.12.019>

Zack, Michael. "An architecture for managing explicated knowledge". *Sloan management review*, 1998, Sept, v. 39, n. 4, pp. 45-58.

Zhao, Jingyuan. "School knowledge management framework and strategies: the new perspective on teacher professional development". *Computers in human behavior*, 2010, v. 26, n. 2, pp. 168-175.
<http://dx.doi.org/10.1016/j.chb.2009.10.009>

Zollo, Maurizio; Winter, Sidney. "Deliberate learning and the evolution of dynamic capabilities". *Organization science*, 2002, v. 13, n. 3, pp. 339-351.
<http://dx.doi.org/10.1287/orsc.13.3.339.2780>

Si te interesan los

INDICADORES EN CIENCIA Y TECNOLOGÍA,
y todos los temas relacionados con la medición de la ciencia, tales como:

Análisis de citas, Normalización de nombres e instituciones, Impacto de la ciencia en la sociedad, Indicadores, Sociología de la ciencia, Política científica, Comunicación de la ciencia, Revistas, Bases de datos, Índices de impacto, Políticas de open access, Análisis de la nueva economía, Mujer y ciencia, etc.

Entonces **INCYT** es tu lista. Suscríbete en:

<http://www.rediris.es/list/info/incyt.html>



PROMOCIÓN DE LA LECTURA EN LAS BIBLIOTECAS PÚBLICAS DE CATALUÑA: EVALUACIÓN Y RESULTADOS



Mònica Baró, Teresa Mañà, Maite Barrios y Júlia Baena



Mònica Baró es profesora titular de la *Facultat de Biblioteconomia i Documentació* de la *Universitat de Barcelona (UB)*. Licenciada en historia, diplomada en biblioteconomía y doctora en documentación. Miembro del *Observatori de Biblioteques, Llibres i Lectura* de la *UB* desde su creación en 2009.

Universitat de Barcelona
Facultat de Biblioteconomia i Documentació
Melcior de Palau, 140. 08014 Barcelona
baro@ub.edu



Teresa Mañà es profesora titular de la *Facultat de Biblioteconomia i Documentació* de la *Universitat de Barcelona (UB)*. Licenciada en filología, diplomada en biblioteconomía y doctora en documentación. Ha sido directora del *Observatori de Biblioteques, Llibres i Lectura* de la *UB* desde su creación en 2009.

mana@ub.edu



Maite Barrios es profesora agregada del *Departamento de Metodología de las Ciencias del Comportamiento* de la *Universitat de Barcelona (UB)*. Licenciada y doctora en psicología. Desarrolla su actividad docente e investigadora en la *Facultad de Biblioteconomía y Documentación* de la *UB* desde 2002. Es miembro del *Observatori de Biblioteques, Llibres i Lectura* de la *UB* desde 2009.

mbarrios@ub.edu



Júlia Baena es técnica superior en diseño gráfico y diplomada en biblioteconomía y documentación. Realiza tareas de apoyo a la investigación del *Observatori de Biblioteques, Llibres i Lectura* y del grupo de investigación *Bescolar*. Colabora en tareas de documentación con el *Departament d'Ensenyament (Generalitat de Catalunya)* i el *Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya*.

julietabaena@gmail.com

Resumen

Se presentan los resultados de una encuesta realizada en 2009 entre las 320 bibliotecas públicas de Cataluña sobre las prácticas de evaluación de las acciones de promoción de la lectura. Los resultados se basan tanto en los datos cuantitativos como en las opiniones y las valoraciones del personal directivo de las bibliotecas en torno a estas actividades.

Palabras clave

Bibliotecas públicas, Evaluación, Promoción de la lectura.

Title: Reading promotion in public libraries in Catalonia: evaluation and results

Abstract

Results of a 2009 survey of the 320 public libraries in Catalonia that explored the practices of evaluating actions taken to promote reading. The results are based on both quantitative data and the opinions and assessments of library administrators concerning these activities.

Artículo recibido el 23-12-11

Aceptación definitiva: 29-02-12

Keywords

Public libraries, Evaluation, Reading promotion.

Baró, Mònica; Mañà, Teresa; Barrios, Maite; Baena, Júlia. "Promoción de la lectura en las bibliotecas públicas de Cataluña: evaluación y resultados". *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 277-281.

<http://dx.doi.org/10.3145/epi.2012.may.08>

Introducción

En un contexto de crisis en el que lo que cuenta son los resultados, cabe preguntarse sobre la evaluación de las acciones de fomento de la lectura en las bibliotecas públicas: ¿se evalúan? ¿cómo? ¿con qué finalidad?

El artículo trata de dar respuesta a las prácticas relativas a la evaluación de las acciones para la promoción de la lectura en las bibliotecas públicas de Cataluña a partir de las opiniones y valoraciones de los profesionales que las dirigen. Dichas prácticas se identificaron en el contexto de un estudio más amplio (Baró; Mañà, 2009) propuesto por la *Subdirección General de Bibliotecas* de la *Generalitat de Catalunya* al *Observatori de Biblioteques, Llibres i Lectura* de la *Facultat de Biblioteconomia i Documentació* de la *Universitat de Barcelona*, con el doble objetivo de 1) valorar los procesos de planificación, seguimiento y evaluación de la calidad de las actividades de promoción de la lectura llevadas a cabo por las bibliotecas públicas catalanas y 2) definir indicadores e instrumentos para que los centros pudieran valorar el impacto de estas acciones, establecer los procesos de evaluación y modificar sus propuestas en función de los resultados (Baró et al., 2010). Para realizar este estudio, se visitaron 21 bibliotecas y se entrevistaron bibliotecarios, usuarios y profesionales que trabajan en el ámbito de la promoción de la lectura. Además, se distribuyó un cuestionario online a todas las bibliotecas del sistema para obtener datos significativos, tanto cuantitativos como cualitativos, parte de cuyos resultados exponemos a continuación.

<http://www.ub.edu/obl>

Aun cuando en el entorno de la biblioteca pública la evaluación es una práctica cada vez más extendida que se aplica a distintos aspectos como la eficacia de los servicios, el uso de la colección y otros, la investigación y la literatura profesional específicas sobre los resultados de las acciones de promoción de la lectura son muy limitadas (Farmer; Stricevic, 2011), y suelen abordarse en un contexto más amplio de evaluación de servicios bibliotecarios en general. La bibliografía disponible proviene, esencialmente, de los países anglosajones y se relaciona con la evaluación de las acciones desarrolladas por las bibliotecas públicas en el contexto de los distintos planes gubernamentales de promoción de la lectura. En estos casos, algunos autores establecen la dificultad de evaluar la eficacia de dichas acciones en términos cualitativos tales como la modificación de hábitos lectores, incremento del nivel lector, etc. (Train; Elkin, 2001) y se centran en la evaluación de los procesos en sí mismos, valorando la eficacia de los métodos en relación con el cumplimiento de los objetivos estratégicos de dichos planes y de cada una de las acciones (Wilson; Train, 2004). En otros casos se evidencia la dificultad de contar con herramientas

fiabiles (y contrastables) para la recopilación de datos y el establecimiento de indicadores (Greenwood; Davies, 2004). En nuestro caso, la revisión bibliográfica permitió comprobar la ausencia de estudios propios y centrados en estos aspectos y establecer los aspectos clave que articularían el cuestionario.

Éste se elaboró y se distribuyó mediante un software de encuestas online y constaba de 21 preguntas. La encuesta se dirigió a principios de 2009 al total de las 348 bibliotecas públicas catalanas, de las cuales respondieron 197 (56,6%) en un plazo de dos meses. Este porcentaje de respuesta ofrece una elevada fiabilidad de los datos obtenidos.

Se han obtenido los indicadores descriptivos en función del tipo de variable (porcentaje o bien media y desviación típica) y se ha realizado una regresión logística para valorar el efecto de distintos factores sobre la práctica de evaluar todas o parte de las actividades de promoción de la lectura que se programan. Dichos factores son: actividades programadas por los servicios externos de apoyo a las bibliotecas, número de personas que trabajan en las bibliotecas, número total de actividades programadas, así como la interacción entre el número de personas que trabajan en las bibliotecas y el número de actividades programadas.

Evaluación de la promoción de la lectura

Como el conjunto de bibliotecas españolas, las bibliotecas públicas de Cataluña dedican gran parte de su tiempo y de sus energías en diseñar y desarrollar acciones de índole diversa que tienen como objetivo la promoción de la lectura. Según los datos estadísticos de la *Subdirección General de Bibliotecas*, en 2010 se realizaron 34.966 actividades en 348 bibliotecas públicas de Cataluña, lo que da una media de 100 actividades/centro. Sin embargo, según las respuestas de las 197 bibliotecas que contestaron a nuestra encuesta, la media de actividades fue de 75 por centro, lo que aun así supone más de una a la semana [mediana = 50, desviación típica = 72,14]. La variabilidad entre centros es muy elevada, tal como indica el valor de la desviación típica.

Aunque se observa una relación directa y moderada entre el número de personas que trabajan en una biblioteca y el número de actividades que se programan ($r = 0,49$; $p < 0,001$), parece que su evaluación no tiene relación ninguna con el volumen de personal, ni con el número de actividades realizadas ni, tampoco, con el hecho de que éstas se programen a partir de ofertas externas a la biblioteca.

Sobre la eficacia de dichas prácticas, planean dudas bien fundamentadas, máxime cuando los estudios evidencian que las prácticas lectoras de los españoles se sitúan en los puestos de cola. Pero lo cierto es que aunque la mayoría

de las bibliotecas públicas preguntadas considera necesario evaluar estas actividades (80,5%), ésta no es todavía una práctica generalizada y tan sólo un 48,4% (n = 93) de ellas manifiesta evaluar todas las actividades de promoción de la lectura que organizan, un 41,7% (n = 80) evalúa tan sólo algunas de ellas y un 9,9% (n = 19) no evalúa ninguna actividad.

El estudio muestra que no existen modelos ni indicadores específicos para evaluar las actividades de promoción de la lectura y que, en la mayoría de casos, suele recurrirse a la contabilización del público asistente como principal indicador de éxito (97,3%), aunque también se considera como indicador el grado de satisfacción de los usuarios (88,4%). Otros datos objetivos y cuantificables son utilizados por un número mucho menor de centros: el 38% contabiliza el número de préstamos de documentos relacionados con las actividades, y el 34% tiene en cuenta el aporte de nuevos públicos a la biblioteca. Todavía son menos (16,2%) los centros que computan el número de nuevos carnets realizados después de cada sesión.

En la figura 1 se muestran los indicadores que se utilizan para evaluar las actividades de promoción de la lectura y su frecuencia de uso.

Las bibliotecas procuran recabar otras informaciones cualitativas para determinar el éxito de dichas actividades. Así, para conocer la opinión de los usuarios, las conversaciones informales con los asistentes son el procedimiento más habitual. También se tienen en cuenta las valoraciones del personal de la biblioteca, pero no suelen obtenerse ni analizarse las valoraciones del conductor de las actividades realizadas, como se indica en la tabla 1.

La evaluación no se realiza de manera sistemática. Se preguntó sobre las causas que en opinión de los responsables de las bibliotecas obstaculizan dicha evaluación. En primer lugar se situaría la falta de instrumentos preparados, es-

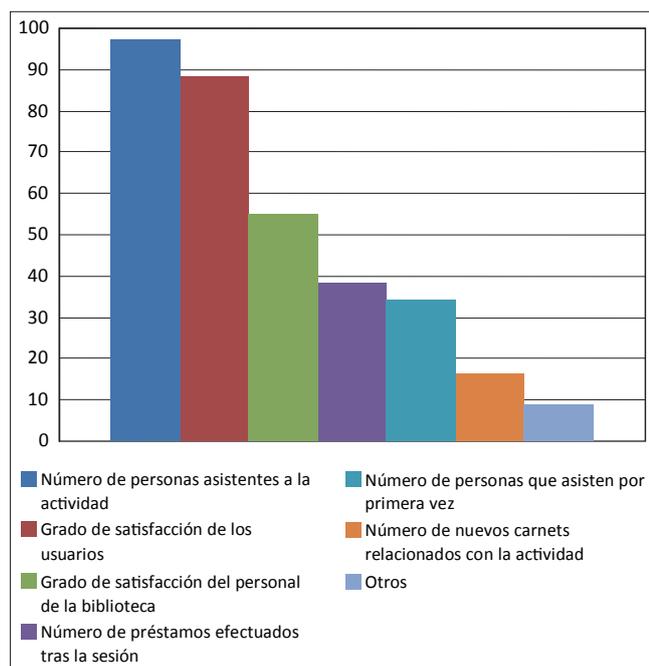


Figura 1. Frecuencia de uso de determinados indicadores para evaluar las actividades de promoción de la lectura

	Media	DT
Conversaciones informales con los usuarios	6,06	1,06
Valoraciones del personal de la biblioteca	5,87	1,38
Cuestionarios de evaluación del conductor de la actividad	3,25	2,17
Cuestionarios de satisfacción del público	3,03	1,97
Entrevistas formales con los usuarios	2,64	1,99

Escala de valoración de 1 a 7 puntos. A mayor puntuación, más de acuerdo con la afirmación (DT = desviación típica)

Tabla 1. Aplicación de distintos métodos para evaluar las actividades de promoción de la lectura

tandarizados y eficaces. El hecho de que los usuarios se muestren reacios a responder encuestas y cuestionarios es visto también como un impedimento aunque en menor medida. La dificultad de establecer relaciones causa-efecto entre las actividades y la promoción efectiva de la lectura es la razón aducida en algunos casos puesto que resulta difícil medir su eficacia en este sentido. Es destacable que algunos encuestados manifiestan claramente que no es necesario evaluar dichas actividades. Los resultados pueden verse en la tabla 2.

	Media	DT
No se dispone de instrumentos eficaces para la evaluación	4,32	1,93
Los usuarios son reacios a responder encuestas y cuestionarios	3,94	1,92
No se puede establecer una relación directa entre las actividades y la promoción real de la lectura	3,53	1,80
No es necesario evaluar este tipo de actividades	2,24	1,62

Escala de valoración de 1 a 7 puntos. A mayor puntuación, más de acuerdo con la afirmación (DT = desviación típica)

Tabla 2. Razones de la no evaluación de las actividades de promoción de la lectura

La relación de posibles causas de la no evaluación se ampliaba con preguntas abiertas. La razón más alegada era la falta de tiempo o de recursos humanos en las bibliotecas unipersonales (46,8%), seguida por la falta de instrumentos para medir y recoger datos (18,7%) y, aunque parezca sorprendente, había una cierta cantidad de respuestas que manifestaban que no habían pensado en ello (9,3%).

Beneficios y factores de éxito

El cuestionario preguntaba sobre aspectos de valoración de las actividades de promoción de la lectura por parte de las bibliotecas en términos de beneficios y eficacia. También sobre los factores que inciden en el éxito de las acciones de promoción de la lectura y los que lo dificultan.

Los profesionales consideraban que las actividades eran bien valoradas por los usuarios destinatarios, y esta valoración positiva se sumaba a los efectos beneficiosos que las prácticas de promoción podían implicar para las bibliotecas. Ante todo, las actividades ofrecen la posibilidad de crear alianzas con otros equipamientos y servicios culturales y educativos de la localidad. Otro beneficio que se le supone,

y en el que una gran parte de los encuestados está de acuerdo, es la posibilidad de atraer nuevos públicos a la biblioteca. Curiosamente, el incremento de préstamo no se halla entre los principales beneficios a pesar de que tiene una alta consideración. Por otro lado, se detecta un rastro de escepticismo en torno a estas actividades puesto que existe un cierto grado de acuerdo en que tienen un coste elevado, generan mucho trabajo y no cumplen directamente con sus objetivos. Los resultados se presentan en la tabla 3.

Actividades de promoción de la lectura	Media	DT
Son bien valoradas por los usuarios	5,82	1,09
Permiten crear alianzas con el entorno cultural local	5,81	1,15
Atraen nuevos públicos	5,06	1,37
Generan más préstamos	4,80	1,36
Son un buen instrumento a pesar de que los resultados son difíciles de medir	4,62	1,48
Gravan excesivamente el presupuesto	3,93	1,65
Sirven para dar a conocer la biblioteca, pero no para fomentar la lectura	3,56	1,66
Generan mucho trabajo para unos resultados inciertos	3,33	1,56

Escala de valoración de 1 a 7 puntos. A mayor puntuación, más de acuerdo con la afirmación (DT = desviación típica)

Tabla 3: Grado de acuerdo con las aportaciones de las actividades de promoción de la lectura

Las actividades, en su conjunto, reportan beneficios pero ¿cuándo consideran las bibliotecas que una actividad obtiene un buen resultado? Y, ¿cuáles son las claves para el éxito?

Las bibliotecas recurren a los más variados sistemas para evaluar las actividades de promoción lectora y generan sus propias escalas de valoración de los resultados de las actividades. Hay unanimidad en considerar la satisfacción del público como el valor más aceptado. También se considera positivo que los usuarios hayan solicitado la repetición de aquella actividad o que una actividad haya atraído nuevo público. Otros indicadores que suscitan un grado de acuerdo elevado tienen que ver con los efectos generados por la actividad, como el incremento del préstamo de materiales más directamente relacionados con el tema de la misma, o el hecho de que el número de asistentes alcance o se aproxime a lo esperado. Otras consideraciones, como que se hayan movilizado los documentos relacionados con la actividad o se hayan registrado desideratas en el mismo tema, no parecen suscitar tanto acuerdo. En la tabla 4 se muestran las opciones ofrecidas en esta pregunta.

Se considera que la adecuación al público al que se dirige constituye el factor principal de éxito de estas actividades; con una media similar se sitúan, por un lado, que el responsable ejecutor de la actividad tenga determinadas cualidades, como empatía, seguridad, dominio del tema, etc., que le permitan conectar y atraer al público, y por otro, el horario de la actividad. Les sigue la relación con el entorno, ya sea porque se tratan temas o autores locales o por la fama del invitado. Otros factores que parecería que deberían influir en el éxito de público como la gratuidad, la originalidad de

Cuando...	Media	DT
...el público ha salido contento	6,29	0,99
...el público ha solicitado repetirla	6,25	0,90
...ha atraído nuevo público a la biblioteca	6,11	0,91
...ha dinamizado el fondo más relacionado con la propuesta	5,83	1,04
...se ha conseguido reunir el número de personas esperado	5,67	1,04
...se han recogido desideratas relacionadas con el tema	5,06	1,53
...la acción se ha desarrollado sin incidentes	4,34	1,57
...se ha ajustado al presupuesto	4,30	1,43

Escala de valoración de 1 a 7 puntos. A mayor puntuación, más de acuerdo con la afirmación (DT = desviación típica)

Tabla 4. ¿Cuándo considera que una acción de promoción de la lectura ha tenido éxito?

la propuesta o que ésta invite a la participación del público, aun siendo importantes, no se consideran determinantes para el éxito. En la tabla 5 se exponen los distintos factores de éxito y su valoración por parte de los centros.

	Media	DT
Adecuación al público	6,14	0,78
Cualidades del conductor (empatía, seguridad, conocimiento...)	6,04	0,99
Horario de programación	6,01	0,93
Relación de la actividad con el entorno (temas/autores locales, implicación de las entidades...)	5,91	1,05
Fama del conductor (mediático, local...)	5,89	1,15
Temática (actualidad, interés)	5,88	1,03
Público a quien se dirige (infantil, jóvenes, adultos)	5,86	1,07
Medio de difusión que se utiliza	5,78	1,07
Gratuidad	5,76	1,29
Implicación de entidades locales	5,34	1,49
Adecuación del espacio donde se realiza	5,34	1,18
Grado de implicación por parte del lector (comentario de lecturas, participación en talleres...)	5,31	1,18
Variedad	5,10	1,21
Originalidad	4,93	1,24

Escala de valoración de 1 a 7 puntos. A mayor puntuación, más de acuerdo con la afirmación (DT = desviación típica)

Tabla 5. Grado en que estos factores parecen determinantes para el éxito de una actividad

¿Qué es lo que de verdad funciona? Larga vida a los clubs de lectura

Como información complementaria, se preguntaba a los encuestados qué tipo de propuesta o de formato de actividad consideraban más eficaz para la promoción de la lectura en las bibliotecas. Las 25 posibilidades planteadas reciben todas una puntuación por encima del valor medio de la escala, y con una DT no superior a 1,5, lo que indica que los profesionales consideran que la promoción lectora requiere una gran variedad de propuestas dirigidas a los distintos

públicos y que, en alguna medida, su eficacia deriva de la complementariedad. Aun así destaca la unanimidad en considerar a los clubs de lectura (para adultos, pero también los destinados a jóvenes y niños) los formatos más idóneos para la promoción de la lectura. A continuación, pero con una cierta diferencia, se sitúan los encuentros con autores y las actividades de narraciones de cuentos para niños (no así las narraciones para adultos, consideradas mucho menos eficaces). Se citan otras acciones cuya eficacia se considera menor: los concursos literarios, el *bookcrossing* o práctica de dejar libros en lugares públicos —en este caso las bibliotecas— para que los recojan otros lectores, las actividades relacionadas con proyecciones cinematográficas (cine forum) y los talleres no específicamente relacionados con lectura y escritura. En estos casos, existen más de dos puntos de diferencia con la actividad mejor valorada, como puede verse en la tabla 6.

Conclusión

Las bibliotecas tienen una gran confianza en los beneficios que pueden generar las actividades de promoción a la lectura aunque en muy pocos casos miden su impacto. El hecho de no disponer de instrumentos y mecanismos para recabar indicadores que les permitan evaluar de manera más o me-

nos objetiva las acciones de promoción de la lectura no es un freno a estas prácticas.

Por otro lado, no parece que los métodos de evaluación de la práctica lectora incorporen elementos de mayor calado como la frecuencia de la lectura, la comprensión lectora, el desarrollo de la capacidad crítica, la autonomía en la selección, etc. limitándose al cómputo cuantitativo (número de público asistente, número de libros prestados) y a las encuestas de satisfacción.

Aunque la bibliografía profesional ya advierte de la dificultad de llevar a cabo estos procesos de evaluación, es innegable que se debería poder ensayar un protocolo que, al menos, determinara hasta qué punto las acciones emprendidas por la biblioteca tienen efectos (aunque mínimos o difícilmente cuantificables) para una mejora de la práctica lectora y cuáles son simplemente acciones para acercar la biblioteca a sus usuarios. Evidentemente, la mejora será a largo plazo y esperamos que los estudios que pretendan cuantificar la evaluación de estas acciones evidencien estos cambios, algo que tras más de treinta años de promoción lectora no ha sucedido.

Bibliografía

Baró, Mónica; Casas-Poves, Juana; Mañà, Teresa; Reyes-Camps, Lourdes. “La promoció de la lectura a les biblioteques públiques de Catalunya: un terreny pròsper que cal endreçar”. En: *12 Jornades catalanes d’informació i documentació*, 2010.

http://www.cobdc.org/jornades/12JCD/materials/comunicacions/REYES_promocio_lectura_BBPP.pdf

Baró, Mónica; Mañà, Teresa. *Estudi sobre les accions de foment de la lectura a les biblioteques públiques de Catalunya*. Barcelona: DG de Coop. Cultural, Generalitat de Catalunya, 2009.

http://www20.gencat.cat/docs/Biblioteques/Tematic/Documents/Arxiu/Activitats/Estudi_foment_lectura.pdf

Farmer, Lesley; Stricevic, Ivanka. *El uso de la investigación para promover la alfabetización y la lectura en las bibliotecas: guía para los bibliotecarios*. La Haya: IFLA, 2011.

<http://www.ifla.org/files/hq/publications/professional-report/126.pdf>

Greenwood, Helen; Davies J. Eric. “Designing tools to fill the void: a case study in developing evaluation for reading promotion projects”. *Performance measurement and metrics*, 2004, v. 5, n. 3, pp. 106-111.

<http://dx.doi.org/10.1108/14678040410570120>

Train, Briony; Elkin, Judith. “Measuring the unmeasurable: reader development and its impact on performance measurement in the public library sector”. *Library review*, 2001, v. 50, n. 6, pp. 295-304.

<http://dx.doi.org/10.1108/EUM000000005598>

Wilson, Kerry; Train, Briony. *Reader development evaluation framework*. Sheffield: The University of Sheffield, The British Council, 2004.

<http://www.shef.ac.uk/content/1/c6/07/01/24/CPLIS%20-%20British%20Council%20RD%20Evaluation%20Framework.pdf>

	Media	DT
Clubs de lectura para adultos	6,30	0,91
Encuentros con autores	5,87	0,98
Clubs de lectura para jóvenes	5,82	1,38
Narraciones de cuentos para niños	5,76	1,15
Clubs de lectura para niños	5,75	1,43
Exposición de novedades	5,73	1,04
Intercambio de opiniones con el bibliotecario	5,65	1,18
Servicios de alerta de novedades	5,49	1,13
Visitas infantiles y escolares	5,47	1,19
Presentaciones de libros	5,22	1,27
Exposiciones temáticas de libros	4,98	1,40
Opiniones y recomendaciones por escrito de los lectores	4,98	1,21
Blog con valoraciones y opiniones de los lectores	4,92	1,33
Conferencias, charlas y mesas redondas	4,85	1,26
Rutas literarias	4,85	1,23
Guías de lectura	4,75	1,25
Espectáculos (representaciones teatrales, marionetas...)	4,71	1,50
Talleres de escritura	4,66	1,35
Lecturas en voz alta por parte de los lectores	4,65	1,39
Narraciones orales para adultos	4,60	1,32
Actos relacionados con las exposiciones	4,44	1,29
Concursos literarios	4,34	1,45
Bookcrossing	4,30	1,47
Cine forum	4,26	1,39
Talleres diversos	4,16	1,31

Escala de valoración de 1 a 7 puntos. A mayor puntuación, más de acuerdo con la afirmación (DT = desviación típica)

Tabla 6. ¿En qué grado estas acciones fomentan la lectura?

Mi **Biblioteca**

Actualidad bibliotecaria

Fotografía momentos de lectura

Opinión

Recomendaciones de lectura
Escolares

Públicas

Reportaje

Entrevistas

Recursos Universitarias

Con pasado

Cartas a Mi Bil

Bibliotecarios

Con futuro

Noticias

Agenda de Formación

Nuevas Tecnologías

Gestión documental

www.mibiblioteca.org



CLASIFICACIÓN DE INICIATIVAS DE CROWDSOURCING BASADA EN TAREAS



Enrique Estellés-Arolas y Fernando González-Ladrón-De-Guevara



Enrique Estellés-Arolas es licenciado con premio extraordinario en ingeniería informática (2007), e investiga el fenómeno del *crowdsourcing* en el *Departamento de Organización de Empresas* de la *Universidad Politécnica de Valencia*. Otras áreas de su interés son las nuevas tecnologías aplicadas a la educación, las herramientas colaborativas web 2.0 y el uso de las etiquetas para marcar recursos.

*Departamento de Organización de Empresas
Universidad Politécnica de Valencia
Camino de Vera, s/n. 46022 Valencia
kikeestelles@gmail.com*



Fernando González-Ladrón-De-Guevara, doctor ingeniero industrial, es profesor de sistemas de información y gestión de operaciones en el *Depto. de Organización de Empresas* de la *Univ. Politécnica de Valencia*. Sus áreas de interés incluyen el e-learning y la implantación de sistemas de información. Ha realizado estancias docentes en la *Universidades de Ansbach* (Alemania), *Santander* (Colombia), *San Martín de Porres* (Perú), y en la *Escuela de Negocios Sustentare* (Brasil).

*Departamento de Organización de Empresas
Universidad Politécnica de Valencia
Camino de Vera, s/n. 46022 Valencia
fgonzal@omp.upv.es*

Resumen

Las iniciativas de *crowdsourcing* planteadas por organizaciones de ámbitos diversos como la música, el diseño o la catalogación son cada vez más frecuentes. A pesar de este auge, la ausencia de un fundamento teórico consistente genera problemas como la existencia de tipos o clasificaciones de *crowdsourcing* que se superponen y entremezclan o la ausencia de una definición compartida. A partir de una revisión sistemática de la bibliografía se analizan las tipologías considerando la naturaleza de las tareas que debe realizar la 'multitud' como criterio, y se propone una nueva tipología integradora.

Palabras clave

Crowdsourcing, Tipología, Clasificación, Multitud, Tarea.

Title: Tasks-based classification of *crowdsourcing* initiatives

Abstract

Crowdsourcing initiatives by organizations working in areas like music, design or cataloguing are becoming more frequent. Nonetheless, the absence of a consistent theoretical background creates problems, such as the existence of diverse crowdsourcing classifications that overlap and interweave, or the lack of a common definition. This paper analyses different typologies considering the nature of the tasks to be performed by the crowd as the main criterion and proposes a new integrative typology.

Keywords

Crowdsourcing, Typology, Classification, Crowd, Task.

Estellés-Arolas, Enrique; González-Ladrón-De-Guevara, Fernando. "Clasificación de iniciativas de *crowdsourcing* basada en tareas". *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 283-291.

<http://dx.doi.org/10.3145/epi.2012.may.09>

1. Introducción

El *crowdsourcing* hace referencia a un conjunto de iniciativas de tipo participativo que se nutren de otros fenómenos como la innovación abierta (Chesbrough, 2003) o la inteligencia colectiva (Schenk; Guittard, 2011).

El periodista americano **Jeff Howe** lo definió en 2006 como una convocatoria abierta iniciada por una empresa o institución –normalmente realizada por un empleado– dirigida a un grupo de individuos indefinido (“la multitud” o *crowd*), con frecuencia grande (Howe, 2006) con el fin de externalizar una función. Diversos autores han tratado hasta el momento de elaborar una definición, unos centrándose en el uso del *crowdsourcing* como un proceso de resolución de problemas (Brabham, 2008b; Vukovic, 2009), otros como una forma de externalizar tareas (Oliveira; Ramos; Santos, 2009) o como una manifestación particular de la “innovación abierta” (Sloane, 2011). Estellés y González (2012) presentan una definición que permite identificar cualquier tipo de iniciativa de *crowdsourcing* en base a 8 elementos: tarea concreta a realizar, multitud que participará con sus aportaciones (*crowdworkers*), beneficio para dicha multitud, iniciador, beneficio para el iniciador, utilización de un proceso participativo, uso de una convocatoria abierta y flexible, y uso de internet como infraestructura fundamental.

“Crowdsourcing es una iniciativa participativa de innovación abierta o de inteligencia colectiva”

El *crowdsourcing* se lleva a cabo en internet, con el apoyo de las aplicaciones web 2.0 que facilitan la conexión de miles de usuarios que comparten información y resuelven problemas de forma colaborativa (Burger-Helmchen; Pénin, 2010; Vukovic; Bartolini, 2010a).

Las tareas que realizan los colaboradores pueden abarcar desde la catalogación de documentos hasta la innovación que mejora un proceso o un bien. Atendiendo a su complejidad, pueden ser de tres tipos:

- simples, normalmente repetitivas, que no requieren de un nivel cognitivo alto, como por ejemplo el etiquetado de una imagen;
- complejas que necesiten de una capacidad intelectual y de inventiva mayor, como la resolución de un problema de una empresa; y
- creativas, donde la singularidad de la aportación del usuario es fundamental, como en el diseño de un logo (Schenk; Guittard, 2009).

En muchos casos se trata de tareas modulables, hecho que posibilita su realización por varios usuarios en paralelo, produciendo un ahorro económico y de tiempo (Mazzola; Distefano, 2010; Kleeman; Voss; Rieder, 2008). Por todo ello, empresas como *Doritos* (*SuperBowl*, 2011), organizaciones públicas como la *Unión Europea* (ECMT, 2011) o incluso individuos aislados, como el músico español Carlos Jean (*PlanB*, 2011), se interesan por el potencial del *crowdsourcing* (Howe, 2008; Vukovic; Bartolini, 2010a).

Sin embargo el *crowdsourcing* no dispone de una base teórica que fundamente su estudio (Denyer; Tranfield; Van-Aken, 2008), aunque este problema está subsanándose. Ya existen puntos de acuerdo entre los autores, como que todas las iniciativas de este tipo deben tener, como mínimo, dos elementos: una multitud a priori indefinida y heterogénea (Geerts, 2009; Schenk; Guittard, 2009) y la utilización de una llamada abierta a todo el mundo (Pénin, 2008; Geerts, 2009; Burger-Helmchen; Pénin, 2010), coincidiendo con los elementos enumerados por Estellés y González (2012). Se han generado también diversas clasificaciones basadas en criterios como la perspectiva organizativa (Geiger; Seedorf; Schader, 2011).

El presente artículo pretende participar en la creación de esta base teórica: su objetivo es aportar una tipología del *crowdsourcing* basada en la tarea a realizar por parte de la multitud. Para ello se analizan tipologías planteadas por diversos autores, identificando divergencias y puntos en común y definiendo una tipología integradora.

2. Metodología

El trabajo se ha llevado a cabo en tres fases: 1) revisión sistemática de la bibliografía existente (Delgado-Rodríguez; Doménech; Llorca, 2010), 2) creación de un repositorio documental con los documentos hallados, y 3) descripción de sus categorías, ilustrándolas con ejemplos, comparándolas y detectando relaciones entre ellas. Para este fin se ha elaborado e interpretado una parrilla de análisis (Codina, 1997; Pinto-Molina et al., 2007).

2.1. Revisión sistemática: búsqueda de información

Se han realizado consultas en siete bases de datos: *ACM*, *Scopus*, *Emerald*, *SAGE*, *Wiley*, *SpringerLink* y *ScienceDirect*. El criterio de selección ha sido la ocurrencia del término *crowdsourcing* tanto en el título como en las palabras clave del documento. Se han obtenido 151 documentos. Una búsqueda adicional en *Google Scholar* por “classification of crowdsourcing” OR “crowdsourcing classification” ha permitido obtener nueve documentos, a partir de las bibliografías de los cuales se han hallado 28 documentos más. La composición del repositorio documental obtenido se describe en la tabla 1. La mayoría de los documentos (66%) aparecen en actas de congresos, lo que sugiere el carácter preliminar de la investigación existente sobre este objeto de conocimiento.

2.2. Filtrado de documentos

Se han descartado documentos que no aportan clasificación alguna de las iniciativas de *crowdsourcing*, en total once do-

Tipo de documento	Búsqueda en bases de datos	Búsqueda en Google Scholar	Total
Ponencias en congresos	108	11	119
Artículos en revista	43	11	54
Monografías	0	1	1
Otros	0	5	5
Total	151	28	179

Tabla 1. Composición del repositorio documental

cumentos. De estos, se han rechazado los que no utilizan el criterio del tipo de tarea sino otros como la naturaleza del *crowd* (Schenk; Guittard, 2009) o la recompensa (Corney et al., 2009). También se han descartado tipologías centradas en un área o subsector específico, como la de Ooman y Aroyo (2011) focalizada en galerías de arte, librerías, archivos, etc.; la de La-Vecchia y Cisternino (2010), centrada en modelos de negocio; y la de Geiger, Seedorf y Schader, (2011) con una perspectiva organizativa. Finalizado el proceso de filtrado se han obtenido seis documentos que cumplieran los requerimientos citados.

3. Descripción de las tipologías

A continuación se presentan ordenadas cronológicamente las tipologías encontradas, junto con ejemplos que las ilustran. Cada uno de sus subtipos estará identificado por un código que se utilizará después en la parrilla de análisis de subtipos.

3.1. Reichwald & Piller (2006)

Agrupar las tareas de *crowdsourcing* con dos enfoques:

1. Innovación abierta (RP1). Tareas de cooperación entre la empresa iniciadora y sus clientes en la elaboración de nuevos productos y que suponen la generación de conocimiento.
2. Actividades operativas de soporte (RP2). Mejora de procesos operativos para la personalización masiva de bienes (Heizer; Render, 2010).

Comprenden desde tareas sencillas que requieren un nivel cognitivo bajo, como la búsqueda de información en internet, hasta tareas complejas que exigen competencias específicas como la búsqueda de una solución para un problema científico, pasando por tareas creativas, como el diseño de un logo.

3.2. Howe (2008)

Propone los siguientes tipos de tareas (Howe, 2006):

1. De inteligencia colectiva (*crowdwisdom*) con 3 subtipos:
 - Predicción de mercados (H1.1): una comunidad de inversores particulares votan diversas alternativas a partir de la información descriptiva suministrada, como en el caso de *Iowa Electronic Markets*.
 - Competición (*crowdcasting*) (H1.2): se recompensa a quien resuelva un desafío, como los planteados en la plataforma *Innocentive*.
<http://www.innocentive.com>
 - Tormenta de ideas online (*crowdstorming*) (H1.3): similar a la que se produjo en el proyecto *Idea Jam* de IBM.
2. Creativas (*crowdproduction*) (H2). El promotor de la iniciativa (*crowdsourcer*) externaliza actividades que necesitan la energía creativa de los colaboradores para obtener un nuevo producto o servicio (una base de datos o cualquier tipo de contenido generado por los usuarios). *Wikipedia* o *iStock* son ejemplos de este tipo de tareas.
3. Recoger opiniones de los usuarios (*crowdvoting*) (H3). Un ejemplo característico es *Threadless*, empresa de camisetas que pide a los usuarios que elijan sus diseños preferidos para lanzarlos al mercado.

4. Obtención de fondos económicos (*crowdfunding*) (H4). Se pide una cantidad determinada de dinero a cambio de una recompensa: *MyFootballClub* es un club de fútbol en el que, a cambio de una cuota anual, los inversores deciden sobre el fichaje de jugadores o el precio de las entradas.

Algunos autores consideran esta clasificación como no demasiado rigurosa, dado el solapamiento de los tipos planteados (Geerts, 2009).

3.3. Kleeman, Voss & Rieder (2008)

Comprende siete tipos:

1. Participación de consumidores en el desarrollo colaborativo de un producto, como en el proyecto *Fiat Mio* o en *Idea Storm* de Dell (K1).
2. El diseño de un nuevo producto, que depende casi por completo de las aportaciones de los usuarios. Así ocurre en *Spreadshirt* o *Fluevog* (K2).
3. Ofertas competitivas sobre ciertas tareas o problemas bien definidos, similares a *Innocentive* (K3).
4. Llamadas abiertas permanentemente donde los *crowdworkers* presentan información o documentación a lo largo de un período indeterminado de tiempo, como en el caso de *iReport*. Esta es una iniciativa en la que la *CNN* ha puesto un conjunto de herramientas online a disposición de cualquier reportero aficionado para recoger imágenes (K4).
5. Periodismo de comunidad (*community reporting*): los usuarios informan sobre nuevos productos o tendencias sobre algún tipo de comunidad online, como ocurre con *Trendwatching.com* (K5).
6. Valoración de productos por consumidores o perfiles de consumidor. Un ejemplo es *Amazon Reviews* (K6).
7. Soporte de clientes. Los usuarios de un servicio resuelven los problemas y las dudas de otros usuarios, como en la *Universidad de Indiana* donde el teléfono 24 horas de asistencia técnica ha sido sustituido por un foro que tanto empleados como usuarios utilizan para resolver las dudas planteadas (K7).

“Crowdcasting (EG1): un individuo, empresa u organización plantea a la multitud un problema o tarea, siendo recompensado quien lo resuelva antes o mejor”

3.4. Brabham (2008a)

Propone una clasificación de tareas para resolver problemas que comprende cuatro grupos con distintos objetivos:

1. Descubrimiento y gestión de conocimiento (B1). Su objetivo es encontrar y reunir de manera coherente cierto conocimiento disperso. Un ejemplo es el proyecto *Peer to Patent - Community Patent Review*. En él, una comunidad online se encarga de informar sobre patentes existentes que puedan estar relacionadas con solicitudes a la oficina de patentes de EUA (Ghafele; Gibert; DiGiammarino, 2011).

2. Obtención de una respuesta correcta (B2). Se difunde un problema del que se busca una solución, esperando que la aporte un experto que podría encontrarse en la Red. Estas tareas se dan en plataformas como *Innocentive*, que permite difundir problemas de I+D con carácter científico a un conjunto de especialistas.

3. Diseño y valoración de productos por parte de los usuarios (B3). Este tipo de tareas son útiles cuando se desea conocer la opinión o las preferencias de los usuarios sobre un producto. Un ejemplo es la citada empresa *Threadless*.

4. Participación distribuida (B4). Las tareas son realizadas por una comunidad online y suelen implicar el procesamiento de grandes cantidades de datos. Un ejemplo son las propuestas en *Amazon Mechanical Turk*, plataforma donde cualquier empresa puede contratar una comunidad de usuarios para realizar trabajos, como la traducción de textos o la indización de imágenes.

<http://www.mturk.com>

3.5. Geerts (2009)

Propone cuatro tipos que tienen como punto de partida la clasificación de **Howe** (2008):

1. *Crowdcasting* (G1). Un grupo de usuarios compiten para obtener una recompensa al proporcionar la mejor solución a un problema, como en *Innocentive*.

2. *Crowdstorming* (G2). Mediante foros, como los de *Ideastorm* de *Dell*, los participantes discuten, preguntan o proponen enfoques alternativos para resolver un problema de forma colectiva. Se suelen combinar distintas aportaciones para obtener el resultado final.

3. *Crowdproduction* (G3). Los colaboradores tienen como objetivo la elaboración conjunta de un bien: una base de datos para investigación, el contenido de una wiki, el etiquetado de recursos en red, etc.

4. *Crowdfunding* (G4). El objetivo de los colaboradores es invertir. **Geerts** (2009) distingue entre dar dinero a través de mediadores o por iniciativa individual. En el primer caso, aparecen plataformas como *Kiva*, que permite financiar a empresarios del tercer mundo. En las iniciativas particulares la multitud suele ser recompensada participando en la toma de decisiones relevantes, como en el equipo de fútbol *MyFootballClub*.

3.6. Burger-Helmchen & Pénin (2010)

Se proponen tres tipos de tareas, difíciles de distinguir en algunos casos:

1. Tareas innovadoras (BH1). La multitud constituye tan sólo un revestimiento adicional que no resuelve los problemas. Para la empresa iniciadora es más importante recoger el conocimiento de un pequeño número de especialistas en diferentes campos que la participación de un número elevado de profanos (**Pisano; Verganti**, 2008).

2. Tareas rutinarias (BH2). Se trata de tareas modulares que no requieren competencias específicas, únicamente implican el uso de tiempo, como por ejemplo, la búsqueda de direcciones de correo electrónico de un determinado segmento de clientes para una actividad de marketing electrónico. Los colaboradores aportando su tiempo, información y capacidad de cómputo contribuyen a disminuir costes y aumentar la velocidad de ejecución de la tarea. Aquí el tamaño de la multitud sí importa: cuanto más grande, más tareas podrán ser ejecutadas de forma paralela y en menos tiempo.

3. Tareas de contenido (BH3). Los colaboradores aportan su tiempo, información y capacidad de cómputo para generar servicios basados en información (*Wikipedia*). Además del tamaño de la *crowd*, importa su heterogeneidad y diversidad.

4. Comparación de tipologías y desarrollo de una nueva

En la tabla 2 se presenta una parrilla de análisis que compara cada uno de los componentes de las tipologías con el resto. Se trata de una matriz no simétrica que leída por filas, destaca, para cada caso, los tipos (columnas) que no están representados. Es decir, la celda [2,3] (fila 2, columna 3) indica los elementos H3 y H4 de la tipología **Howe** que no están incluidos o representados en la tipología **Reichwald & Piller**.

“*Crowdcollaboration* (EG2): se produce comunicación entre los individuos de la multitud y la empresa iniciadora del proceso queda al margen”

Lo primero que se observa es que la tipología de **Reichwald & Piller** (2006), al ser tan genérica —y la más antigua— no abarca muchos de los elementos del resto y, además, su tipo RP2 no aparece en tres de las tipologías consideradas. Por todo esto, no va a ser tenida en cuenta en el análisis comparativo (fondo azul).

	Reichwald & Piller	Howe	Kleeman	Brabham	Geerts	Burger-Helmchen y Pénin
Reichwald & Piller		H3 y H4	K4, K5, K6 y K7	B1 y B4	G4	BH2 y BH3
Howe	RP2		K7	B4		BH1 y BH2
Kleeman	RP2	H4		B4	G4	
Brabham		H4	K7		G2 y G4	
Geerts		H3	K7	B1		
Burger-Helmchen & Pénin	RP2	H4			G2 y G4	

Tabla 2. Comparación de tipologías: elementos distintivos faltantes

Con respecto al resto, se pueden observar algunos elementos que, aun siendo actividades de *crowdsourcing*, no son tenidos en cuenta por todos los autores. Destacan fundamentalmente dos: el *crowdfunding* (H4, G4) (no presentes en **Kleeman, Brabham y Burger-Helmchen y Penin**) y el soporte entre clientes (K7), diferenciándolo del *crowdstorming* (**Howe, Brabham y Geerts**). El resto de los elementos analizados coinciden en mayor o menor grado, siendo las tipologías de **Brabham y Geerts** las que presentan un mayor carácter integrador.

El *crowdcontent* (EG3) se diferencia del *crowdcasting* en que no es una competición, sino que cada individuo trabaja de manera individual y al final se reúne el resultado de todos

En base a los resultados anteriores y a la revisión bibliográfica realizada se propone la siguiente clasificación que intenta recoger las aportaciones y subsanar las carencias de las anteriores. Para cada tipo planteado se mencionan los elementos previos relacionados:

1. *Crowdcasting* (EG1). En este tipo de iniciativas, un individuo, empresa u organización plantea a la multitud un problema o tarea, siendo recompensado quien lo resuelva antes o mejor. *Innocentive* es un ejemplo paradigmático: en esta plataforma se permite la propuesta de tareas como la elaboración de un tratamiento que permita reducir el coeficiente de fricción en las piezas de metal hechas de acero inoxidable, premiando dicha propuesta con 10.000 dólares (**Doan; Ramakrishnan; Halevy, 2011**). Este nuevo tipo engloba a H1.2, K3, B2, G1 y BH1.

2. *Crowdcollaboration* (EG2). Considera las iniciativas en las que se produce una comunicación entre los individuos de la multitud, mientras la empresa iniciadora del proceso queda relativamente al margen. Los individuos aportarán su conocimiento para resolver problemas o plantear ideas de forma colaborativa (**Doan; Ramakrishnan; Halevy, 2011**) y normalmente no existe una recompensa económica.

Podemos encontrar dos subtipos que se diferencian en el objetivo final:

a. *Crowdstorming* (EG2.1). Sesiones de tormenta de ideas online, en las que se plantean soluciones y la multitud participa con sus comentarios y votos, como en el caso de la plataforma *Ideajam*. Este subtipo está relacionado con H1.3, K1, K2, G2 y BH1.

<http://www.ideajam.net>

b. *Crowdsupport* (EG2.2). Los propios clientes son los que solucionan las dudas o problemas de otros, sin necesidad de acudir al servicio técnico o posventa de atención al cliente. La diferencia con EG2.1 es que el *crowdsupport* busca ayudar, como es el caso de *Getsatisfaction*, una plataforma que permite a compañías como *Microsoft* realizar este tipo de tareas. Este subtipo incorpora el tipo K7.

<http://www.getsatisfaction.com>

3. *Crowdcontent* (EG3). La gente aporta su mano de obra y su conocimiento para crear o encontrar contenido de diversa naturaleza (**Doan; Ramakrishnan; Halevy, 2011**). Se diferencia del *crowdcasting* en que no es una competición, sino que cada individuo trabaja de manera individual y al final se reúne el resultado de todos. Se pueden encontrar tres subtipos que se diferencian en su relación con los contenidos:

a. *Crowdproduction* (EG3.1). La multitud debe crear contenido colaborando con otros, como en el caso de la *Wikipedia*, o de manera individual, realizando tareas de dificultad variable como la traducción de fragmentos cortos de texto o el etiquetado de imágenes, como en algunas tareas de *Amazon Mechanical Turk*. Este subtipo incorpora a H2, K4, K5, B3, B4, G3, BH2 y BH3.

b. *Crowdsearching* (EG3.2). Los colaboradores buscan contenidos disponibles en internet con algún fin. Aunque existen proyectos que se basan en este tipo de tareas, como el *Peer to Patent Review*, también existen otras de menor tamaño como algunas planteadas en *Amazon Mechanical Turk*. Este subtipo contempla los tipos B1 y BH2.

<http://www.peertopatent.org>

c. *Crowdanalyzing* (EG3.3). Es parecido al *crowdsearching* (EG3.2), con la diferencia de que la búsqueda no se realiza en internet, sino en documentos multimedia como imágenes o vídeos. Un ejemplo sería el proyecto *stardust@home*, en el que cualquier persona puede buscar muestras de polvo interestelar, analizando imágenes en 3 dimensiones de la sonda espacial *Stardust*. Este subtipo surge de los mismos tipos que el *crowdsearching*, refinado tras la consulta de los artículos recogidos en la revisión sistemática de la bibliografía.

4. *Crowdfunding* (EG4). Un individuo o una organización buscan la financiación por parte de la multitud a cambio de alguna recompensa. En el mundo del cine, por ejemplo, la película española "El cosmonauta" está siendo financiada de esta manera: los productores ofrecen a los que les financian promoción comercial o aparecer en los títulos de crédito. En el mundo del deporte destaca el caso del equipo inglés de fútbol *Myfootballclub*. En este caso, la multitud participa aportando su dinero. Este tipo abarca H4 y G4.

<http://www.elcosmonauta.es>

<http://www.myfootballclub.co.uk>

Crowdfunding (EG4): un individuo o una organización buscan la financiación por parte de la multitud a cambio de alguna recompensa

5. *Crowdopinion* (EG5). Se intenta conocer la opinión de los usuarios sobre un tema o producto. Es el caso de *Modcloth*, tienda de ropa inglesa donde cualquier usuario registrado puede opinar sobre productos que todavía no han salido a la venta, obteniéndose así información sobre su potencial aceptación en el mercado. La gente aporta su opinión o criterio para realizar valoraciones (**Doan; Ramakrishnan; Halevy, 2011**). Este tipo se corresponde con H3, K6 y B3. También se correspondería con H1.1, subtipo que **Howe** (2008) denomina *market research*. En este caso se trata de

iniciativas de *crowdvoting* donde la opinión del usuario no se manifiesta mediante un voto, sino mediante la compra y venta de acciones vinculadas a algún resultado próximo como una elección presidencial. Para este tipo de iniciativas de *crowdvoting* se utilizan plataformas especializadas llamadas “mercados de predicción online”, como *Intrade* o *inkling markets*.

<http://www.modcloth.com>

<http://www.intrade.com>

<http://inklingmarkets.com>

‘ **Crowdopinión (EG5):** se intenta conocer la opinión de los usuarios sobre un tema o producto ’

La tabla 3 proporciona información sobre la cobertura y nivel de integración que presenta la tipología propuesta respecto a los elementos componentes de las anteriores, mencionados por filas. Cada celda asocia cada uno de los tipos “previo” con el correspondiente de la nueva tipología. Puede observarse que el elemento que se detecta con más frecuencia es el tipo EG3.1 *crowdproduction* y se ha elaborado el nuevo componente EG3.3 *crowdanalyzing* para incorporar la tarea de búsqueda e interpretación de información en documentos multimedia. Finalmente, todos los elementos previos están reflejados por al menos un componente.

5. Comprobación de la tipología propuesta

Para testear la validez de la propuesta realizada, se han escogido 15 casos al azar, a partir de una lista de 84 iniciativas de *crowdsourcing* (Wikipedia, 2011).

Los ejemplos seleccionados son:

- *99designs*: plataforma web donde las empresas plantean sus necesidades de diseño gráfico para que sean resueltas por la multitud a cambio de una recompensa económica.
- *Article One Partners*: comunidad de expertos en tecnología que buscan información sobre el estado del arte de un tema relacionado con una patente nueva.
- *BlueServo*: los colaboradores pueden visualizar las cámaras de la frontera de EUA y México para detectar inmigrantes ilegales.
- *GoldCorp*: empresa minera que puso a disposición de la multitud su información técnica y material cartográfico, premiando a aquellos que encontraron nuevos yacimientos de oro.
- *IBM*: recogió más de 37.000 ideas a través de sesiones de *crowdstorming* donde participaron clientes, empleados y familiares de empleados.

- *The guardian*: investigó el escándalo de los miembros del *Parlamento* del Reino Unido y permitió el acceso a 700.000 documentos para que fueran examinados.
- *Juratis*: plataforma web estadounidense que permite que los usuarios puedan preguntar y resolver dudas legales.
- *Lánzanos*: plataforma española de *crowdfunding* que facilita que cualquiera presente un proyecto y sea financiado por la multitud.
- *Pepsi*: lanzó una campaña de publicidad en la que los usuarios podían diseñar una lata de refresco. El ganador recibía una recompensa económica.
- *reCaptcha*: utiliza el sistema *Captcha* para ayudar a digitalizar libros de texto, a la vez que protege los sitios web de accesos inadecuados (*anti-bot*).
- *setiQuest*: analizar señales recibidas del espacio para buscar signos de civilizaciones avanzadas.
- *SocialAttire*: votar diseños de ropa.
- *Starmind*: se plantean problemas de consultoría de empresas cuya resolución implicará una recompensa económica.
- *TopCoder*: se plantean retos sobre desarrollo de software y creaciones digitales.
- *Userfarm*: primera plataforma internacional de video elaborado mediante *crowdsourcing*.

Como se puede comprobar en la tabla 4, todos los casos seleccionados se ajustan a alguno de los tipos propuestos en el presente artículo.

‘ El *crowdsourcing* es un fenómeno reciente que ha surgido con fuerza y que puede ser utilizado en cualquier ámbito: empresarial, institucional, educativo, etc. ’

6. Conclusión

El *crowdsourcing* es un fenómeno reciente que ha surgido con fuerza y es susceptible de ser utilizado en cualquier ámbito: empresarial, institucional, educativo, etc. Las iniciativas de este tipo proliferan de forma notable y jugarán un papel cada vez más importante en la web del futuro. Con todo, adolece de un adecuado soporte de investigación relativo a su propia definición y a la descripción y clasificación de sus manifestaciones (**Denyer; Tranfield; Van-Aken, 2008**).

Como cualquier objeto que es sometido a un análisis para ser clasificado, el *crowdsourcing* presenta una serie de características que pueden ser empleadas como criterios de clasificación (**Doan; Ramakrishnan; Halevy, 2011**). En este artículo se ha escogido “la tarea a realizar” como criterio

	Elemento previo/propuesto						
Howe	H1.1/EG5	H1.2/EG1	H1.3/EG2.1	H2/EG3.1	H3/EG5	H4/EG4	
Kleeman	K1/EG2.1	K2/EG2.1	K3/EG1	K4/EG3.1	K5/EG3.1	K6/EG5	K7/EG2.2
Brabham	B1/EG3.2	B2/EG1	B3/EG3.1/EG5	B4/EG3.1			
Geerts	G1/EG1	G2/EG2.1	G3/EG3.1	G4/EG4			
Burger-Helmchen y Penin	BH1/EG1//EG2.1	BH2/EG3.1/EG3.2	BH3/EG3.1				

Tabla 3. Encaje de la nueva tipología con las tipologías estudiadas

	EG1	EG2.1	EG2.2	EG3.1	EG3.2	EG3.3	EG4	EG5
99designs	X							
Article One Partners					X			
BlueServo						X		
GoldCorp	X							
IBM		X						
The guardian					X			
Juratis			X					
Lánzanos							X	
Pepsi	X							
reCaptcha				X				
setiQuest						X		
SocialAttire								X
Starmind	X							
TopCoder	X							
Userfarm				X				

Tabla 4. Contraste de la tipología planteada con los casos seleccionados

fundamental ya que es el elemento que genera más diferencias: producirá lo que el iniciador de la actividad de *crowdsourcing* necesita y condiciona el resto de características.

Mediante una revisión sistemática de la bibliografía se han obtenido tipologías de *crowdsourcing* con el criterio de las tareas a realizar. Tras su análisis se ha propuesto una tipología que recoge e integra las anteriores siendo coherente con ellas. Se propone el tipo *crowdanalyzing* para atender a nuevas realidades colaborativas de generación de contenido dado que el tratamiento de los documentos en internet (numerosos y de calidad dispar) puede beneficiarse de la “inteligencia colectiva”. Además, esta tipología ha sido contrastada mediante quince casos de iniciativas de *crowdsourcing*.

Existen algunas limitaciones en el estudio: por un lado, la revisión sistemática de la literatura no ha cubierto obviamente todos los documentos que hablan de este tema, por lo que es posible que alguna tipología no haya sido tenida en cuenta; por otro lado, al ser un concepto dinámico (Schenk; Guittard, 2009), esta tipología tiene una validez temporal limitada por la aparición de nuevos modelos de negocio que hagan uso de la inteligencia colectiva.

Además, existen todavía aspectos del *crowdsourcing* en los que no existe un acuerdo explícito y que convendría que fueran tratados, como su relación con la co-creación o la innovación abierta. Incluso, hay ámbitos donde no ha sido ampliamente utilizado, como la educación superior, donde podría aportar importantes beneficios para todos los agentes implicados, dando lugar a nuevas posibilidades de estudio. Con todo, se ha aportado una chispa en el debate sobre este fenómeno.

7. Bibliografía

Brabham, Daren C. “Crowdsourcing as a model for problem solving: an introduction and cases”. *Convergence: the intl journal of research into new media technologies*, 2008a, Febr., v. 14, n. 1, pp. 75-90.

http://www.clickadvisor.com/downloads/Brabham_Crowdsourcing_Problem_Solving.pdf
<http://dx.doi.org/10.1177/135485650708442>

Brabham, Daren C. “Moving the crowd at iStockphoto: the composition of the crowd and motivations for participation in a crowdsourcing application”. *First monday*, 2008b, v. 13, n. 6.

<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2159/1969>

Burger-Helmchen, Thierry; Pénin, Julien. “The limits of crowdsourcing inventive activities: what do transaction cost theory and the evolutionary theories of the firm teach us?” En: *Workshop on open source innovation*, 2010, Estrasburgo, Francia.
http://cournot.u-strasbg.fr/users/osi/program/TBH_JP_crowdsourcing%202010%20ENG.pdf

Codina, Lluís. “Una propuesta de metodología para el diseño de bases de datos documentales (Parte II)”. *El profesional de la información*, 1997, dic., v. 6, n. 12, pp. 20-26.

http://www.elprofesionaldeinformacion.com/contenidos/1997/diciembre/una_propuesta_de_metodologia_para_el_diseo_de_bases_de_datos_documentales_parte_ii.html

Corney, Jonathan R.; Torres-Sánchez, Carmen; Jagadeesan, Prasanna; Lynn, A.; Regli, William. “Outsourcing labour to the cloud”. *Intl journal of innovation and sustainable development*, 2010, v. 4, n. 4, pp. 294-313.

<http://dx.doi.org/10.1504/IJSD.2009.033083>

Chesbrough, Henry W. *Open innovation: the new imperative for creating and profiting from technology*. Harvard Business Press, 2003, ISBN: 978 15 785 1837 7

Delgado-Rodríguez, Miguel; Sillero-Arenas, María; Gálvez-Vargas, Ramón. “Metaanálisis en epidemiología (Primera parte): características generales”. *Gaceta sanitaria*, 1991, v. 5, n. 27, pp. 265-272.

<http://www.elsevier.es/sites/default/files/elsevier/pdf/138/138v05n27a13140889pdf001.pdf>

Denyer, David; Tranfield, David; Van Aken, Joan-Ernst. “Developing design propositions through research synthesis”. *Organization studies*, 2008, marzo, v. 29, n. 3, pp. 393-413.

Doan, Anhai; Ramakrishnan, Raghu; Halevy, Alon. “Crowdsourcing systems on the world wide web”. *Communications of the ACM*, 2011, v. 54, n. 4, pp. 86-96.

<http://cacm.acm.org/magazines/2011/4/106563-crowdsourcing-systems-on-the-world-wide-web/fulltext>
<http://dx.doi.org/10.1145/1924421.1924442>

ECMT. European Commission for Mobility and Transport. *Door-to-door in a click*.

http://ec.europa.eu/transport/its/multimodal-planners/index_en.htm

Estellés-Arolas, Enrique; González-Ladrón-de-Guevara,

Fernando. "Towards an integrated crowdsourcing definition". *Journal of information science*, 2012, April, v. 38, n. 2, pp. 189-200.

<http://dx.doi.org/10.1177/0165551512437638>

FBI. Federal Bureau of Investigation. *Cryptanalysts: help break the code*.

<http://forms.fbi.gov/code>

Geerts, Simone. *Discovering crowdsourcing: theory, classification and directions for use*. Tesis de máster. Technische Universiteit Eindhoven, Netherlands, 2009.

<http://alexandria.tue.nl/extra2/afstversl/tm/Geerts%202009.pdf>

Geiger, David; Seedorf, Stefan; Schader, Martin. "Managing the crowd: towards a taxonomy of crowdsourcing processes". En: *Procs of the 7th Americas conf on information systems*, Detroit, Michigan, August 4th-7th, 2011.

http://schader.bwl.uni-mannheim.de/fileadmin/files/publikationen/Geiger_et_al._-2011_-_Managing_the_Crowd_Towards_a_Taxonomy_of_Crowdsourcing_Processes.pdf

Ghafele, Roya; Gibert, Benjamin; DiGiammarino, Paul. "How to improve patent quality by using crowdsourcing". *Innovation management*, Sept. 2011.

<http://www.innovationmanagement.se/2011/09/29/how-to-improve-patent-quality-by-using-crowd-sourcing/>

Grier, David-Alan. "Not for all markets". *Computer*, 2011, mayo, v. 44, n 5, pp. 6-8.

<http://www.computer.org/portal/web/The-Known-World/home/-/blogs/not-for-all-markets>

Heizer, Jay; Render, Barry. *Principles of operations management*. Prentice Hall, 2010. ISBN: 978 01 361 1446 8

Howe, Jeff. "Crowdsourcing: a definition". *Crowdsourcing: why the power of the crowd is driving the future of business*. June 2006.

http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html

Howe, Jeff. *Crowdsourcing: why the power of the crowd is driving the future of business*. Great Britain: Business Books, 2008, ISBN: 978 03 073 9620 7

Kleeman, Frank; Voss, G. Günter; Rieder, Kerstin. "Un(der)paid innovators: the commercial utilization of consumer work through crowdsourcing". *Science, technology & innovation studies*, 2008, v. 4, n. 1, p. 5-26.

<http://www.sti-studies.de/ojs/index.php/sti/article/viewFile/81/62>

La-Vecchia, Gioacchino; Cisternino, Antonio. "Collaborative workforce, business process crowdsourcing as an alternative of BPO". En: Floarian, Daniel; Facca, Federico. *Current trends in web engineering, ICWE 2010 workshops*. Viena, July 5-6, 2010, pp. 425-430. ISBN: 978 3 642 16984 7

Mazzola, Daniele; Distefano, Alessandra. "Crowdsourcing and the participation process for problem solving: the case of BP". En: *VII Conf of the Italian Chapter of AIS. Information technology and innovation trends in organizations*, 2010.

<http://www.cersi.it/itais2010/pdf/041.pdf>

Oliveira, Fabio; Ramos, Isabel; Santos, Leonel. "Definition

of a crowdsourcing innovation service for the European SMEs". En: Daniel F. et al. (eds.) *Current trends in web engineering. Lecture notes in computer science*, 2010, v. 6385, pp. 412-416.

http://dx.doi.org/10.1007/978-3-642-16985-4_37

Oomen, Johan; Aroyo, Lora. "Crowdsourcing in the cultural heritage domain: opportunities and challenges". En: *Procs of 5th intl conf on communities & technologies – C&T*. Queensland Univ. of Technology, Brisbane, Australia, 29 June-2 July 2011.

<http://www.cs.vu.nl/~marieke/OomenAroyoCT2011.pdf>

Pénin, Julien. "More open than open innovation? Rethinking the concept of openness in innovation studies". Working papers of BETA, Bureau d'Economie Théorique et Appliquée, 2008, UDS, Estrasburgo.

<http://www.beta-umr7522.fr/productions/publications/2008/2008-18.pdf>

Pinto-Molina, María; Alonso-Berrocal, José-Luis; Cordón-García, José-Antonio; Fernández-Marcial, Viviana; García-Figuerola, Carlos; García-Marco, Javier; Gómez-Camarero, Carmen; Francisco-Zazo, Ángel; Doucer, Anne-Vinciane. "Análisis cualitativo de la visibilidad de la investigación de las universidades españolas a través de sus páginas web". *Revista española de documentación científica*, 2007, v. 27, n. 3, pp. 345-370.

<http://redc.revistas.csic.es/index.php/redc/article/view/157/211>

<http://dx.doi.org/10.3989/redc.2004.v27.i3.157>

Pisano, Gary P.; Verganti, Roberto. "Which kind of collaboration is right for you". *Harvard business review*, 2008, v. 86, n. 12, pp. 78-86.

PlanB. *El Plan Ballantine's de Carlos Jean*, marzo 2011.

<http://prensa.elplanb.tv/>

Reichwald, Ralf; Piller, Frank T. *Interaktive wertschöpfung. Open innovation, individualisierung und neue formen der arbeitsteilung*. Wiesbaden: Gabler Verlag, 2006. ISBN: 978 38 349 0972 5

Schenk, Eric; Guittard, Claude. *Crowdsourcing: what can be outsourced to the crowd, and why?*, 2009. Technical report. http://halshs.archives-ouvertes.fr/docs/00/43/92/56/PDF/Crowdsourcing_eng.pdf

Schenk, Eric; Guittard, Claude. *Le crowdsourcing: modalités et raisons d'un recours à la foule*.

<http://marsouin.infini.fr/ocs2/index.php/frontieres-numeriques-brest2009/frontieres-numeriques-brest2009/paper/viewFile/60/8>

Schenk, Eric; Guittard, Claude. "Towards a characterization of crowdsourcing practices". *Journal of innovation economics*, 2011, v. 1, n. 7, pp. 93-107.

http://www.cairn.info/article.php?ID_ARTICLE=JIE_007_0093

Sloane, Paul. "The brave new world of open innovation". *Strategic direction*, 2011, v. 27, n. 5, pp. 3-4.

<http://dx.doi.org/10.1108/02580541111125725>

SuperBowl. *Crash the SuperBowl*, Nov. 2011.

<http://www.crashthesuperbowl.com>

Vukovic, Maja. "Crowdsourcing for enterprises". En: *Procs of the 2009 congress on services - I*. IEEE Computer Society, Washington DC, 2009, pp. 686-692.
<http://dx.doi.org/10.1109/services-i.2009.56>

Vukovic, Maja; Bartolini, Claudio. "Towards a research agenda for enterprise crowdsourcing". En: Margaria, Tiziana; Steffen, Bernhard (eds). *Leveraging applications of formal methods, verification, and validation. Lecture notes in computer science*, 2010a, v. 6415, pp. 425-434.
http://dx.doi.org/10.1007/978-3-642-16558-0_36

Vukovic, Maja; Bartolini, Claudio. "Crowd-driven processes: state of the art and research challenges". En: Maglio, Paul; Weske, Mathias; Yang, Jian; Fantinato, Marcelo. *Servi-*

ce-oriented computing. Lecture notes in computer science, 2010b, v. 6470, p. 733.
http://dx.doi.org/10.1007/978-3-642-17358-5_79

Vukovic, Maja; Kumara, Soundar; Greenshpan, Ohad. "Ubiquitous crowdsourcing". En: *Procs of the 12th ACM intl conf*, 2010, pp. 523-526.
<http://dx.doi.org/10.1145/1864431.1864504>

Wexler, Mark N. "Reconfiguring the sociology of the crowd: exploring crowdsourcing". *Intl journal of sociology and social policy*, 2011, v. 31, n. 1, pp. 6-20.
<http://dx.doi.org/10.1108/01443331111104779>

Wikipedia. *List of crowdsourcing projects*, Febr. 2011. http://en.wikipedia.org/wiki/List_of_crowdsourcing_projects



Más información:

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



LINKED DATA Y LINKED OPEN DATA: SU IMPLANTACIÓN EN UNA BIBLIOTECA DIGITAL. EL CASO DE EUROPEANA



Ana Ríos-Hilario, Diego Martín-Campo y Tránsito Ferreras-Fernández



Ana Ríos-Hilario, licenciada y doctora en documentación, es profesora del *Depto. de Biblioteconomía* de la *Univ. de Salamanca*, donde imparte, entre otras, la asignatura de descripción de documentos digitales en el *Máster en sistemas de información digital*. Su labor docente e investigadora se centra en la descripción y acceso documental de los recursos digitales, tema sobre el que ha realizado publicaciones y ponencias. Dirige trabajos de investigación sobre modelos de metadatos y proyectos de datos enlazados. Participa en el *Grupo estándares: sistemas de representación y organización del conocimiento* del proyecto *I+D+i Infoscopos*.

*Departamento de Biblioteconomía y Documentación
Universidad de Salamanca
Francisco de Vitoria 6-16. 37008 Salamanca, España
anarihi@usal.es*

Diego Martín-Campo, diplomado en biblioteconomía y licenciado en documentación por la *Universidad de Salamanca*. Cuenta con estudios de posgrado en tratamiento de información digital (*Máster en sistemas de información digital*), y realiza el doctorado en información y documentación. Trabaja como documentalista en la descripción y difusión de documentos digitales en el *Centro Internacional de Tecnologías Avanzadas* de la *Fundación Germán Sánchez Ruipérez*, y como técnico para el *Repositorio Documental Gredos* de la *Universidad de Salamanca*.

*Departamento de Biblioteconomía y Documentación
Universidad de Salamanca
Francisco de Vitoria, 6-16. 37008 Salamanca, España
diegomartin@usal.es*

Tránsito Ferreras-Fernández es licenciada en filología francesa y máster en sistemas de información digital por la *Universidad de Salamanca (Usal)*. Pertenece al cuerpo de ayudantes de archivos, bibliotecas y museos de la *Usal*. Es coordinadora técnica del *Repositorio Institucional Gredos*. Investiga el impacto del acceso abierto, el desarrollo e implementación de esquemas de metadatos en repositorios *DSpace* y a la preservación digital de repositorios. Participa en grupos de trabajo de proyectos nacionales como *Recolecta*, e internacionales como *Europeana*.

*Universidad de Salamanca
Repositorio Institucional Gredos
Libreros, 68. 37008 Salamanca, España
transiff@usal.es*

Resumen

Se realiza un análisis de los conceptos *linked data* (LD) y *linked open data* (LOD) y su aplicación a un caso real: *Europeana*. Para LD se parte del informe final publicado recientemente por el *Library Linked Data Incubator Group (LLD-XG)*, describiendo seguidamente el caso de *Europeana* y detallando las tareas realizadas y por desarrollar. Posteriormente, se examina la aplicación de LOD a través de la implantación del *Europeana data model* (EDM) haciendo referencia a cada uno de sus componentes. Se concluye con una valoración de la aplicación de estas técnicas en el ámbito de las instituciones culturales.

Artículo recibido el 09-02-12
Aceptación definitiva: 16-05-12

Palabras clave

Linked data (LD), Linked open data (LOD), Library linked data incubator Group (LLD-XG), *Europeana data model* (EDM).

Title: *Linked data and linked open data and its implementation in a digital library: the case of Europeana*

Abstract

An analysis of the implementation of linked data (LD) and linked open data (LOD) techniques in *Europeana* is presented. The final report recently published by the *Library Linked Data Incubator Group* (XG-LLD) is examined, and its application in *Europeana* is analyzed. Details of the work already done and the activities to be developed are provided. Then, the article examines the LOD implementation following the *Europeana data model* (EDM) for each of its components. Finally an assessment on the use of these techniques in the field of cultural institutions is done.

Keywords

Linked data (LD), Linked open data (LOD), *Library Linked Data Incubator Group* (LLD-XG), *Europeana data model* (EDM)

Ríos-Hilario, Ana; Martín-Campo, Diego; Ferreras-Fernández, Tránsito. "Linked data y linked open data: su implantación en una biblioteca digital. El caso de *Europeana*". *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 292-297.

<http://dx.doi.org/10.3145/epi.2012.may.10>

1. Introducción: sobre datos abiertos, enlazados y datos abiertos enlazados

Los movimientos *open* (*open access*, *open source*, *open data*) están siendo objeto de atención por la comunidad bibliotecaria en los últimos años, pues aportan numerosos beneficios ante el gran crecimiento de información, ya que aumentan las posibilidades de descubrir, usar y reutilizar los recursos en las instituciones culturales.

Open access se define como el libre acceso a la información y conocimiento a través de internet, sin barreras económicas y sin restricciones derivadas de los derechos de copyright (Ferreras-Fernández, 2011). Las instituciones culturales se benefician del software *open source*, reduciendo los costes iniciales y de mantenimiento al eliminar —en principio— la figura de los proveedores tecnológicos. El software OS facilita a los profesionales un mayor control sobre sus sistemas pudiendo obtener un mejor conocimiento sobre los intereses y expectativas de los usuarios, pues pueden adaptar los programas libremente a sus necesidades. Finalmente, *open data* es un movimiento que promueve la liberación de datos, generalmente no textuales y en formatos reutilizables no propietarios como CSV (*comma separated values*), procedentes de organizaciones (Peset; Ferrer-Sapena; Subirats-Coll, 2011). En este contexto, las bibliotecas juegan un importante papel dentro de la web semántica, al exponer grandes colecciones de datos que los usuarios pueden reutilizar y enriquecer.

Igualmente podíamos hablar de los movimientos "data", entre los que se encuentran *open data*, *linked data* (LD) y la suma de ambos, *linked open data* (LOD). LD (*datos vinculados o enlazados*) se refiere a los datos publicados de acuerdo con los principios diseñados para facilitar los vínculos entre conjuntos de datos, vocabularios controlados y otros elementos (Berners-Lee, 2011). LD utiliza URIs (*uniform resource identifiers*) como identificadores únicos globales para todo tipo de recursos. Los datos enlazados se expresan utilizando estándares como RDF (*resource description fra-*

mework), que especifica las relaciones entre los recursos, las relaciones que se pueden utilizar para navegar entre ellos, o integrar la información de múltiples fuentes.

Por su parte, el objetivo del LOD es la construcción de una base de conocimiento común mediante la publicación en la Web de bases de datos en RDF y el establecimiento de enlaces RDF entre datos de diferentes fuentes.

En este sentido, nos parecen muy oportunas las palabras de Zeinstra y Keller (2011) cuando manifiestan que "los datos pueden estar abiertos pero no enlazados. Los datos pueden estar enlazados pero no abiertos. La web semántica puede solamente funcionar con datos que estén tanto abiertos como enlazados".

Europeana es un portal de internet que actúa como una interfaz de millones de libros, pinturas, películas, objetos de museo y registros de archivos digitalizados de toda Europa (figura 1). Esta biblioteca es también una plataforma para el intercambio de conocimiento que promueve la colaboración entre los agentes culturales.

Los datos pueden estar abiertos pero no enlazados. Los datos pueden estar enlazados pero no abiertos. La web semántica puede solamente funcionar con datos que estén tanto abiertos como enlazados (Zeinstra y Keller)

El objetivo del presente artículo es estudiar los conceptos de LD y LOD a través de un caso real de biblioteca digital: *Europeana*. A su vez, se pretende analizar el informe aprobado en octubre de 2011 por el *Library Linked Data Incubator Group* (LLD-XG, Grupo Incubadora de Datos Enlazados Bibliotecarios) y el modelo de aplicación LOD a través del *Europeana data model* (EDM). Todo ello realizado desde un

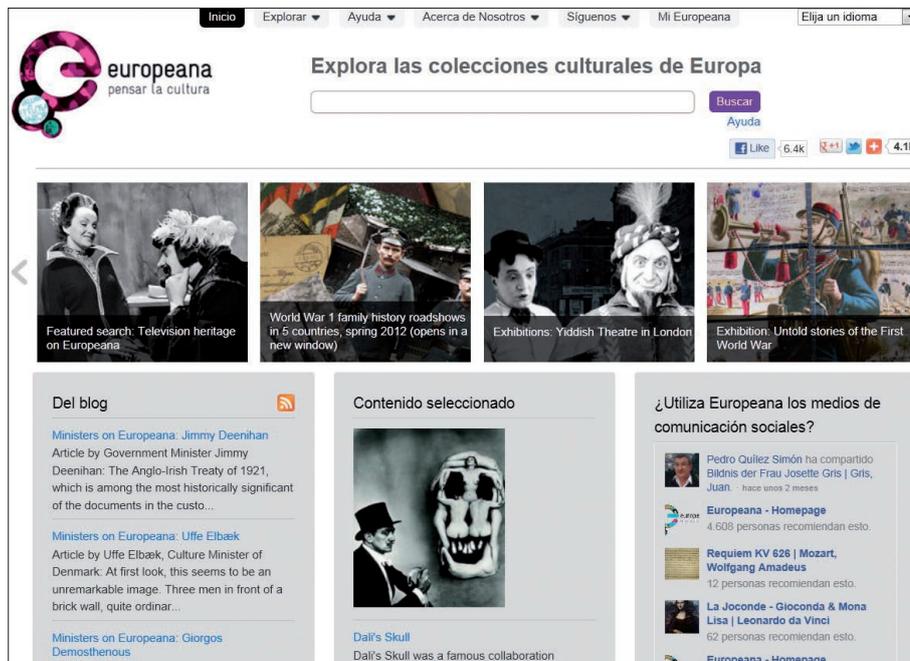


Figura 1. Página de inicio de *Europeana*, <http://www.europeana.eu/portal>

punto de vista eminentemente práctico, centrándonos en la aplicación real de estos conceptos.

2. *Library Linked Data Incubator Group (LLD-XG): análisis de uso en Europeana*

2.1. *LLD-XG: Informe final*

En mayo de 2010 el *World Wide Web Consortium (W3C)* constituyó el *Library Linked Data Incubator Group*. Su labor consistió en estudiar los componentes básicos ya existentes en este ámbito, tales como los modelos de metadatos y los estándares y protocolos de interoperabilidad, animando a las bibliotecas a publicar sus contenidos y en general a reorientar sus métodos de cara a la interoperabilidad de los datos en la Web. La idea se extendió también a otras comunidades, pues el término *library data* se utiliza en el informe en el sentido más amplio, comprendiendo no sólo los datos bibliotecarios sino también los de archivos y de museos –LAM (*libraries, archives and museums*)–.

El resultado del trabajo del *LLD-XG* se presentó en octubre de 2011 con el título *Library Linked Data Incubator Group final report*¹, que incluye a su vez otros dos documentos que lo complementan pero que se publican de modo independiente: *Use cases* (“*Usecase*”, casos de usos) y *Datasets, value vocabularies, and metadata element sets* (“*VocabDataset*”, Conjuntos de datos, vocabularios controlados y conjuntos de elementos de metadatos). Según Saorín (2011) este documento “es relevante para entender la oportunidad que se abre a las bibliotecas de hacer mejor las cosas, alcanzar más impacto digital y apostar por formas complementarias de utilidad social de una forma indirecta”.

El cuerpo del documento se divide en los siguientes apartados: ámbito del informe; beneficios de la técnica de datos vinculados; situación actual; y recomendaciones. Resulta interesante la lectura del apartado recomendaciones en función de los distintos destinatarios establecidos.

Una de las principales actividades del *Grupo* fue la recopilación de casos de uso que demuestran la implementación exitosa de las tecnologías de web semántica en las diferentes instituciones culturales. El conjunto de casos prácticos recopilados en el *Usecase* se organizaron en las siguientes agrupaciones:

- datos bibliográficos
- datos de autoridad
- vocabularios
- archivos y datos heterogéneos
- citas
- objetos digitales
- colecciones
- usos sociales

En el siguiente apartado estudiaremos esta publicación a través del análisis detallado de *Europeana*.

VocabDataset es un complemento del informe final que pretende identificar “un conjunto de recursos útiles para la creación o consumo de datos enlazados en el campo de las bibliotecas”. Se concibe como un punto de partida para que los profesionales puedan encontrar, comprender y explorar algunos ejemplos de conjuntos de datos, vocabularios controlados y conjuntos de metadatos.

“ Los estándares de la web semántica y los principios de *linked data* pueden utilizarse para hacer que los valiosos activos de información que la biblioteca crea y mantiene sean más visibles y reutilizables en la Web ”

2.2. Caso de uso: *Europeana*

El *LLD-XG* incluye al portal *Europeana* dentro del grupo “archivos y datos heterogéneos”. La siguiente figura recoge la estructura del estudio (figura 2). Los apartados de los que consta el análisis son idénticos en cada uno de los casos recopilados por el *LLD-XG*, con lo cual, a través de este caso específico podremos conocer también el modo de proceder del *Grupo* en la indagación de los casos prácticos. Seguidamente pasaremos a detallar los datos más relevantes obtenidos en el examen de *Europeana*.

Actualmente, la labor de *Europeana* se ve dificultada por la gran variedad de formatos de metadatos utilizados. Es posible exigir a los proveedores presentar los metadatos utilizando un modelo unificado, lo que generalmente se traduce en la pérdida de información útil. También es difícil para la recolección, mantener y explotar los enlaces que conectan los objetos a la información contextual, tanto dentro de las colecciones de objetos agregados por el proyecto como entre estos objetos y otros recursos externos (por ejemplo,

Figura 2. Metadata aggregation for accessing digitized objects, Use Case http://www.w3.org/2005/Incubator/llid/wiki/Use_Case_Europeana

La recopilación de los casos de uso por el LLD-XG demuestra la implementación exitosa de las tecnologías de web semántica en las bibliotecas y otros sectores afines

Europeana no parte de cero, puesto que para beneficiarse de la técnica de LD ya ha realizado las siguientes tareas: experimentación con la búsqueda semántica RDF para las colecciones, alineamiento de los sistemas de organización del conocimiento y utilización de agregadores de datos que exploran los datos enlazados. Además, este portal emplea modelos de referencia y vocabularios controlados que pueden ser útiles y aplicarse dentro del contexto de los datos enlazados.

3. Aplicación de LOD a través de *Europeana data model* (EDM)

Una de las características fundamentales de *Europeana* es que dispone de un modelo de datos específico, *Europeana data model* (EDM) basado en LOD. Con este modelo y las especificaciones funcionales se inicia el 1 de mayo de 2011 la denominada *fase Danubio* (figura 3) en la que actualmente se encuentra el proyecto (**Agénjo-Bullón**, 2012). Según estas especificaciones, el objetivo de esta fase es “proveer de una sofisticada funcionalidad a la biblioteca digital *Europeana* con más de 10 millones de ítems procedentes del patrimonio cultural y científico de Europa”.

El deseo de *Europeana 2011* que era llegar a ser una institución distribuidora, facilitadora e innovadora (**Bloomberg**, 2010) no hay duda de que ya se está cumpliendo.

Wikipedia). La reutilización de los objetos y sus datos adjuntos del mismo modo se ve obstaculizada por la rigidez de los medios para acceder a ellos (portal y API –*application programming interface*–).

La utilización de LD y otras técnicas de web semántica pueden aplicarse en *Europeana* para recolectar objetos mejor contextualizados, enriquecer los metadatos ya existentes, y para proporcionar un acceso directo a los mismos lo que permite su reutilización por terceros. Igualmente, su uso posibilita la coexistencia de varios niveles de generalidad en una misma base de conocimiento y una mejora en los procesos de búsqueda.

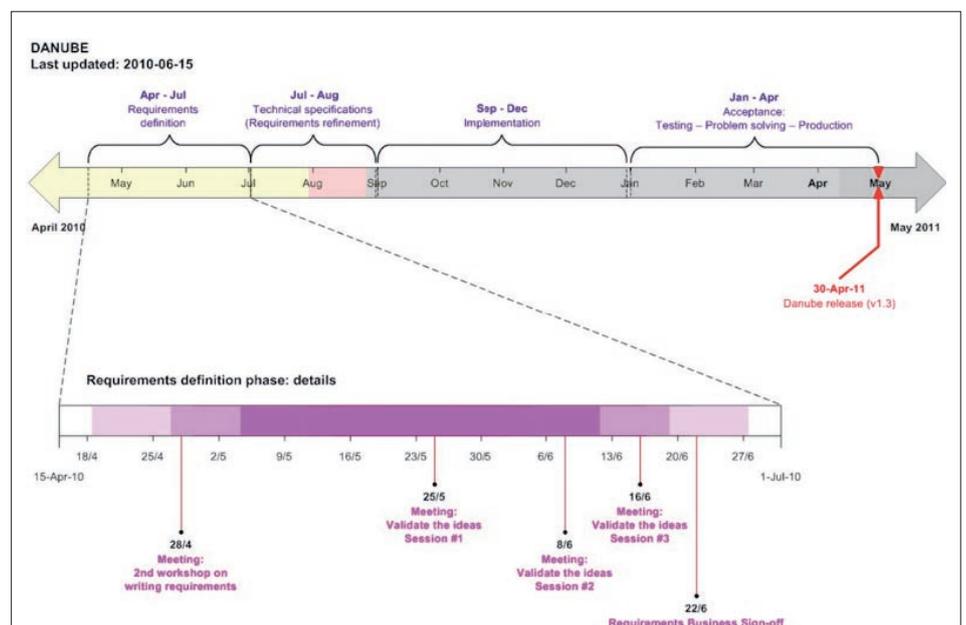


Figura 3. Cronograma de la fase Danubio http://version1.europeana.eu/c/document_library/get_file?uuid=42fdf604-8bd7-4fe4-8d3e-36cce6fcad65&groupld=10602



Figura 4. *Europeana data model*

La figura 4 sintetiza el cambio del modelo *Europeana semantic elements* (ESE) a EDM.

Componentes de *Europeana data model*

Europeana pretende ofrecer una contextualización semántica rica en las representaciones de objetos, para realizar sobre ellos operaciones semánticas complejas que no podrían ser soportadas en la interfaz tradicional de una biblioteca digital. Para ello las representaciones de objetos de esta biblioteca necesitan estar conectadas sistemáticamente a LOD, bien en internet o en recursos de contextualización semántica propios de *Europeana*, como tesauros y vocabularios estructurados migrados al estándar SKOS (*Simple knowledge organization system*). Por tanto, la estrategia técnica de la institución fue concebida para contribuir al paradigma LD y, al mismo tiempo, beneficiarse de él.

La adopción de EDM permite a *Europeana* ser compatible con la web semántica utilizando los estándares del W3C. EDM está constituido por la reutilización de namespaces como RDF, RDFS (*resource description framework schema*), OAI-ORE (*Open archives initiative object reuse and exchange*), SKOS y DCterms (*Dublin core metadata initiative terms*).

LD utiliza la arquitectura de la Web como estructura principal y RDF como lenguaje de representación. Dicho estándar permite representar información estructurada acerca de cualquier recurso mediante simples declaraciones triples (sujeto, predicado, objeto), es decir, los recursos se describen por medio de conexiones entre ellos. Por otra parte, EDM permite convertir cualquier objeto de interés en un recurso, identificándolo mediante una URI.

El enfoque de web semántica permite la combinación de varias ontologías en las mismas descripciones, entendidas esas ontologías como conceptualizaciones (Gruber, 1993). EDM reutiliza algunas de las ontologías de referencia ya existentes, como el estándar SKOS. En esta línea, ya se ha publicado la *Lista de encabezamientos de materia para bibliotecas públicas* que

establece un *closeMatch* con los *Library of Congress subject headings (LCSH)* y recientemente con *Rameau (Répertoire d'autorité-matière encyclopédique et alphabétique uni-fié)* y los *Schalgwortnormdatei (Agenjo-Bullón, 2012)*.

En cuanto al esquema de descripción *Dublin Core (DC)* proporciona un modelo compacto para describir las características esencia-

les de objetos culturales o científicos (creadores, relaciones con otros recursos, indización por materias, etc.) ajustándose a un amplio rango de necesidades de acuerdo con la web semántica. DC se utilizó también como base del esquema ESE, lo que permite a los proveedores de datos seguir utilizándolo en su versión 3.4., para enviar sus datos a *Europeana*. DC se emplea en muchas aplicaciones permitiendo así que los datos en EDM sean más susceptibles de compartir y reutilizar.

Para el intercambio y reutilización de objetos, se han escogido las especificaciones de OAI-ORE (figura 5) como marco estructural para el modelo de la ontología EDM, ya que, las representaciones de objetos típicas de *Europeana* son entidades compuestas de varias partes, como por ejemplo metadatos de atributos, una vista en miniatura y una página estática en html. OAI-ORE permite un enfoque para la identificación y descripción de estos conjuntos de recursos web.

4. Conclusiones

La aplicación de las técnicas de LD y LOD en las instituciones culturales es ya un hecho, tal y como recogen los casos de uso recopilados por el LLD-XG. Además, el informe de dicho grupo contribuye a sentar las bases en la materia a través de sus recomendaciones.

La implementación de LD en las bibliotecas como *Europeana*, va a permitir un enriquecimiento en todos sus niveles tanto de los datos, modelos de descripción y procesos de búsqueda. Así, y a tenor del trabajo ya realizado, podemos concluir que esta gran biblioteca digital se está convirtiendo

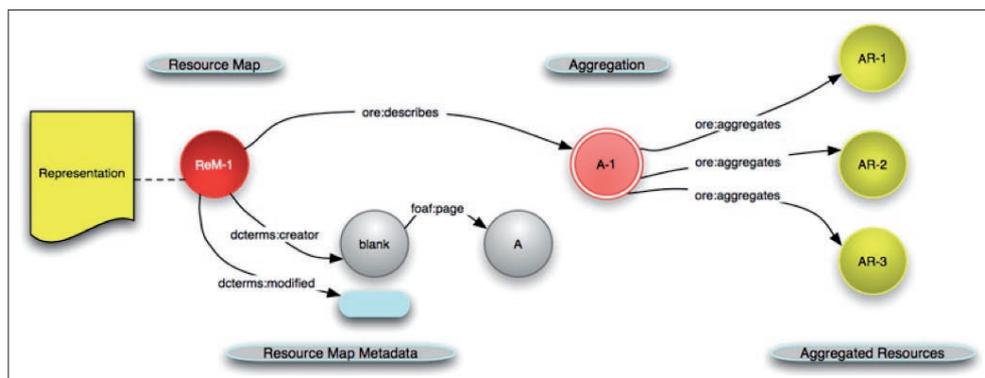


Figura 5. Ejemplo de agregación con OAI-ORE, <http://www.openarchives.org/ore/1.0/primer>

en abanderada de la implementación tecnológica de LD y LOD.

Del estudio realizado extraemos las siguientes apreciaciones sobre la aplicación de dichas técnicas en el entorno de las instituciones culturales. Existen esquemas y modelos de metadatos que evolucionan hacia la Web, normas y protocolos para la construcción de la interoperabilidad más allá del ámbito bibliotecario que cambian de una visión integrada a un entorno de red. Una gran cantidad de datos estructurados ya están disponibles en los sistemas bibliotecarios y podrían publicarse como LD utilizando las técnicas de web semántica. Las instituciones del patrimonio cultural podrían ser un importante proveedor de bases de datos de autoridad (personas, materias...) para la web de datos enlazados.

Sin embargo, también hay que salvar una serie de obstáculos: la adopción real de los principios mismos de datos enlazados en estas comunidades tendrán que coordinarse. Por otra parte, LD a menudo se está construyendo a un nivel de investigación y desarrollo, y trasladar sus resultados a la práctica parece difícil (*Library Linked Data Incubator Group final report*, 2011).

Recientemente **Tomás Saorín** (2011) planteó algunas cuestiones sobre LOD y las oportunidades que representa este nuevo paradigma para las bibliotecas. En este sentido afirmaba “es momento de acordar cuál ha de ser la posición en la que las bibliotecas pueden ser más útiles y eficientes en el mundo *linked open data*”. Nosotros consideramos que es una oportunidad para nuestro sector. Es hora de ponerse manos a la obra.

5. Nota

1. El documento *VocabDataset* fue traducido al castellano primero por Saorín y Pastor:
<http://skos.um.es/Incubator/llid/XGR-llid-vocabdataset>

En marzo de 2012 se publicó la versión española de todos los documentos del informe realizada por Agenjo y Hernández.
http://www.larramendi.es/i18n/cms/elemento.cmd?grupo=estaticos/paginas&elemento=documentacion_tecnica_lam.html

6. Referencias

Agenjo-Bullón, Xavier. “*Europeana* en fase Danubio”. *Blok de BiD*, 25 enero 2012.
<http://www.ub.edu/blokdebid/es/node/157>

Berners-Lee, Tim; Fielding, Roy; Masinter, Larry. *RFC 3986 - Uniform resource identifier (URI): generic syntax*. The Internet Society, 2005.

<http://tools.ietf.org/html/rfc3986>

Berners-Lee, Tim. *Linked data*. World Wide Web Consortium, 18 junio 2009.
<http://www.w3.org/DesignIssues/LinkedData.html>

Bloomberg, Ruth; Dekkers, Makx; Gradmann, Stefan; Lindquist, Mats; Meghini, Carlo; Verleyen, Julie. *Functional specification for the Europeana Danube release*. Europeana Think Culture, 2010.
http://pro.europeana.eu/documents/844813/851994/D3_2_Final.pdf

Europeana Semantic Elements, v.3.4, 2011-03-31.
<http://www.europeana.eu/schemas/ese>

Europeana
<http://www.europeana.eu>

Ferreras-Fernández, Tránsito. “Open access en España: los repositorios institucionales, 2011”. En: *V Jornadas de e-learning en la formación para el empleo en las administraciones públicas*, Valladolid, 15-17 sept. 2011.
<http://eprints.rclis.org/handle/10760/16355#.T7evSZ91Dsw>

Gruber, Thomas R. “Toward principles for the design of ontologies used for knowledge sharing”. *Intl journal human-computer studies*, 1993, n. 43, pp. 907-928.
<http://tomgruber.org/writing/onto-design.pdf>

Library Linked Data Incubator Group final report.
<http://www.w3.org/2005/Incubator/llid/XGR-llid-20111025>

Marcos-Martín, Carlos; Soriano-Maldonado, Salvador-Luis. “Reutilización de la información del sector público y open data en el contexto español y europeo. *Proyecto Aporta*”. *El profesional de la información*, 2011, mayo-junio, v. 20, n. 3, pp. 291-297.
<http://dx.doi.org/10.3145/epi.2011.may.07>

Peset, Fernanda; Ferrer-Sapena, Antonia; Subirats-Coll, Imma. “Open data y linked open data: su impacto en el área de bibliotecas y documentación”. *El profesional de la información*, 2011, marzo-abril, v. 20, n. 2, pp. 165-173.
<http://www.elprofesionaldelainformacion.com/contenidos/2011/marzo/06.pdf>
<http://dx.doi.org/10.3145/epi.2011.mar.06>

Saorín, Tomás. “Cómo *linked open data* impactará en las bibliotecas a través de la innovación abierta”. *Anuario ThinkEPI*, 2012, v. 6, pp.
<http://www.thinkepi.net/como-linked-open-data-impacta-ra-en-las-bibliotecas-a-traves-de-la-innovacion-abierta>

Zeinstra, Maarten; Keller, Paul. *Open linked data and Europeana*.
http://pro.europeana.eu/c/document_library/get_file?uuid=374c381f-a48b-4cf0-bbde-172cf03672a2&groupId=10602



HERRAMIENTAS WEB PARA LA MEDICIÓN DE LA INFLUENCIA DIGITAL: ANÁLISIS DE *KLOUT* Y *PEERINDEX*



Javier Serrano-Puche



Javier Serrano-Puche, profesor de teoría de la comunicación en la *Universidad de Navarra*, investiga los fundamentos teóricos de la comunicación, y sobre todo del periodismo, con interés por el ámbito digital 2.0. En especial estudia las redes sociales, tanto por sus implicaciones para la expresión de la identidad personal como en la conformación de la esfera pública. Es autor de *La verdad recobrada en la escritura* (2011), una biografía intelectual de Leonardo Sciascia. Sobre este escritor siciliano también ha publicado numerosos ensayos en diversas revistas españolas y extranjeras.

Universidad de Navarra
Departamento de Comunicación Pública
Edif. Biblioteca, 31080 Pamplona, España
jserrano@unav.es

Resumen

Cada vez más es reconocida la relevancia en el entorno digital de los “líderes de opinión” o usuarios influyentes: aquellos que por medio de su actividad online (publicación de *tweets* y entradas en blogs, actualización de su estado en las redes sociales, recomendación de lecturas...), cumplen con la función de crear contenidos o de filtrarlos hacia personas sobre las que tienen ascendencia. Por eso han proliferado las herramientas que evalúan la influencia de una persona o marca a través de la monitorización de su uso de los medios sociales. Se analizan las dos más importantes: *Klout* y *PeerIndex*. Se explican los parámetros que constituyen la base de sus algoritmos de medición, así como sus carencias. La comprensión de cómo se ejerce y cómo puede medirse la influencia digital es una de las cuestiones más interesantes del fenómeno 2.0.

Palabras clave

Influencia digital, Medios sociales, Redes sociales, Web 2.0, Analítica web, *Klout*, *PeerIndex*, *Twitter*, *Facebook*.

Title: Digital influence measurement tools: an analysis of *Klout* and *PeerIndex*

Abstract

The digital realm increasingly recognizes the relevance of “opinion leaders” or influential users: those who, through their online activity (publishing tweets and blog posts, updating social networks status, recommending readings, etc.) create or filter content to people over whom they exert personal influence. Hence the proliferation of tools that assess the influence of a person or brand by monitoring use of social media. The article discusses the two most important tools: *Klout* and *PeerIndex*, explaining the parameters that form the basis of their measurement algorithms as well as the deficiencies of these models of analysis. Understanding how digital influence is exercised and how to measure it is one of the most interesting questions of the “2.0” phenomenon.

Keywords

Digital influence, Social media, Social network sites, Web 2.0, Web analytics, *Klout*, *PeerIndex*, *Twitter*, *Facebook*.

Serrano-Puche, Javier. “Herramientas web para la medición de la influencia digital: análisis de *Klout* y *PeerIndex*”. *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 298-303.

<http://dx.doi.org/10.3145/epi.2012.may.11>

1. Introducción

Internet ha trastocado muchos de los elementos que hasta hace poco regían el proceso de comunicación pública. Frente a la unidireccionalidad en la difusión, la linealidad del discurso, la periodicidad de los mensajes o cierta pasividad por parte de la audiencia receptora —características

propias de la comunicación de masas—, con la generalización de la Red han entrado en juego nuevos paradigmas. En este ecosistema digital rigen factores diferentes, tales como la centralidad del usuario en el proceso comunicativo, la universalización del lenguaje multimedia y del hipertexto, las interacciones en tiempo real, la necesaria gestión de la abundancia informativa, la desintermediación de los proce-

Artículo recibido el 05-12-11
Aceptación definitiva: 20-02-12

sos comunicativos y la revalorización del conocimiento por encima de la información (Orihuela, 2004).

Los cambios derivados del auge de internet permiten mayor integración social de los copartícipes de la comunicación. Así se constata con la popularización y consolidación de la web 2.0, caracterizada por una mayor interactividad, participación y colaboración de sus usuarios. La muestra más representativa del fenómeno “dos punto cero” son las redes sociales, que permiten tanto consolidar relaciones ya establecidas en el mundo físico como construir nuevos vínculos (Boyd; Ellison, 2007). Constituyen el cauce mediante el cual el usuario agrega y comparte su actividad en internet, construyendo y expresando así su identidad online y, en definitiva, convirtiéndose en “una nueva suerte de ágora en la que reinventamos nuestra actividad social más natural: la relacional” (Fumero-Reverón, 2011). Su impacto en la configuración del escenario de la comunicación digital es indudable, hasta el punto de que “la *capa social* invade los espacios web: todo se vuelve social y definir qué es red social se vuelve aún más complejo (y quizá innecesario)” (*The cocktail analysis*, 2011). Se llega a hablar de una “sociedad red” en la era de la información (Castells, 2001).

Es precisamente en las redes sociales donde mejor puede apreciarse la vigencia de los nuevos paradigmas comunicativos. Aunque la web 2.0 propicia una mayor democratización en el acceso y producción de contenidos digitales, el ciudadano se enfrenta a un enorme flujo informativo que puede conducirle a la saturación (Gitlin, 2005) y al que resulta imprescindible dotar de orden y sentido, como tradicionalmente han venido haciendo los medios periodísticos. En la actualidad tenemos acceso a multitud de datos, pero con frecuencia nos falta el contexto y la capacidad de interpretación. Es una de las paradojas de la comunicación digital, puesto que “el incremento de la cantidad de información no lleva consigo mayor calidad en los conocimientos” (Pérez-Latre, 2011).

Una de las paradojas de la comunicación digital es que tenemos acceso a multitud de datos, pero con frecuencia nos falta el contexto y la capacidad de interpretación

Pese a que internet ha propiciado la desintermediación de los procesos comunicativos, en el ámbito digital emerge con fuerza la figura de líderes de opinión o *influentials*, tanto en el área de la actualidad informativa como en el complejo entramado de decisiones sobre entretenimiento, consumo, tecnología, tendencias sociales, etc. Por medio de su actividad *online* (publicación de *tweets* y entradas en blogs, actualización de su estado en redes sociales, recomendación de lecturas y vídeos, etc.) y de manera más o menos consciente, los usuarios influyentes cumplen con la función de generar contenidos, o de filtrarlos hacia otros usuarios sobre los que tienen cierta ascendencia.

En definitiva, y por expresarlo en términos espaciales, la horizontalidad en el acceso a la información derivada de la irrupción de la Red no anula una indispensable verticalidad

jerárquica, más necesaria si cabe frente a la ingente cantidad de información disponible.

2. Influencia personal en los procesos comunicativos

La constatación de que también en el ámbito digital existen figuras con capacidad de liderazgo en determinados asuntos, hasta el punto de modificar el comportamiento o la actitud de las personas bajo su área de influencia, nos remite a una cuestión clásica en la investigación sobre los efectos de los medios de comunicación. En estudios pioneros como *The people's choice* (Lazarsfeld; Berelson; Gaudet, 1944) y *Personal influence* (Katz; Lazarsfeld, 1955), sus autores detectaron que en todos los grupos sociales existen individuos particularmente activos, más sensibles que el resto a los intereses del propio grupo y más proclives a expresar abiertamente sus opiniones. Demostraron, por un lado, la existencia de un “flujo de la comunicación en dos pasos”, por medio del cual la información que parte desde los medios masivos es recibida y a su vez retransmitida por esos líderes de opinión a las demás personas de los grupos sociales. Por otro, constataban la primacía de la influencia personal por encima de la influencia de la comunicación de masas en el ámbito de las decisiones políticas y, sobre todo, en el ámbito de compras domésticas y ocio.

Es necesario estudiar los procesos de influencia personal y de liderazgo en el medio digital

Desde estas investigaciones, han sido más de 4.000 los estudios centrados en la figura de los *influentials* (Berry; Keller, 2003; Weimann, 2004). Se han ampliado tanto los temas tomados en consideración (planificación familiar, ciencia, cuidados de la salud, etc.) como las corrientes teóricas que se aproximan al fenómeno: difusión de las innovaciones (Beal; Bohlen, 1957; Rogers, 1962); teoría social cognitiva (Bandura, 1982); difusión de contenidos e ideas (Gladwell, 2002); teoría de *agenda setting* (Brossius; Weimann, 2006)... La generalización del ámbito digital como espacio de comunicación reclama la necesidad de estudiar los procesos de influencia personal y de liderazgo en este nuevo entorno mediático.

3. Medición de la influencia en el ámbito digital

Conocer la dinámica de los procesos de comunicación es una de las cuestiones de mayor interés, y para ello han surgido aplicaciones que intentan medir la influencia en las redes sociales de una determinada identidad digital (persona o marca). Son numerosas las que se limitan a una labor de métrica aplicada a la actividad del usuario en *Twitter* (*Twitalyzer*, *Twitter Grader*, *TweetLevel*, *Retweetrank*). Más interesante, en cambio, son las que trabajan con algoritmos que cuantifican más de una red social, ofrecen una categorización del tipo de liderazgo del usuario, determinan cuál es su área de influencia y, en conjunto, muestran plena adecuación con el estilo comunicativo 2.0, incluida una dosis de

“ludificación” (*gamification*) que mejora su usabilidad. Es el caso de *Klout* y *PeerIndex*.

<http://klout.com>

<http://www.peerindex.com>

3.1. Klout, ¿ranking estándar de influencia?

Fundada en agosto de 2008, *Klout* es actualmente la herramienta de medición de mayor aceptación social (alcanzó los 100 millones de usuarios conectados en septiembre de 2011). Se ha erigido así para muchos en el ranking “estándar sobre la influencia”, como reza su lema.

Klout mide de 0 a 100 la influencia personal del usuario en función de su actividad hasta en 12 redes sociales

Mide en una escala de 0 a 100 la influencia personal del usuario, en función de su presencia y actividad en las redes sociales. En la actualidad puede monitorizar doce de ellas (*Twitter, Facebook, Google+, LinkedIn, Foursquare, YouTube, Instagram, Tumblr, Blogger, WordPress, Last.fm, Flickr*); corresponde al usuario decidir cuáles conecta a su perfil. Basándose en más de 30 algoritmos que la empresa no ha dado a conocer, la puntuación se actualiza diariamente. En palabras de su creador, **Joe Fernández**, “la premisa central que sustenta nuestros algoritmos siempre ha sido que la influencia es la capacidad para dirigir la acción. No eres más influyente porque tuitees más o uses más *Facebook*, lo eres porque tienes una audiencia influyente que engancha e interactúa con tu contenido” (**Fernández**, 2011a). De ahí que el resultado final del índice *Klout* se obtenga mediante la agregación de tres factores:

1) Alcance real: número de personas sobre las que se ejerce influencia (que no coincide con el número de *followers*). Responde a la cuestión “¿a cuánta gente influyes?”. Para ello analizan cómo las personas con las que un usuario está conectado interactúan con él y con su contenido.

2) Amplificación: “¿Cuánto influyes en la gente sobre la que tienes influencia?”. Se mide tomando en cuenta cómo los mensajes provocan conversación o son replicados (retuiteados, marcados como “me gusta” en *Facebook*, merecedores de un +1 en *Google+*, etc.). Cada vez que uno de sus seguidores realiza una de estas acciones es considerada como una señal de autoridad y calidad del contenido del usuario.

3) Impacto en la Red: determina la influencia que a su vez tienen los seguidores que están dentro del alcance real del usuario evaluado. Mide, por tanto, la “calidad” de sus *followers* (“¿Cómo de influyentes son ellos?”).

La puntuación media de los usuarios de *Klout* ronda los 20 puntos y una cifra superior a 50 ya sitúa al usuario en el percentil 95 (**Fernández**, 2011b). Junto con dicha puntuación, *Klout* categoriza al usuario dentro de un estilo de liderazgo. Para ello, toma en consideración cuatro parámetros:

- tendencia del usuario a compartir y crear contenido;
- propensión a hacer un uso ocasional o constante de las redes sociales;

- disposición a escuchar o a participar en la Red;
- carácter variado o específico del contenido de sus mensajes.

Como representación gráfica de los estilos de liderazgo se ha optado por la disposición de dichos parámetros a modo de planos (los dos primeros como ejes horizontales paralelos y, en perpendicular respecto a ellos, los dos últimos como ejes verticales). El resultado es una cuadrícula de dieciséis cuadrados, en la que cada tipo de liderazgo halla su ubicación en función de su relación con dos de aquellos polos, como se aprecia en la figura 1.

Las esquinas de la cuadrícula acogen los perfiles de *influentials* más claramente definidos:

- curador: muy participativo y propenso a compartir contenidos;
- famoso: crea contenido de tipo muy variado;
- observador: escucha y utiliza las redes sociales de manera ocasional;
- especialista: tiene una presencia online constante, difundiendo contenido específico.

Completan el cuadro los demás tipos de liderazgo (gran emisor, creador de tendencias, sindicador, alimentador, líder de opinión, experto, aficionado, conversador, socializador, tejedor de redes, explorador, activista). De cada uno de ellos se ofrece una breve descripción (**Berry**, 2010). Esta clasificación parece mejorable, pues resulta un tanto inoperante una taxonomía tan prolija: la necesidad de completar la cuadrícula con dieciséis especies de líderes lleva a que apenas se adviertan diferencias entre algunos estilos de influencia.

Además de dicha categorización, y basándose en un análisis de contenido de los mensajes digitales del usuario, *Klout* determina los temas sobre los que ejerce su liderazgo, elaborando así listados con las personas o marcas más influyentes en cada uno de los temas (educación, periodismo, negocios, entretenimiento, tecnología, etc.). Estos listados cumplen una función de detección de potenciales prescriptores para

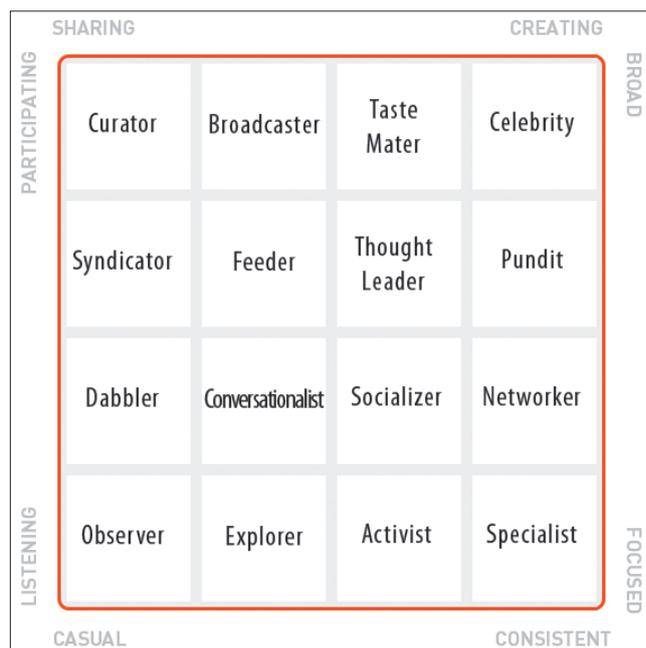


Figura 1. Estilos de influencia de *Klout*

las marcas externas interesadas, pues *Klout* permite a éstas que ofrezcan productos y ventajas en exclusiva (*perks*)¹ a los expertos y personas más influyentes en los temas relacionados con aquellas marcas.

Ante las dudas éticas que esta iniciativa pueda generar, *Klout* defiende que el usuario beneficiado no necesariamente ha de promocionar o hablar bien de la marca en cuestión, pues le asegura que tiene libertad para “decir al mundo que amas el producto, que no te gusta o bien no decir nada” (*Klout*, 2011).

La existencia de los *perks* y el creciente interés que ha generado en las marcas parece indicar la consolidación del índice *Klout* como referencia de la influencia personal digital. Las marcas buscan así identificar a los líderes de opinión sobre las materias que más les afectan a ellas, sacando provecho de diferentes modos: obteniendo una mayor viralidad en la promoción de sus nuevos productos o servicios, facilitando la entrada de éstos en el mercado, conociendo qué aspectos de dichos productos deben ser mejorados para una mayor satisfacción del usuario, etc. Es una dinámica de influencia y difusión de opiniones que recuerda algunas de las conclusiones de **Katz** y **Lazarsfeld** (1955). La modificación de opiniones y conductas es más factible en el ámbito de las decisiones de compra y consumo que en el de los asuntos políticos.

3.2. *PeerIndex*: autoridad y reputación online

Creada en julio de 2009, se define como una herramienta para la medición y comprensión del “capital social” que una persona ha conseguido en el medio online. Es un concepto que se asienta, a diferencia de la propuesta de *Klout*, no tanto sobre la influencia sino sobre la noción de autoridad y reputación.

En las *faqs* e información de ayuda de su web, *PeerIndex* toma distancia con el modelo de medición de *Klout*, que ha recibido numerosas críticas (**Pelzer**, 2010; **Livingston**, 2010; **Sheehy**, 2011) fundadas en estudios y experimentos con las métricas de *Klout*, que señalan que en realidad es más bien un baremo de la notoriedad y exposición de los usuarios en las redes sociales (**Méndez-Parra**, 2011). Pese a que desde la compañía afirmen lo contrario, parece evidente la relevancia que en *Klout* adquiere el factor cuantitativo (número de *tweets* enviados y de *followers* del usuario, frecuencia en la actualización del *status*, etc.). Indicadores todos ellos controlables por el usuario, que podría con cierta facilidad ver aumentado su índice de influencia.

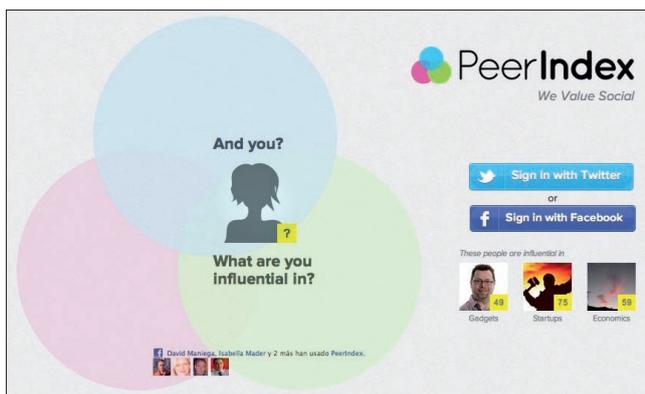


Figura 2. *PeerIndex*

Frente a esto, desde *PeerIndex* sostienen que el mero hecho de ser popular, sea de manera más o menos natural o forzada, no es representativo de la autoridad de un usuario. Ellos buscan medir la reputación del usuario en los temas específicos en torno a los que desempeña su actividad online (puede monitorizar *Twitter*, *Facebook*, *LinkedIn*, *Quora* y blogs). Para ello establece ocho grandes temas que permiten dibujar una especie de “huella digital” de la resonancia de una persona dentro de cada uno de esos ámbitos. Estas áreas (que a su vez incluyen más de 8.000 temas específicos) son:

- 1) arte, medios y entretenimiento;
- 2) tecnología e internet;
- 3) ciencias y medio ambiente;
- 4) salud;
- 5) deportes;
- 6) actualidad, política y sociedad;
- 7) negocios y economía;
- 8) ocio y estilo de vida.

Es una clasificación más sistemática y holística que el repertorio de temas de *Klout*. De hecho *PeerIndex* sólo aplica su ranking a los temas que previamente tiene catalogados y ubicados dentro de las áreas. Para eso ha de detectar la existencia de una comunidad constituida y en conversación en torno a una de dichas cuestiones.

Desde *PeerIndex* defienden que el mero hecho de ser popular no es representativo de la autoridad de un usuario

Indica en un índice de 0 a 100 la puntuación total del usuario según tres componentes. Si en *Klout* eran alcance real, amplificación e impacto en la Red, en este caso son:

- 1) Autoridad: “la medida de la confianza”, el modo en que las demás personas confían en las recomendaciones y opiniones del usuario evaluado, de modo general y en temas concretos. Es una posición relativa respecto al resto de usuarios en un determinado tema. Se calcula en una escala de 0 a 100 (una cifra superior a 40 ya coloca al usuario entre las autoridades de un tema).
- 2) Audiencia: equivalente a los parámetros de alcance real y amplificación de *Klout*, puesto que mide el impacto de las acciones del usuario sobre sus seguidores y la manera en que éstos enganchan con su contenido.
- 3) Actividad: representa la frecuencia de actualización del *status* del usuario en relación con el resto de la comunidad *PeerIndex* en torno a un tema. Según sea la naturaleza del tema (y exija una conversación en torno a él más o menos continua), una actividad excesiva podría considerarse *spam* y penalizar al usuario, o viceversa: que su intervención en el debate sea tan esporádica que llegue a ser irrelevante.

A través de un análisis semántico del contenido de los urls compartidos por el usuario en sus *tweets*, *PeerIndex* determina los 5 temas más relevantes para él. Esta información aparece en su perfil de usuario junto con las fuentes a las que recurre con mayor frecuencia. Sin embargo, y es algo

	<i>Klout</i>	<i>PeerIndex</i>
Fundación	Agosto de 2008 (California, EUA)	Julio de 2009 (Londres, Reino Unido)
Concepto central	Influencia	Autoridad / reputación
Número de usuarios	100 millones	45 millones
Redes sociales monitorizadas	<i>Twitter, Facebook, Google+, LinkedIn, Foursquare, YouTube, Instagram, Tumblr, Blogger, WordPress, Last.fm, Flickr</i>	<i>Twitter, Facebook, LinkedIn, Quora, Blogger, WordPress</i>
Factores que componen el índice	Alcance real, amplificación e impacto en la Red	Autoridad, audiencia, actividad
Clasificación de perfiles	16 estilos de liderazgo	8 grandes áreas temáticas
Valor añadido	<i>Perks</i> (ventajas y productos en exclusiva para el usuario)	<i>PeerPerks</i>

Tabla 1. Análisis comparativo de *Klout* y *PeerIndex*

reprochable al conjunto de parámetros medidos por esta aplicación, es bastante confuso el modo con que es explicada toda la información. Al igual que *Klout*, por razones de competitividad empresarial no hay transparencia en cómo se llevan a cabo los algoritmos de cálculo.

4. Conclusiones

La medición de la influencia personal en internet requiere, en primer lugar, un conocimiento profundo del entorno online y de las interacciones comunicativas que tienen lugar en él. El marco general que ofrece el índice *Klout* es una buena aproximación. Así lo reflejan sus estilos de liderazgo, en los que se reconoce el valor de la conversación y la construcción de redes entre los contactos sociales, la capacidad de filtrado del conocimiento, la habilidad para generar compromiso y mover a la acción, la detección y creación de nuevas tendencias... Son elementos distintivos del fenómeno 2.0 y como tales han de ser punto de referencia para la comprensión de esta realidad. Por su parte, también es muy atractiva la apuesta de *PeerIndex* por revalorizar el concepto de autoridad, que nunca perderá vigencia en las interacciones humanas, con independencia del ámbito en que se desarrollen.

La mayor objeción que puede hacerse a ambos es que su materia de análisis en realidad no se puede reducir a un número. Por decirlo con el acertado quiasmo de **Albert Einstein**: “no todo lo que cuenta puede ser contado y no todo lo que puede ser contado cuenta”. En el ejercicio y desarrollo de la influencia y de la autoridad intervienen numerosos elementos que escapan a las mediciones. Es el factor humano, tan imprevisible como rico en valores intangibles.

Por otra parte, y como consecuencia de su carácter comercial, estas iniciativas sufren la carencia de un bagaje conceptual verdaderamente sólido sobre qué son la influencia y la autoridad y qué parámetros sirven para cuantificarlas. Más aún, sus propios creadores han reconocido que el campo de las métricas aplicadas a las redes sociales está todavía dando sus primeros pasos y es mejorable (**Azhar**, 2011). Muestra de ello son los sucesivos ajustes que ha experimentado el sistema de medición de *Klout*, el último de los cuales suscitó airadas quejas por parte de centenares de usuarios que vieron caer drásticamente su puntuación (**Rust**, 2011). Este hecho sirve para constatar la relevancia adquirida por las herramientas de medición de la influencia, especialmente en el entorno del *marketing* digital. Así lo demuestra el auge de nuevos competidores como *Kred* o *Traack*. Pero en

este fenómeno también está el deseo de los usuarios por medirse frente a los demás y verse con un índice alto de influencia. En definitiva, este tipo de servicios se beneficia del creciente interés por la gestión del ego o *vanity management*.

La influencia no puede reducirse a un algoritmo que, además, es mejorable. Entre otras razones porque ni procesa todo el universo de usuarios (sólo aquellos que se han registrado en *Klout* o *PeerIndex*) ni tampoco todo el conjunto de plataformas sociales. Tampoco se puede descartar que el objetivo de medición de la influencia se vea ocasionalmente desplazado por otros intereses lucrativos, derivados de la naturaleza de *Klout* y *PeerIndex* como iniciativas comerciales privadas.

“Queda el reto de poner en relación la influencia digital de una persona con su influencia en el mundo presencial”

Por todo lo anterior, es deseable que desde el mundo académico surjan propuestas teórico-empíricas de medición de la influencia personal digital, acordes con el rigor y seriedad exigibles a las investigaciones científicas. Queda por delante el reto de hallar la manera de poner en relación la influencia digital de una persona con su influencia en el mundo presencial, examinando en toda su riqueza las múltiples concreciones en que puede encarnarse dicha correspondencia: liderazgo equivalente en ambos planos, ausencia de liderazgo en uno de esos ámbitos, influencia en asuntos de temática diversa según el entorno sea online u offline, etc.

5. Nota

1. Abreviatura del inglés *perquisite*, bono, ventaja o beneficio marginal, adicional a un salario.

6. Bibliografía

Azhar, Azeem. “Don’t worry. Be circumspect (and happy)”. *PeerIndex. Understand your social capital.* October 28th, 2011. <http://blog.peerindex.com/dont-worry-be-circumspect-and-happy>

Bandura, Albert. *Teoría del aprendizaje social.* Madrid: Espasa-Calpe, 1982. ISBN: 84 2396 506 6

Beal, George M.; Bohlen, Joe M. "The diffusion process". *Special report n. 18, Agriculture extension service*, 1957, n. 1, pp. 56-77.

Berry, Megan. "Better know the Klout classes". *Measuring online influence. The official Klout blog*. August 12th, 2010. <http://corp.klout.com/blog/2010/08/better-know-the-klout-classes/>

Berry, Jon; Keller, Ed. *The influentials: one American in ten tells the other nine how to vote, where to eat, and what to buy*. New York: Free Press, 2003. ISBN: 978 0743227292

Boyd, Danah; Ellison, Nicole B. "Social network sites: definition, history, and scholarship". *Journal of computer-mediated communication*, 2007, v. 13, n. 1, article 11. <http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html>

Brosius, Hans-Bernd; Weimann, Gabriel. "Is there a two-step flow of agenda-setting?" *International journal of public opinion research*, 1996, v. 6, n. 4, pp. 323-341. <http://dx.doi.org/10.1093/ijpor/6.4.323>

Castells, Manuel. *La era de la información. Vol. I: La sociedad red*. México, DF: Siglo XXI, 2001. ISBN: 84 2064 246 0

Fernández, Joe. "A new era for Klout scores". *Measuring online influence. The official Klout blog*. October 19th, 2011. <http://corp.klout.com/blog/2011/10/a-new-era-for-klout-scores>

Fernández, Joe. "Understanding the Klout score, part I". *Measuring online influence. The official Klout blog*. November 18th, 2011. <http://corp.klout.com/blog/2011/11/understanding-the-klout-score-parti>

Fumero-Reverón, Antonio. "IRC 2.0. Medios para la información, la relación y la comunicación en la web 2.0". *El profesional de la información*, 2011, nov.-dic., v. 20, n. 6, pp. 605-609. <http://www.elprofesionaldelainformacion.com/contenidos/2011/noviembre/01.pdf>
<http://dx.doi.org/10.3145/epi.2011.nov.01>

Gitlin, Tod. *Enfermos de información: de cómo el torrente mediático está saturando nuestras vidas*. Barcelona: Paidós, 2005. ISBN: 84 4931 713 4

Gladwell, Malcolm. *The tipping point: how little things can make a big difference*. Boston: Little, Brown & Company, 2000. ISBN: 978 0316346627

Katz, Elihu; Lazarsfeld, Paul F. *Personal influence: the part played by people in the flow of communication*. Illinois: Free Press, 1955. ISBN: 978 1412805070

Klout. "Why do brands want to give me free stuff?" *Frequently asked questions*, 2011. <http://klout.com/corp/faq>

Lazarsfeld, Paul F.; Berelson, Bernard; Gaudet, Hazel. *The people's choice: how the voter makes up his mind in a presidential campaign*. New York: Duell, Sloan and Pearce, 1944. ISBN: 978 0231031585

Livingston, Geoff. "5 primary problems with Klout". *Geoff Livingston*. Dec. 16th, 2010. <http://geofflivingston.com/2010/12/16/five-primary-problems-with-klout>

Méndez-Parra, Rafael. "Redefiniendo a Klout: un juego social de notoriedad". *Mis apis por tus cookies*, oct. 2011. <http://www.misapisportuscookies.com/2011/10/redefiniendo-a-klout>

Orihuela, José-Luis. "eCommunication: the 10 paradigms of media in the digital age". En: Salaverría, Ramón; Sádaba, Charo (eds.). *Towards new media paradigms: content, producers, organisations and audiences*. Pamplona: Ediciones Eunat, 2004, pp. 129-135. ISBN: 84 7768 150 3

Pelzer, Adriaan. "Klout is broken". *Blog of RAAK (Digital & Social Media Agency London)*, Dec. 2th, 2010. <http://wewillraakyou.com/2010/12/klout-is-broken>

Pérez-Latre, Francisco J. "Paradojas de la comunicación digital". *Aceprenta*, 29 julio 2011. <http://www.aceprenta.com/articles/paradojas-de-la-comunicacion-digital/>

Rogers, Everett. *Diffusion of innovations*. New York: Free Press, 1962. ISBN: 978 0743222099

Rust, Ash. "A more accurate, transparent Klout score". *Measuring online influence. The official Klout blog*, Oct. 26th, 2011. <http://corp.klout.com/blog/2011/10/a-more-accurate-transparent-klout-score>

Sheehy, Chris. "Klout makes you as much a marketing maven, as Farmville makes you a farmer". *Social media today*, Sept. 21th, 2011. <http://socialmediatoday.com/sidewalkbranding/362391/klout-makes-you-much-marketing-maven-farmville-makes-you-farmer>

The cocktail analysis. *Tercera oleada del Observatorio de las Redes Sociales*, febrero 2011. <http://www.slideshare.net/TCAnalysis/tca-observatorio-redes-sociales2011publico>

Weimann, Gabriel. *The influentials: people who influence people*. Albany: State University of New York Press, 1994. ISBN: 978 0791421413

7. Sitios web

Klout
<http://klout.com>

Kred
<http://kred.ly>

PeerIndex
<http://www.peerindex.com>

Traackr
<http://traackr.com>

TweetLevel
<http://tweetlevel.edelman.com>

Twitalyzer
<http://twitalyzer.com>

Twitter Grader
<http://tweet.grader.com>

Retweetrank
<http://www.retweetrank.com>



GEOIBLINE Y SU COMPARACIÓN CON OTRAS BASES DE DATOS GEOGRÁFICAS



Eva Novotná



Eva Novotná, doctora en filosofía y en gestión, es directora de la *Biblioteca de Geografía* y de la *Colección de mapas* de la *Facultad de Ciencias*, de la *Univerzita Karlova (Universidad Carolina)*, Praga. Gerente del proyecto *Bibliografía geográfica online de la República Checa*. Ha trabajado durante más de 20 años en bibliotecas especializadas, 15 de ellos como gerente. Dirige proyectos de automatización de bibliotecas, así como de autoridades y bibliografías. Co-dirige el proyecto *Tecnología para el procesamiento de las colecciones de mapas de la República Checa*. Autora de artículos, exposiciones especiales, catálogos y del libro *Geografía de apoyo a la información* (2009) y co-autora del libro *Bibliografía geográfica online de la República Checa* (2011).

Geografické knihovny, Mapové sbírky
Geografická sekce PŘF, Univerzity Karlovy
Albertov, 6. 128 43 Praha 2, República Checa
novotn48@natur.cuni.cz

Resumen

Se describe la creación de la base de datos *Geobibline* y se compara con otras similares en geografía. Históricamente, las bases de datos tradicionales se crearon con el trabajo bibliográfico metódico llevado a cabo generalmente por una sola institución, y también, en las últimas décadas por bibliotecas que cooperan en un catálogo colectivo. Es el caso que presentamos: gracias a la colaboración entre 20 bibliotecas especiales checas se ha creado en sólo 4 años una base de datos que contiene 162.000 registros bibliográficos sobre temas geográficos de la República Checa, inicialmente sólo del siglo 20. Adicionalmente, la colaboración con el proyecto *Temap* ha permitido ampliar la cobertura temporal hasta el inicio de los libros impresos (1450). La tipología de los documentos extractados es amplia e incluye la literatura gris. A los registros se unen los objetos (mapas, artículos, portadas de periódicos), las direcciones url de los textos completos y también enlaces a *Google Books*. La base de datos realizada se ha comparado con otras equivalentes de Alemania, Polonia y EUA usando unos encabezamientos de materias concretos, comparables a nivel internacional. La sostenibilidad de *Geobibline* está garantizada gracias a la importación automática de datos y a su conexión con *Temap*.

Palabras clave

Bases de datos bibliográficas, Bibliotecas geográficas, *Geobibline*, *Geodok*, *Geobib*, *Bibliografia geografii Polskiej*, República Checa.

Title: *Geobibline* and how it compares with other geographic databases

Abstract

The development of the *Geobibline* database is described. Historically, traditional databases were created using bibliographic methods usually carried out by a single institution, but in recent decades multiple libraries have often cooperated to create joint catalogues. This is the case presented in this article which describes the collaboration of 20 Czech special libraries resulting in the creation of a database that contains 162,000 bibliographic records on geographical topics related to the Czech Republic. The project, completed in just four years, initially covered only 20th century materials, however, collaboration with the *Temap* project has expanded coverage back to early printed books (1450 A.D.). The typology of the documents referenced is broad and includes gray literature. Records are linked to objects (maps, articles, front pages of newspapers) and there are full text links to *Google Books*. The database has been compared with counterparts in Germany, Poland, and the United States using internationally comparable specific subject headings. The sustainability of *Geobibline* is ensured by the automatic import of data and their connection with *Temap*.

Keywords

Bibliographic databases, Geographical libraries, Library cooperation, *Geobibline*, *Geodok*, *Geobib*, Polish geographic bibliography, Czech Republic.

Novotná, Eva. "Geobibline y su comparación con otras bases de datos geográficas". *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 304-311.

<http://dx.doi.org/10.3145/epi.2012.may.12>

Artículo recibido el 15-01-12
Aceptación definitiva: 16-03-12

Introducción

El fundador del *Instituto Geográfico* de la *Universidad Carolina (UC)* en Praga, **Václav Švambera** (1866-1939), tenía grandes expectativas sobre las bibliografías geográficas de la República Checa hace ya un siglo (**Švambera; Basl**, 1911). Era lógico: el trabajo bibliográfico se considera la forma básica de sistema de información de cualquier disciplina académica sólida.

Las bibliografías impresas checas creadas en el campo de la geografía a lo largo del s. XX eran muy dispersas –cronológica, temática e institucionalmente-. Los registros bibliográficos de las bases de datos de las principales bibliotecas, catálogos colectivos en checo o en la bibliografía nacional eran insuficientes (**Novotná**, 2009).

Un programa del *Ministerio de Cultura* checo titulado *Abrir el acceso y proteger las fuentes culturales, artísticas y científicas* ayudó a resolver esta crisis de falta de fuentes de información: entre 2008 y 2011 fue posible crear una base de datos en libre acceso (sin necesidad de licencia) llamada *Geobibline* (*Geografická knihovna et al.*, 2008-2011), dirigida a especialistas en geografía, profesores, estudiantes, estudiosos y público en general.

Para implementarla se tuvieron en cuenta experiencias tradicionales extranjeras, tales como el repertorio internacional francés *Bibliographie internationale géographique* (**Prodig**, 1995-), las bibliografías de **Chauncy D. Harris** (1975, 1976, 1985), la monografía *Geographic information* de **Jenny M. Johnson** (2003) y especialmente las bases de datos bibliográficas geográficas *Geodok* (del *Institut für Geographie*, 2011), *Bibliografia geografii Polskiej* (**Gazicka; Peliwo**, 2011) y *Geobib* (*American Geographical Society Library*, 2011).

1. Creación de la base de datos

1.1. Criterios para la selección de documentos

Al principio hubo que establecer los criterios de selección. Además de la atención básica a la geografía, el otro criterio fue el aspecto checo, es decir, documentos de autores checos y sobre la República Checa. Se asumió que las bases de datos temáticas en inglés, alemán y polaco cubrirían la mayoría de las revistas con factor de impacto, revistas evaluadas por pares, y otros documentos en esos idiomas, relacionados con Chequia.

1.2. Secuencia de trabajo

Se extrajeron los registros del catálogo centralizado de la *Universidad Carolina (UC)* y se creó *Geobibline* como base de datos separada. Después de una serie de negociaciones con la *Biblioteca Nacional (NK)* y el *Centro de Computación de la UC (ÚVT UK)*, se tomó la decisión de importar los registros bibliográficos de los tipos de documentos que no estaban incluidos en la base de datos original de la *Biblioteca geográfica* o lo estaban sólo parcialmente. Los primeros fueron los documentos cartográficos de la *NK* –al ser depositaria del depósito legal la colección de estos documentos era la más grande y completa (en 2008 contenía 6.000 registros cartográficos relacionados con Chequia)-. Además se importaron artículos de temas geográficos de la bibliografía nacional realizada por la *NK* desde mediados de los 90. Pos-

teriormente se copiaron los registros de monografías, documentos electrónicos, tesis y revistas de la *Biblioteca de Geografía* de la *Facultad de Ciencias* de la *Universidad Carolina (FCUC)* a la base de datos *Geobibline*.

Para llevar a cabo las búsquedas online en las bases de datos de la *NK*, a principios de 2008 el *ÚVT UK* prestó un cliente de interfaz gráfica que almacena los registros seleccionados en el servidor de la *NK*, desde donde se envían al servidor *ÚVT UK*. Las búsquedas fueron realizadas utilizando encabezamientos de materia en combinación con el lugar de publicación y el tipo de documento.

En 2009 se hicieron búsquedas adicionales en la *NK*, inicialmente con un proceso semi-automático, que más tarde fue automatizado por completo. Las bibliografías originales empezaron a recopilarse en las bibliotecas participantes en 2008. Otros tipos de documentos se exportaron de forma automática todos los días desde la *Biblioteca de Geografía* en la *FCUC* identificados a través del campo <GEO>. La primera vez los registros de las otras bibliotecas se importaron offline pero desde 2011 la importación ha sido totalmente automatizada a intervalos regulares, no sólo de la *NK*, sino también de la *Biblioteca de Moravia*, las bibliotecas de la *Academia de Ciencias*, bibliotecas de universidades, científicas estatales y otras organizaciones profesionales. Cada mes la base de datos aumenta en varios cientos de registros procedentes de fuentes externas. A principios de 2012 la base de datos *Geobibline* contenía 162.000 registros bibliográficos, con un incremento anual de alrededor de 20.000 registros.

1.3. Metodología

La metodología básica para la creación de la base de datos fue elaborada y actualizada de forma progresiva. Los registros siguen las reglas de catalogación nacionales e internacionales, fijándose el nivel mínimo del registro MARC21. Se utilizaron encabezamientos de materia en checo. Para los participantes en el proyecto se creó una interfaz y un acceso a la base de datos a través del protocolo Z39.50.

1.3.1. Método del campo atributo GEO

Hubo que resolver un problema fundamental derivado del trabajo simultáneo en dos bases de datos (el catálogo centralizado de la *UC* y *Geobibline*): había que catalogar los documentos una sola vez, al mismo tiempo que se exportaban (preferiblemente de forma automática) a la otra base de datos. Al final, se optó por introducir atributos en los campos como la solución más simple. De acuerdo con el atributo <GEO> predefinido en el catálogo de la sub-biblioteca de geografía de la *FCUC*, se identificaba el registro bibliográfico y se exportaba automáticamente del catálogo centralizado de la *UC* a la base de datos *Geobibline*.

1.3.2. Comprobación de duplicados

El problema de los registros duplicados comenzó a resolverse para los registros de monografías recién importados y para los nuevos registros cartográficos que las bibliotecas participantes podrían haber enviado más de una vez.

Antes de la primera importación era necesario tener ya resuelto el tema de los registros duplicados. En el sistema de bibliotecas *Aleph* de la *UC*, la comprobación automática se

<http://www.geobibline.cz>

hizo utilizando al menos dos campos de control y el campo <001> (MARC21). El sistema tenía que marcar los registros duplicados e informar de los errores, a continuación el administrador de la base de datos marcaba los duplicados reales y el sistema posteriormente terminaba importando los registros. El proceso se depuró y automatizó con éxito. Un correo electrónico automático informa de todos los registros duplicados encontrados (Novotná, 2011).

1.3.3. Textos completos

Desde el principio la base de datos *Geobibline* fue entendida como un sistema que en el futuro daría acceso a los textos completos. La mayoría de los editores estuvieron de acuerdo, aunque no tienen demasiada experiencia en tratar el asunto de los derechos de autor, y en muchos casos no existen contratos entre los editores y los autores.

Un problema era la inexistencia de textos completos electrónicos, pero la situación cambió pronto al comenzar muchas editoriales a publicar electrónicamente.

Licencias

Los derechos de autor y los temas legales fueron tratados de forma gradual. Se empezó a negociar con los propietarios de derechos de autor que ya habían expuesto sus documentos en internet, preguntándoles sobre la posibilidad de enlazar a los textos completos de su web. Más tarde pedimos al representante legal de la editorial o de la institución científica que había publicado el documento su consentimiento por escrito para guardar los datos en el repositorio de la *Facultad* y publicarlos a través de la base de datos de *Geobibline*. Con el tiempo empezamos a firmar contratos de licencia continua de provisión de ficheros electrónicos de los textos completos. Los contratos también incluyen la autorización de la editorial de hacer accesible y tener en la base de datos *Geobibline* todos los volúmenes publicados con anterioridad. Se respeta la política editorial de cada revista en particular, especialmente en relación al embargo por un período de tiempo específico.

La información sobre copyright, modificada a petición del titular de los derechos de autor, se incluye en el registro bi-

bliográfico en el campo 500 del MARC 21. El usuario siempre está informado acerca del copyright, pues antes de abrir el texto completo puede leer lo siguiente:

“Advertencia sobre restricciones definidas por la *Ley de Derecho de Autor (Ley N. 121/2000 Coll.)*. Bajo las condiciones definidas por la *Ley de Derecho de Autor*, las bibliotecas y archivos están autorizados a realizar libremente fotocopias u otras reproducciones de trabajos de autor. La condición específica es que esas fotocopias u otras reproducciones no pueden ser utilizadas para otros propósitos que no sean el estudio privado o la investigación. Si un usuario solicita un documento y realiza una

fotocopia o reproducción en conflicto con las condiciones definidas por la *Ley de Derecho de Autor*, será responsable de infracción de copyright. Esta biblioteca se reserva el derecho de rechazar un pedido de copias, si en su opinión el cumplimiento de dicho pedido supondría una infracción de copyright” (Información de copyright en *Geobibline*).

Además de los artículos, los textos completos de las tesis académicas también van siendo accesibles. La base de datos proporciona acceso a las tesis de la *Universidad Masaryk*, de Brno y de la *UC*. El usuario externo tiene que registrarse para poder acceder a los textos completos. Aquí también se le informa sobre las condiciones de la licencia:

“Toda la información obtenida mediante la búsqueda en las tesis publicadas por la *UC* de Praga, no puede ser usada con ánimo de lucro ni ser declarada como estudio, investigación o actividad creativa de otra persona que no sea el autor” (condiciones de licencia del repositorio de la *UC*).

La *Biblioteca Nacional* ha elaborado un contrato para el suministro de recursos online, de acuerdo con el cual los proveedores pueden dar el consentimiento a la biblioteca para que los conserve permanentemente y dé acceso a los mismos. Este contrato se tomó como modelo para *Geobibline*. También se redactó un modelo de contrato de licencia para los autores (o sus herederos propietarios de los derechos) con el fin de obtener su consentimiento para la publicación. En la República Checa los derechos expiran a los 70 años de la muerte del autor. Teóricamente es posible publicar sin permiso expreso obras de autores fallecidos antes de 1942, pero aun así pueden surgir obstáculos legales.

Con el *Aleph digital asset module (ADAM)* se enlazaron a las referencias bibliográficas los textos completos de (principalmente) artículos, portadas de revistas o muestras especiales de manuscritos. *ADAM* es un módulo especial del software *Aleph 500* que permite integrar objetos digitales con sus metadatos (fichas descriptivas) en el entorno de la biblioteca. Los 7.360 objetos digitales adjuntos (mayo de 2012) se pueden buscar por su texto completo y luego visualizarse (ÚVT, 2008).

Otra opción es conectar los artículos disponibles online mediante sus urls en el campo 856 de MARC. Desde 2010 se realizan actualizaciones y controles periódicos de urls rotos mediante una función de *Aleph*. En mayo de 2012 la base de datos contiene más de 12.700 direcciones url.

1.4. Cooperación entre bibliotecas

El proyecto se inició en la *Biblioteca de Geografía* de la *FCUC*, de Praga, en la biblioteca de la *Universidad Masaryk*, de Brno y en el *Departamento de Geografía* de la *Facultad de Ciencias, Universidad J. E. Purkyne*, de Usti nad Labem. Las dos primeras bibliotecas trabajaban con el software *Aleph*, por lo que la exportación de datos fue fácil de configurar. La tercera biblioteca utiliza el software *Rapid Library* y la situación fue mucho más complicada. Tomó mucho tiempo depurar las importaciones.

Poco a poco fue creciendo el interés por parte de otras bibliotecas, hasta un total de 20. En la práctica esto significa

que las “grandes” bibliotecas (*Nacional, Academia de Ciencias*, estatales de investigación, etc.) suministran registros bibliográficos a las bibliotecas más pequeñas y especializadas. Para compartir los registros se tuvieron que solventar varias dificultades debidas a las diferentes opciones de catalogación e indexación en diversos campos MARC locales.

Con la *Biblioteca Nacional* no hubo problemas técnicos porque trabaja con *Aleph*, pero sí los hubo con las pequeñas bibliotecas que disponen de sistemas diferentes como *ISIS, KP-Win*, etc. Algunos no soportan el formato MARC 21, sólo Unimarc. A pesar de que se trabajara con normas e instrucciones comunes, había diferencias. A menudo surgían “restos” de otros sistemas más antiguos, así como prácticas de catalogación locales (por ejemplo, autoridades locales que no estaban vinculadas a las nacionales). Algunas bibliotecas no utilizaban encabezamientos de materias, y a veces los registros importados eran más difíciles de corregir que si se hubieran creado nuevos.

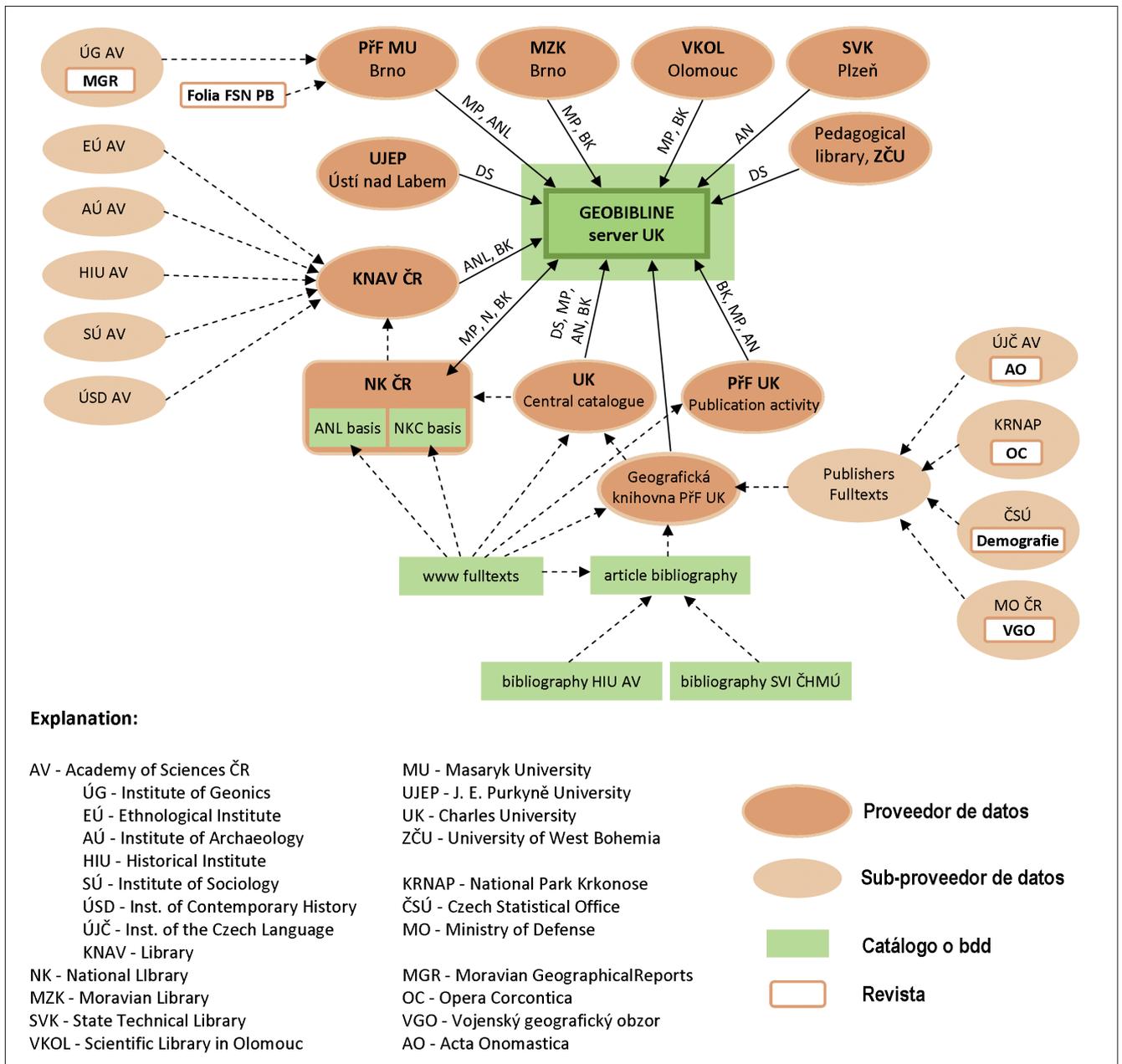


Figura 1. Modelo de proceso de Geobibline

Pero no sólo se importaron registros de los clásicos catálogos de biblioteca. El gran problema fue la importación de los registros de las llamadas "actividades de publicación" de la *Universidad Carolina* que realiza su personal, y que no se incluyen normalmente en el catálogo central. Los registros los habían hecho los mismos autores y estaban llenos de errores. La importación requirió mucho trabajo manual, pero el contenido valía el esfuerzo.

Finalmente, la propia *Biblioteca Nacional* extrajo enlaces url de *Geobibline* para su catálogo. Y también las bibliotecas participantes recibieron *feedback* sobre su catalogación con los errores encontrados.

La base de datos *Geobibline* ha introducido un nuevo modelo de colaboración en el panorama de las bibliotecas checas. Las grandes bibliotecas con colecciones que abarcan por lo general una amplia gama de temas que van desde las humanidades a la tecnología han acordado colaborar estrechamente con las pequeñas bibliotecas de investigación, en su mayoría especializadas en geografía.

2. Acceso de los usuarios

La base de datos se puede consultar online y en DVD. La página principal del opac ofrece entrada tanto a *Geobibline* como a otras bases de datos de instituciones colaboradoras. Se ha creado una versión de consulta en inglés y muchos documentos poseen términos en este idioma, especialmente los trabajos de la universidad.

Los usuarios pueden visualizar, almacenar, enviar por correo-e o localizar los documentos, ya sea en el *Catálogo Colectivo de la República Checa*, en la *NK*, la *Biblioteca de Moravia* en Brno, la *Biblioteca de Investigación* de Olomouc, o en la *Biblioteca del Congreso*.

En 2010 se puso en funcionamiento un servicio de gestión de referencias de la base de datos *Geobibline* que las formatea según *ISO 690* e *ISO 690-2* (Krčál, 2004-2011):

<http://www.citace.com>

3. Comparativa de Geobibline con otras bases de datos internacionales

Hacemos la comparación con otras bases de datos geográficas gratuitas de Polonia, Alemania y Estados Unidos. No se ha hecho con bases de datos comerciales como *GeoRef*, *GeoBase* y otras que tienen restricciones de uso.

3.1. Bibliografía Geografii Polskiej (BGP)

La *Bibliografía geográfica de Polonia* se publica en la *Academia de Ciencias de Polonia* en Varsovia y ha estado disponible en forma impresa desde 1956.

En internet está desde 1999 (Gazicka-Wójtowicz; Peliwo, 2011), cubriendo

el período 1985 a 2006 (lleva, pues, un considerable retraso de producción), y funciona con el software *ISIS*. Contiene 57.203 registros, que a su vez ofrecen información de unos 150.000 títulos de artículos de revista, capítulos de libro, ponencias, mapas y atlas. El crecimiento anual es de 2.500-3.000 registros (Gazicka-Wójtowicz, 2012).

Incluye publicaciones educativas, científicas y populares de geógrafos polacos y artículos geográficos de toda índole. Indiza 150 títulos de revistas polacas y extranjeras.

Los editores son **Dorota Gazicka** y **Jan Peliwo**.

La *Academia de Ciencias de Polonia* mantiene un repositorio con 1.036 objetos digitales:

<http://rcin.org.pl/igipz>

que también aloja al *Repozytorium Cyfrowe Instytutów Naukowych* (repositorio de 16 institutos científicos), que contiene 5.000 publicaciones (mayo de 2012):

<http://rcin.org.pl>

Comparación entre BGP y Geobibline

Los visitantes no polacos pueden echar en falta que haya poca información básica en inglés sobre la base de datos, pues sólo se ofrece un pequeño manual en este idioma. La cobertura temática de la bdd polaca es mucho más amplia que la de *Geobibline*, lo cual se debe a la historia y a los recursos humanos de sus respectivas instituciones. En cambio los registros más antiguos de la polaca son sólo de 1980.

La búsqueda en ambas es bastante similar, aunque el software *ISIS* impone serias limitaciones.

Una vez seleccionados los resultados, hay una opción para ver o imprimir los registros en cuatro tipos de formatos, pero no pueden ser salvados o enviados por correo electrónico. No acepta el formato MARC o la norma *ISO 690*.

Dorota Gazicka, Jan Peliwo
Bibliography of Polish Geography since 1985

The database has been created by Central Library of Geography and Environmental Protection of the Institute of Geography and Spatial Organization Polish Academy of Sciences
 © [Central Library of Geography and Environmental Protection](#) | [more info about database](#) | [List of indexed journals](#) | [User's guide](#) |

Software **Expertus** WWW for CDS/ISIS databases © **SPLENDOR**, Poznań
 In fields '...' - any words' type several words or word roots nearby; [example](#)
 A term can be selected from the index: [example](#)

Any words from record [] INDEX
 Title [] INDEX
 Author [] INDEX

Combine search: and [] Search Cancel

Search limits
 Document type: any articles from journals book chapters, conference paper monographies, atlases maps
 Select year: (no selection - any year) 2004
 2006 2005 2003 2002 2001 2000 1999 1998 1997 1996 1995
 1994 1993 1992 1991 1990 1989 1988 1987 1986 1985

Search resources of other libraries: (no selection - all libraries from list; other localization given if and
<http://www.cbgiios.pan.pl/bazy/bgp>

3.2. Geodok

La bdd geográfica bibliográfica *Geodok 5.0* es producida por el *Institut für Geographie*, de la *Universität Erlangen*, en Nurenberg, en colaboración con la *Asociación de Geógrafos de las Universidades Alemanas*. Las referencias de los documentos más antiguos datan de 1950, pero la mayoría son posteriores a 1980. Se puso en funcionamiento en 1999 y se actualiza mensualmente. Contiene 153.000 referencias (75% de artículos y 25% de monografías), con un crecimiento anual de 5.000 (Huber, 2012; *Institut für Geographie*, 2011). Usa el software *Aleph*.

Indexa cerca de 500 revistas geográficas en alemán, inglés y francés. Los usuarios realizan unos 3 millones de búsquedas anuales en la versión en inglés (Novotná, 2009).

Los resultados se muestran en un formato corto. Los encabezamientos de materia, un enlace al texto completo en *ZDB-OPAC*, *Elektronische Zeitschriftenbibliothek*, y el servicio *SFX* de resolución de enlaces están claramente situados a la derecha de cada registro. Los resultados se pueden seleccionar, imprimir, convertir a pdf, guardar o enviar por correo electrónico. *Geodok* también permite exportar los resultados a *Citavi* y *EndNote*. Existe una fuente RSS que se puede utilizar para enviar consultas de los usuarios.

3.3. Geobib

Está publicada por la *American Geographical Society Library* de la *University of Wisconsin-Milwaukee*, EUA (*American Geographical Society Library*, 2011), y abarca desde 1985 hasta diciembre de 2005, fecha en la que dejó de actualizarse. Contiene 115.000 registros que siguen siendo accesibles (Peschel, 2012).

Tenía el equivalente impreso *Current Geographical Publications* que terminó en 2003, y desde entonces sólo continuó online (*ProQuest*, 2011). Funciona con *Aleph*.

Los registros están indexados mediante el catálogo de materias de la *American Geographical Society Library*.

3.4. Estructura temática

Para comparar los contenidos de las bases de datos se seleccionaron cuatro áreas temáticas:

- Geografía física: geología, geomorfología, meteorología, climatología, hidrología, edafología, biogeografía y ecología;
- Geografía social/humana: geografía social, económica, del transporte, cultural, política, regional y sociología;
- Demografía;

The screenshot shows the GEODOK search interface. On the left, there are search filters for 'Exact Search' with options for Keyword (Title), Author/Editor, and Year. The main search area has a 'Quick Search' box and buttons for 'AND-Search', 'OR-Search', and 'Reset'. Below the search area, there are search results for 'Barcelona' with details like 'Charnock, G. Ribera-Fumaz, R. (2011) Urban development | Urban planning | Catalonia A new space for knowledge and people? Henri Lefebvre, representations of space, and the production of 22@Barcelona'.

<http://www.geodok.uni-erlangen.de>

- Disciplinas auxiliares: cartografía, sistemas de información geográfica (SIG), teledetección y estadísticas (Raska, 2011).

Podría plantearse la objeción de que existe un gran número de definiciones de geografía, la descripción de la temática de los documentos es subjetiva, y la indización varía en bases de datos diferentes: a veces es más profunda, y otras veces parece ser superficial. Esto apareció con el título general de geografía social/humana, por ejemplo, con sólo 30 documentos clasificados por este concepto en la base de datos de Polonia. Es probable que existan muchos más documentos con descripciones más detalladas.

El gráfico 1 con las disciplinas básicas geográficas que fueron seleccionados inicialmente ofrece la comparación final. La bdd checa tiene la mayor incidencia de geografía física

The screenshot shows the WebVoyage search interface. At the top, there are navigation buttons for 'Database', 'Search', 'Headings', 'Titles', 'Patron', 'Login', 'Request', 'History', 'Help', and 'Exit'. Below that, it shows the search results for 'Madrid' in the GEOBIB database. The results are displayed in a table format with columns for '#', 'Title Long', 'Author', and 'Date'. The first result is 'Industrial en la comunidad de Madrid. / Mayoral' by Mayoral, M. M. (2001).

<http://geobib.lib.uwm.edu>

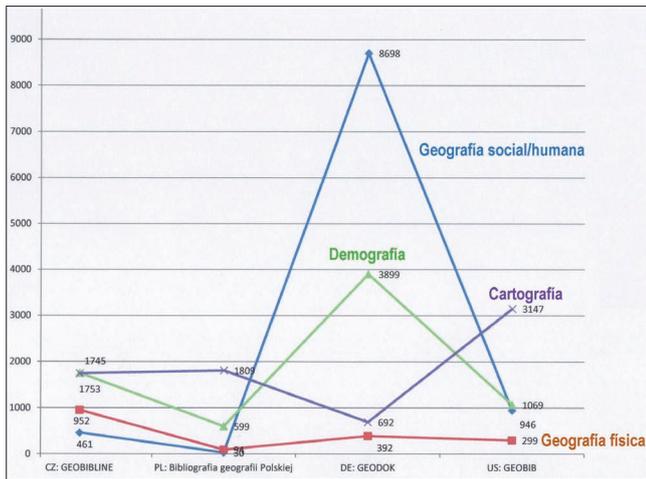


Gráfico 1. Gráfico comparativo de los encabezamientos de materia de geografía social, geografía física, demografía y cartografía en las cuatro bases de datos geográficas internacionales

(952), la alemana es mejor en geografía social (8.698) y demografía (3.899), y la norteamericana es mejor en cartografía (3.147). A partir de esta comparación, la Geodok alemana podría coronarse como “reina” de las bases de datos geográficas.

4. Conclusión

La base de datos Geobibline es ya comparable con los productos extranjeros similares. Se ve que el método de creación en sólo 4 años es un aspecto especialmente interesante (la de EUA lleva 25 años, y la polaca y la alemana llevan 13). Esto fue posible gracias a la colaboración entre 20 bibliotecas. Entre ellas las exportaciones de datos no sólo van de arriba hacia abajo sino también de abajo hacia arriba.

Nombre	Nº de registros	Período
Geobibline	164.000	s. XVI - actual
Bibl. Geografii Polskiej	58.000	1985 - 2006
Geodok	152.000	1950/1980 - actual
Geobib	115.000	1985 - 2005

Poder trabajar con el software Aleph (bdd de Estados Unidos y de la República Checa) permite ciertas funciones que no son posibles con la polaca y la alemana, las cuales ofrecen menos opciones.

Una ventaja de la bdd checa son los adjuntos: objetos y textos completos de los documentos. La alemana también ofrece textos completos a través de la Elektronische Zeitschriftenbibliothek (EZB) en Ratisbona (Regensburg).

Geobibline llena un vacío sobre información geográfica en Centroeuropa, como lo prueba el número de visitas que recibe: 877.000 desde que se puso en funcionamiento en 2009. El mayor interés procede de Estados Unidos, Países Bajos y Alemania.

El proyecto Temap (Tecnología para el acceso a las colecciones de mapas checos), que ha colaborado en la bdd desde 2011, va a incluir más documentos antiguos en texto completo.

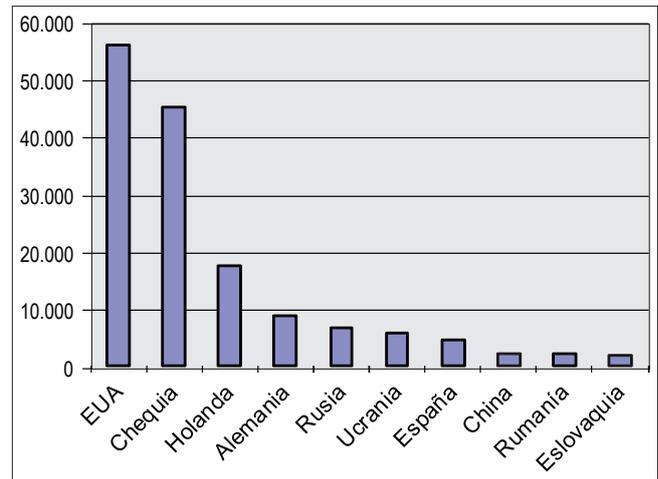


Gráfico 2. Tráfico por países del sitio www.geobibline.cz en 2011

Las bdds geográficas descritas se complementan entre sí, y se distinguen en la tipología de documentos, y en la forma en que son elaboradas. Cada base de datos tiene sus fortalezas y debilidades. En el futuro, sería sin duda beneficioso para los usuarios que se pudieran consultar las cuatro desde un portal único, pero habría que resolver la barrera del idioma. Los documentos de las europeas tendrían entonces que indexarse también con palabras clave en inglés, o habría que usar un diccionario geográfico común multilingüe. Después, a los usuarios se les ofrecería la oportunidad de elegir entrar en una única base de datos cada vez, o consultarlas todas al mismo tiempo. ¿Qué elegirían?

5. Referencias

American Geographical Society Library. *Geobib*. <http://geobib.lib.uwm.edu/cgi-bin/Pwebrecon.cgi?DB=local&PAGE=First>

Elsevier. *Scopus*. <http://www.scopus.com/home.url>

European Science Foundation, ERIH. <http://www.esf.org/research-areas/humanities>

Gazicka-Wójtowicz, Dorota; Peliwo, Jan. (ed.). *Bibliografía geografii Polskiej*. <http://www.cbgi.os.pan.pl/bazy/bgp>

Gazicka-Wójtowicz, Dorota. About *Bibliografía geografii Polskiej* 1985+ (email), 2012.

Geografická knihovna...[et al.]. *Geobibline*. <http://www.geobibline.cz>

Harris, Chauncey D. (ed.) *Guide to geographical bibliographies and reference works in Russian or on the Soviet Union*. Chicago: University of Chicago, 1975.

Harris, Chauncey D. (ed.) *Bibliography of geography*. Chicago: University of Chicago, 1976.

Harris, Chauncey D. et al. (ed.) *A geographical bibliography for American libraries*, Washington, D.C.: Association of American Geographers, 1985. ISBN: 0 89291 193 X

Hüber, Thomas. About *Geodok* (email), 2012.

Johnson, Jenny-Marie. *Geographic information*. Westport, CT: Greenwood Press, 2003. ISBN: 1 57356 392 7

Institut für Geographie. *Geodok*.
<http://www.geodok.uni-erlangen.de>

Krčál, Martin. Citace 2.0.
<http://www.citace.com>

Národní knihovna ČR. ANL.
http://aleph.nkp.cz/F/VI1TUHH6KFAK894RTK7DV3RC7IEKNVKD28JEH17E2LKTLP8BTL-35362?func=file&file_name=base-info

Novotná, Eva. *Informační podpora geografie*. Praha: Univerzita Karlova, Přírodovědecká fakulta, 2009. ISBN: 978 80 86561 42 4

Novotná, Eva, et al. *Geografická bibliografie ČR on-line*. Praha: Vydavatelství VŠCHT, 2011. ISBN: 978 80 7080 773 6

Peschel, Susan. About *Geobib* (Email), 2012.

Prodig. *Bibliographie géographique internationale*, v. 100, tome 1 (1995-).

ProQuest. Ulrichsweb: the global source for periodicals.
<http://www.ulrichsweb.com/ulrichsweb/default.asp?navPage=4&#>

Rada pro výzkum, vývoj a inovace. Seznam recenzovaných neimpaktovaných časopisů vydávaných v ČR.
<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=495942>

Raška, Pavel. "Projekt *Geobibline* a e-learning v geografickém vzdělávání." In: Novotná, E. et al. *Geografická bibliografie ČR on-line*. Praha: Vydavatelství VŠCHT, 2011, pp. 122. ISBN 978 80 7080 773 6

Švampera, Václav; Basl, Antonín (ed.). *Bibliografie české literatury geografické*, v. I., 1906 a 1907. Praha: Geografický ústav České university, 1911.

ÚVT. ADAM.
<http://aleph.cuni.cz/ALEPH-50.html>

Máster en Buscadores: Marketing y posicionamiento.

Modalidad *online*. Ponte a punto

Recuperación de la información, marketing *online* innovador, programas de vigilancia tecnológica y de inteligencia competitiva

Dirección: **Cristòfol Rovira**
Calendario: **diciembre 2012 - junio 2013**
Más información: www.idec.upf.edu/mbusol

PONTE A PUNTO

Másters y Programas de postgrado
en Comunicación y Ciencias Sociales
Universitat Pompeu Fabra. www.idec.upf.edu



ÁREAS DE FORMACIÓN: Comunicación y Periodismo • Derecho • Gestión Cultural y Humanidades • Comunicación y Medios Audiovisuales • Política y Sociedad • Publicidad y Reputación Corporativa • Lenguaje y Traducción • Creatividad en Medios Digitales • Salud y Bienestar Social • Innovación y Gestión de la Información

MAPA DE PROCESOS DE UN SISTEMA DE GESTIÓN DE ACCESIBILIDAD EN UN SERVICIO WEB DE LA ADMINISTRACIÓN PÚBLICA: EL AYUNTAMIENTO DE ZARAGOZA

José-Antonio Salvador-Oliván y María-Jesús Fernández-Ruiz



José-Antonio Salvador-Oliván, doctor en medicina y profesor titular en el *Departamento de Ciencias de la Documentación* de la *Universidad de Zaragoza*, es responsable del *Servicio de Información* y asesor estadístico del *Colegio de Médicos de Huesca* desde 1994. Ha dirigido proyectos de investigación nacionales e internacionales, habiendo obtenido contratos con el *Ayuntamiento de Zaragoza* para implantar el sistema de gestión de accesibilidad en la *Unidad de Gestión de la Web Municipal*, la realización de auditorías internas y estudios de uso y consumo de información.

Universidad de Zaragoza
Departamento de Ciencias de la Documentación
Pedro Cerbuna, 12. 50009 Zaragoza
jaso@unizar.es



María-Jesús Fernández-Ruiz es licenciada en filosofía y letras (geografía) por la *Universidad de Zaragoza*. Documentalista del *Ayuntamiento de Zaragoza*, es la responsable de la *Unidad de Gestión de la Web Municipal* del *Ayuntamiento de Zaragoza* (sede electrónica) desde 1994. Ha obtenido premios a la mejor web de las administraciones públicas que otorga el periódico *El país* y a las mejores prácticas en protección de datos. Ha participado en proyectos de investigación de administración electrónica.

Ayuntamiento de Zaragoza. Web Municipal
Edificio Seminario. Vía Hispanidad, 20. 50071 Zaragoza
mjferuiz@zaragoza.es

Resumen

Se presenta el mapa de procesos que se realizan en la sede electrónica del *Ayuntamiento de Zaragoza*, que forma parte del sistema de gestión de la accesibilidad, requisito para obtener el certificado *Aenor-Marca N* de accesibilidad TIC que garantiza el cumplimiento de las pautas de accesibilidad, aspecto al que está obligado por ley cualquier administración pública. Para identificar los procesos y poder elaborar el mapa, un experto revisó la documentación y mantuvo diversas entrevistas con la dirección y el personal de la organización. Los procesos se clasificaron en tres tipos (estratégicos, clave y de apoyo) apareciendo identificados en el mapa y señalando su secuencia e interacción. En el apartado de resultados se describen las principales actividades realizadas en cada uno de ellos y se analizan sus interrelaciones.

Palabras clave

Sistema de gestión de la accesibilidad, Mapa de procesos, Accesibilidad, Gestión por procesos, Sitios web.

Title: Process map of the accessibility management system in a public administration web service: Saragossa City Council

Abstract

The main objective of this paper is to show the process map for introducing the Accessibility Management System in the Saragossa city council web service. This map is a requirement of the *Spanish Association for Standards and Certification (Aenor)* for the "N" certification of ICT Accessibility, which guarantees users that an internet page is accessible. All public sector organizations are required by law to achieve this norm and display the "N" mark of certification.

To identify the processes and design the process map, an expert reviewed documents and interviewed managers and staff. The processes were classified into strategic, key and support categories and so identified on the process map, which indicates sequences and the relationships between inputs and tasks. The results section describes the main activities and analyzes their inter-relationship.

Keywords

Accessibility management system, Process map, Accessibility, Process management, Web sites.

Artículo recibido el 29-11-11
Aceptación definitiva: 18-01-12

Salvador-Oliván, José-Antonio; Fernández-Ruiz, María-Jesús. "Mapa de procesos de un sistema de gestión de accesibilidad en un servicio web de la administración pública: el Ayuntamiento de Zaragoza". *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 312-317.

<http://dx.doi.org/10.3145/epi.2012.may.13>

Introducción

La accesibilidad web ha adquirido gran importancia en los últimos años, especialmente en los sitios web de las administraciones públicas ya que la legislación vigente, tanto española como europea, obliga a proporcionar información accesible a todos los ciudadanos. El concepto de accesibilidad en el entorno web se refiere a que todas las personas, independientemente de las circunstancias (idioma, cultura, localización geográfica, discapacidades) y dispositivos (hardware y software) que utilicen, puedan acceder correctamente a la información disponible en internet en igualdad de condiciones.

Aenor (Asociación Española de Normalización y Certificación) se ocupa de evaluar si un sitio web cumple las pautas de accesibilidad de la *WAI (Web Accessibility Initiative)* recogidas en la norma *UNE 139803, "Aplicaciones informáticas para personas con discapacidad. Requisitos de accesibilidad para contenidos web"*. Aunque existen tres niveles en estas pautas (A, AA y AAA), *Aenor* certifica sólo aquellos sitios que cumplen con los niveles AA y AAA, expidiendo el certificado de *Accesibilidad TIC*; para conseguir el certificado *Aenor - Marca N de Accesibilidad TIC*, es necesario además que la organización haya implantado y mantenga un sistema de gestión de la accesibilidad (SAG). Según *Aenor*, para la certificación se deberá establecer, documentar, implantar y mantener un sistema de gestión relativo al diseño, desarrollo, gestión de contenidos y mantenimiento de los sitios web. Para ello, la organización debe identificar y documentar los procesos necesarios para que el diseño, desarrollo y mantenimiento del sitio web sea accesible (*Aenor*, 2006).

La implantación de un SAG es similar a un sistema de gestión de calidad basado en la norma *ISO 9001:2000*, que también especifica en su apartado 4.1a) que se deben "identificar los procesos necesarios para el sistema de gestión de la calidad y su aplicación a través de la organización", y en el apartado 4.1b) "determinar la secuencia e interrelación de estos procesos".

La identificación y gestión sistemática de los procesos empleados en una organización y, en particular, las interacciones entre ellos se conoce como "enfoque basado en procesos" (*López-Rey*, 2006) y es especialmente relevante para la implementación de un sistema de gestión de calidad total (*Ahire; Dreyfus*, 2000). Es ampliamente utilizado en la práctica, ya que el análisis y la reorganización de los procesos son prerequisites esenciales para mejorar la eficiencia de una organización (*Heinrich et al.*, 2009). El objetivo de este enfoque es una mejora continua de la efectividad y la satisfacción de todas las partes interesadas: clientes y usuarios del sistema.

Son muchas las organizaciones inmersas en la implantación de sistemas de gestión de calidad basados en las normas *ISO* o *EFQM (European Foundation for Quality Manage-*

ment), donde una definición adecuada de los procesos es una parte necesaria para conseguir resultados de excelencia (*Ruiz-Canela*, 2003). Las dos familias de normas promueven la adopción de un enfoque basado en procesos como principio básico para la consecución de manera eficiente de resultados relativos a la satisfacción del cliente y de las restantes partes del sistema (*Villanueva et al.*, 2005).

La gestión por procesos conlleva básicamente una estructura coherente que representa el funcionamiento de la organización, un sistema de indicadores para evaluar la eficacia y eficiencia de los mismos, y la designación de sus responsables (*Rey-Peteiro*, 2005). La aplicación de este enfoque implica definir las actividades necesarias para conseguir el resultado deseado, establecer los responsables de la gestión de las principales actividades, analizar y medir los resultados, identificar las interfaces de las actividades dentro y entre las funciones de la organización, centrarse en todos los factores (recursos, métodos y materiales) que puedan mejorar las actividades, y evaluar riesgos, consecuencias e impactos de las actividades en los clientes (*ISO*, 2011).

Objetivos

En este contexto y dado que la accesibilidad ha sido siempre una prioridad para el servicio web del *Ayuntamiento de Zaragoza*, éste decidió implantar el sistema de gestión de accesibilidad exigido por *Aenor* para obtener el certificado *Marca N de Accesibilidad TIC*, con el fin de proporcionar servicios y productos con un grado de accesibilidad que alcanzara como mínimo un nivel de adecuación doble A (AA) de las pautas desarrolladas por la *Iniciativa de Accesibilidad Web (WAI)*.

El objetivo de este trabajo es presentar el mapa de procesos que se realizan en la *Unidad de Gestión de la Web* del *Ayuntamiento de Zaragoza*, describiéndolos brevemente y señalando sus interrelaciones e interacción (a un macronivel).

Método

Previamente a la construcción del mapa de procesos, es necesario definir qué es un proceso y su correspondiente mapa. En la norma *ISO 9001:2000* se define un proceso como un conjunto de actividades mutuamente relacionadas que interactúan, y que transforman unos elementos de entrada en elementos de salida. En un contexto de estrategia de información, similar al que nos ocupa, ha sido definido como un conjunto de actividades organizadas en una secuencia lógica (flujo) que incorpora puntos de decisión y de cierre y que están influidas por las normas de la organización o de la empresa (*Gibb et al.*, 2006). En parecidos términos se expresa *Davenport* (1993) al definir un proceso como un ordenamiento específico de actividades a través del tiempo y del lugar, con un comienzo, un final, y unas entradas y salidas bien identificadas.

Un mapa de procesos es su representación gráfica orientada a cumplir la misión de una organización. Como herramienta que es, ha de reflejar la realidad de la organización y ser fácil de comunicar y comprender (Pérez-Fernández-de-Velasco, 2004). Es una ayuda visual que muestra la relación entre las entradas, salidas y tareas (Anjard, 1996) y pone de relieve las interrelaciones internas al representar con flechas los flujos entre los procesos (Duvauferrier *et al.*, 2002).

Para implementar el SAG fue necesario desarrollar un mapa de procesos, y para ello, las primeras acciones consistieron en: (1) identificar los procesos, y (2) determinar la secuencia e interacciones entre ellos.

Para el diseño del mapa se contrató un asesor externo que estudió la documentación existente y mantuvo entrevistas y reuniones con la dirección y todo el personal trabajador del servicio web para, en una primera fase, obtener una visión global de las tareas realizadas en el servicio y conseguir la información necesaria para identificar los procesos y sus responsables. Posteriormente, la segunda fase consistió en definir, estructurar y describir las tareas, actividades y registros utilizados en cada proceso, secuenciarlas en diagramas de flujo y plasmar todo en un procedimiento. La redacción inicial de cada procedimiento dio paso a una fase iterativa de revisión y depuración con el responsable principal de cada proceso y todo el personal implicado en cada procedimiento hasta llegar a su redacción definitiva.

Una vez identificados los procesos y desarrollados los procedimientos, el siguiente paso consistió en determinar la secuencia e interacciones entre los procesos representándolos en un gráfico (mapa). Se clasificaron, en función de sus actividades y funciones, en tres tipos: estratégicos, clave y de apoyo.

Estratégicos: En ellos se analizan las necesidades y condicionantes de la organización, y constituyen las guías adecuadas al resto de procesos para asegurar una respuesta a dichas necesidades y condicionantes. Definen la misión, la estrategia, la política y los objetivos del sistema.

Clave: Contienen las actividades esenciales que conducen al objetivo final de la organización. Son los que distinguen y caracterizan a la organización y son necesarios para realizar el producto, en este caso, información accesible en la web.

De apoyo: Son los responsables de proveer todos los recursos necesarios en cuanto a personas, equipos, materiales y materia prima, para a partir de los mismos, poder generar el valor añadido deseado por los clientes. Son los que dan soporte a los procesos clave para que éstos cumplan adecuadamente con su misión.

En el diseño del mapa se aplicaron criterios presentes en libros sobre sistemas de gestión de calidad y aplicación de las normas ISO (López-Rey, 2006; Cervera-Martínez, 2001; Pérez-Fernández-de-Velasco, 2004) y publicados también en algún otro estudio (Aboal-Viñas *et al.*, 2008):

- a) la representación del mapa comenzaba y terminaba en los ciudadanos y usuarios del servicio;
- b) los símbolos utilizados a un macronivel contenían los procesos e indicaban los flujos y relaciones, y
- c) todo el personal de la organización participó en la concreción de las tareas.

Resultado: mapa de procesos

La figura 1 muestra, a nivel macro, los procesos, su secuencia e interacción, y proporciona una visión global de la actividad realizada en la *Unidad Web* para cumplir su misión: proporcionar información accesible a la ciudadanía. Se describen a continuación, brevemente, los procesos:

a) Procesos estratégicos

Aparecen en la parte superior del gráfico e influyen en todos los demás.

Normativa: estándares y pautas (N:EyPau): El proceso define las normas técnicas, pautas y estándares del W3C recomendadas para obtener un determinado grado de accesibilidad. El seguimiento y cumplimiento de estas normas y recomendaciones están presentes en todos los procesos clave, desde la planificación y diseño hasta la verificación y validación de que la información en la web sigue las pautas y es accesible para todos los ciudadanos. Condiciona también la política de accesibilidad ya que el objetivo es alcanzar como mínimo el nivel doble A (AA).

Legislación (Leg): El proceso define la legislación en materia de accesibilidad, tanto europea como española, que obliga a que los servicios web de la administración pública sean accesibles para cualquier ciudadano. Su obligado cumplimiento afecta a todos los procesos clave y a la política de accesibilidad, ya que el compromiso es cumplir la legislación vigente.

Política de accesibilidad (PAcc): Basado en la legislación vigente y en las recomendaciones y pautas de accesibilidad del W3C, este proceso define la política de accesibilidad del *Ayuntamiento de Zaragoza* para su sitio web. La política, y los objetivos estratégicos, sirven de guía a todos los procesos clave.

Política de seguridad y protección de datos (PSPD): Basado en la legislación vigente, este proceso define la política de seguridad y protección de datos de *Ayuntamiento de Zaragoza* para su sitio web.

Política de datos abiertos (PDA): Desarrollada en cumplimiento de la *Ley 37/2007*, de reutilización de la información del sector público, por la que se insta a las administraciones a poner sus datos de carácter público a disposición de cualquier parte interesada. La política y los objetivos estratégicos sirven de guía a todos los procesos clave.

Gestión presupuestaria (GP): Establece el presupuesto con que cuenta la *Unidad de Gestión Web Municipal*, y su importancia radica en que afecta a todos los procesos.

b) Procesos clave

Están representados en la parte central del mapa. Tanto el punto de inicio como el final es la ciudadanía y los propios servicios del *Ayuntamiento*, en el primer caso representada por las necesidades de información y en el segundo caso por la información propiamente dicha que satisface las necesidades y las reclamaciones o sugerencias realizadas.

Diseño y desarrollo de portales y servicios web (DDPSW): Es el primer proceso y su principal objetivo es, a partir de las necesidades de los usuarios internos del *Ayuntamiento* y de los propios ciudadanos que demandan la creación de

nuevos portales o servicios web, planificar y diseñar la estructura.

En la planificación del portal se determina si es necesario desarrollar una aplicación para gestionar contenidos, por lo que sirve de entrada para el proceso «Desarrollo de sistemas de gestión de contenidos» (DSGCont). El resultado o salida es un nuevo portal o servicio web cuyos contenidos se deberán actualizar y mantener, por lo que sirve de entrada para el proceso «Edición/actualización de contenidos» (E/ACont).

Interactúa con los procesos clave «Evaluación de la accesibilidad» (EAcc), ya que es necesario evaluar la accesibilidad de los contenidos y del nuevo portal, y «Mantenimiento del sitio web» (MSW), por los posibles cambios o modificaciones derivadas de acciones correctivas, preventivas o de mejora que puedan hacerse en los portales. También se relaciona con el proceso «Gestión de reclamaciones y sugerencias» (GRSug), ya que pueden servir de entrada para crear, mejorar o modificar los portales y servicios web.

Este proceso se relaciona con todos los procesos estratégicos, ya que el desarrollo se realiza teniendo presente las normas y pautas de accesibilidad, la legislación vigente y la política de accesibilidad; el presupuesto va a permitir disponer de los recursos necesarios para realizar esta actividad (tanto personas como productos), por lo que también interactúa con los procesos de apoyo «Gestión de proveedores» (GProv) y «Gestión de recursos humanos» (GRHum), así como con «Proveedores de contenidos» (PCont) ya que proporcionan la materia prima de los portales, como es la información.

La documentación utilizada en este proceso, así como los registros que se crean, se gestionan según las indicaciones de los procedimientos «Gestión de los documentos» (GDoc) y «Gestión de los registros» (GReg).

Desarrollo de sistemas de gestión de contenidos (DSGCont):

En la fase de planificación de los nuevos portales se establece la necesidad de elaborar las herramientas y aplicaciones que faciliten la creación de los portales y la introducción de contenidos, así como su gestión y mantenimiento. La entrada del proceso es por lo tanto la fase de planificación de DDPSW, y el resultado es la herramienta de gestión que servirá inicialmente para crear el portal e introducir la primera información.

Se relaciona con los procesos clave EAcc, ya que hay que validar las salidas de la herramienta, y MSW, en el caso de que se produzcan cambios o modificaciones de las aplicaciones existentes.

Interactúa también con los procesos estratégicos y todos los de apoyo, por las mismas razones comentadas en el caso anterior.

Edición/actualización de contenidos (E/ACont): Su objetivo es introducir, actualizar y modificar la información en los portales y servicios existentes en la web, bien mediante los sistemas de gestión de contenidos desarrollados en el proceso anterior o con otros editores de páginas web.

Se relaciona con el proceso EAcc, ya que la información introducida se valida y verifica que cumple con los requisitos mínimos de accesibilidad establecidos, y con MSW, ya que puede dar lugar a acciones correctivas sobre las causas de las no conformidades detectadas.

Se relaciona asimismo con los procesos estratégicos y los de apoyo, especialmente con PCont, al proporcionar la información que será publicada en la web, y con GRHum, ya que la formación, experiencia y competencia de los gestores de contenidos es fundamental para minimizar los errores y no conformidades con la accesibilidad.

Evaluación de la accesibilidad (EAcc): Tiene como objetivo verificar y validar el cumplimiento de los estándares web y pautas de accesibilidad tanto en el diseño y desarrollo de portales y servicios web, como en el desarrollo de SGCont y E/ACont, por lo que es transversal a estos tres procesos clave.

Se interrelaciona con todos los procesos estratégicos, sobre todo con N:EyP y PAcc, que definen el nivel mínimo de accesibilidad que deben cumplir los contenidos.

También se interrelaciona con los procesos de apoyo, muy especialmente con el de GProv y PCont, ya que en ambos casos es necesario que cumplan con los requisitos de accesibilidad que les indica la *Unidad de Gestión Web Municipal*, tanto en lo que a productos (software) se refiere como a información.

Mantenimiento del sitio web (MSW): Proceso transversal que afecta a los procesos clave DDPSW, DSGCont, y E/ACont, ya que su principal objetivo es controlar los cambios necesarios tanto para corregir no conformidades y errores como para adaptarse a cambios tecnológicos y funcionales y posibles mejoras que afecten al diseño y estructura de los portales, al desarrollo de herramientas de gestión de contenidos y a los propios contenidos.

Se interrelaciona con los procesos de apoyo, especialmente con GProv, GDoc y GRHum, ya que los cambios pueden exigir compras de nuevos productos o servicios, el cambio de documentos esenciales como la guía de estilo, y formación en las novedades implementadas.

Gestión de reclamaciones y sugerencias (GRSug): Este proceso tiene como objetivo atender y resolver las reclamaciones, quejas y sugerencias realizadas por los clientes y relacionadas con la accesibilidad del sitio web. El análisis de las reclamaciones sirve como revisión y evaluación del sitio web ya que permite identificar fallos en el servicio y llevar a cabo las acciones correctivas y/o preventivas pertinentes, estando por tanto relacionado con los procesos DDPSW, E/ACont, EAcc, y MSW.

Se interrelaciona con el proceso de apoyo GProv, ya que algunas reclamaciones o quejas se envían a los servicios que suministran la información.

c) Procesos de apoyo

Se sitúan en la parte inferior del gráfico y se encargan de la gestión y de aquellas tareas cuya responsabilidad recae en otros servicios.

Gestión de documentos (GDoc): Su objetivo es la gestión de los documentos internos y externos creados y utilizados en el SAG. Se interrelaciona con todos los procesos clave, ya que en todos ellos se generan documentos, así como con los procesos estratégicos y con el resto de los de apoyo, especialmente con GReg.

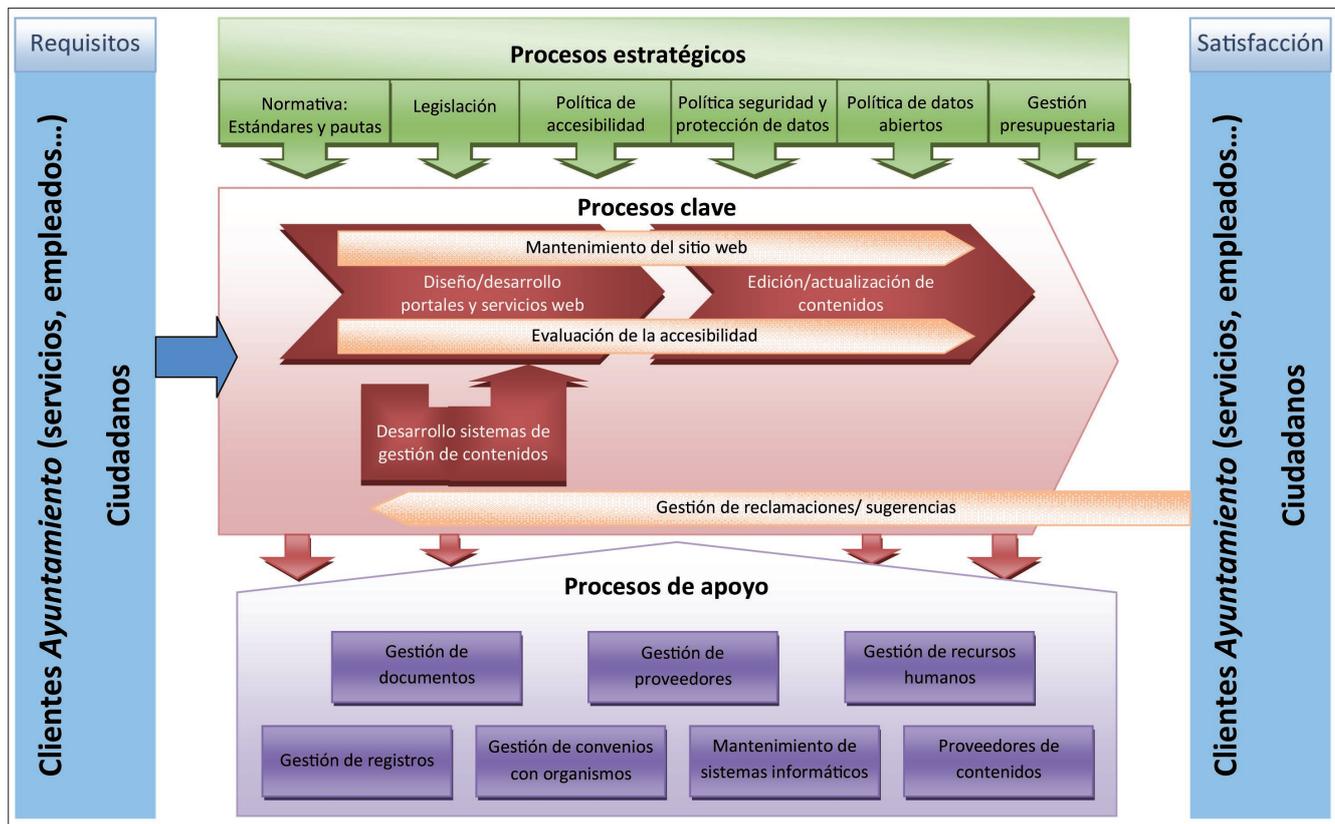


Figura 1. Mapa de procesos del sistema de gestión de la accesibilidad del servicio web del Ayuntamiento de Zaragoza

Gestión de registros (GReg): Establece el control de los registros generados en el SAG, por lo que está relacionado claramente con todos los procesos clave, ya que en todos se generan registros.

Gestión de recursos humanos (RHum): Tiene como objetivo identificar las necesidades y capacitación del personal necesario para realizar las tareas de diseño, desarrollo y mantenimiento del sitio web, así como su formación en accesibilidad. Sirve de apoyo a todos los procesos clave, ya que de la capacitación y formación del personal dependerá que el sistema de gestión sea eficaz.

Gestión de proveedores (GProv): Tiene como objetivo seleccionar, evaluar y hacer seguimiento a los proveedores de servicios y productos necesarios para llevar a cabo los procesos clave, es decir, todo lo que afecte al diseño, desarrollo y mantenimiento del sitio web. Sirve de apoyo a todos los procesos clave, ya que todos requieren de estos productos o servicios.

Proveedores de contenidos (PCont): Determina la relación con los servicios o áreas del Ayuntamiento de Zaragoza y con organismos externos que proporcionan o envían información a la Unidad de Gestión Web Municipal para ser publicada en la web o que publican ellos mismos.

Se interrelaciona directamente con DDPSW, DSGCont, E/ACont y EAcc, ya que los dos primeros desempeñan un papel fundamental al ser demandantes de un nuevo portal o servicio web, en el tercero por enviar contenidos o publicarlos en la web, y en el último porque validan aspectos como la visualización y navegación. También se relaciona con GR-Sug, ya que algunas de las sugerencias se derivan a los propios servicios.

Mantenimiento de sistemas informáticos (MSI): Es externo a la Unidad de Gestión Web Municipal, y determina el mantenimiento del hardware y software empleados para la creación y funcionamiento de la web del Ayuntamiento de Zaragoza, siendo su responsable el Servicio de Informática.

Gestión de convenios con organismos (GCOrg): Indica la relación con otros organismos externos (universidad, colegios profesionales, etc.) que proporcionan recursos humanos o de otro tipo que intervienen en los procesos clave.

Conclusiones

La identificación de los procesos y su descripción y documentación pormenorizada no sólo sirve para obtener un certificado de calidad o accesibilidad que garantice compromisos mínimos exigidos en el certificado. Lo más importante es que sirve para estructurar las actividades que se realizan en un servicio, identificar duplicidades en tareas y registros, detectar problemas adoptando medidas oportunas preventivas y correctoras. En definitiva, mejora la organización, gestión y eficiencia de un servicio y todo ello redundará en una mayor calidad del producto, que en este caso es la información accesible, y por tanto, en un mejor servicio a la ciudadanía o cliente, que es el consumidor de esta información.

Nuestra experiencia ha sido satisfactoria. Ha resultado esencial el compromiso de todo el equipo de la web municipal y su participación e implicación en el mantenimiento y mejora del sistema. A pesar de que el primer año resultó laborioso y complejo, la actualización posterior de los procedimientos siempre que se ha introducido alguna modificación o añadido alguna novedad, no ha supuesto un gran esfuerzo. Hay

que resaltar el aspecto pragmático de un sistema de gestión de calidad con unos procesos bien definidos, que supone que el personal nuevo disponga por escrito de una documentación esencial y útil donde se describe el trabajo y las instrucciones sobre cómo llevarlo a cabo, algo que facilita su formación inicial en el uso del sistema.

Desde el punto de vista de la gestión se han visto mejorados todos los aspectos relacionados con el registro de las actividades, normalizándose los formatos y disponiendo del archivo de todos los registros. Se ha favorecido el control y supervisión de todas las etapas en cada proceso, lo que ha repercutido en una mayor facilidad para llevar a cabo la evaluación del servicio prestado, identificación de errores y aplicación de medidas correctoras y preventivas. Un sistema de documentación eficiente y de registro de la información es uno de los elementos más importantes para asegurar el éxito de un sistema de gestión de calidad o de accesibilidad y también es uno de los que al personal de una organización le cuesta más adaptarse, pero es esencial ya que las auditorías para obtener el certificado se basan en la información recogida en los procedimientos y en los registros.

Desde el punto de vista del cliente se ha mejorado la calidad del servicio y del producto al ofrecer información más accesible tanto para personas con algún tipo de discapacidad como para aquellas conectadas desde cualquier dispositivo.

También se facilita la reutilización de la información por parte de la ciudadanía, las empresas y otros organismos, lo que ofrece un aumento de la transparencia de la administración, el incremento de la participación ciudadana y la posibilidad de crecimiento económico en distintos sectores.

En resumen, la implantación de un sistema de gestión de accesibilidad (que es un sistema de gestión de calidad específico) no sólo supone una mejora puntual en la organización y sistematización de las actividades llevadas a cabo en una organización, sino que es también algo dinámico donde se busca la mejora continua del sistema mediante tres elementos clave: los miembros de la organización con su participación activa proponiendo nuevas ideas y mejoras, los usuarios a través del servicio de quejas y reclamaciones, y la dirección con la revisión y propuesta de acciones correctivas y preventivas.

Bibliografía

Aboal-Viñas, José-Luis; Lado-Lema, María-Eugenia; Amigo-Quintana, Manuel; Hervada-Vidal, Xurso; Gómez-Amorín, Ángel; Fernández-Abreu, Carlos. “Mapa de procesos en organizaciones de salud pública: la experiencia de la Dirección Xeral de Saúde Pública de Galicia”. *Gaceta sanitaria*, 2008, v. 22, n. 3, pp. 275-279.
<http://scielo.isciii.es/pdf/ga/v22n3/especial2.pdf>

Aenor. *Reglamento particular de la marca Aenor N para accesibilidad TIC para sitios web. RP A90.01*. 2006, 21 pp.
http://www.aenor.es/documentos/certificacion/reglamentos/w_RP_A90_01_2006-10-04.pdf

Ahire, Sanjay; Dreyfus, Paul. “The impact of design management and process management on quality: an empirical investigation”. *Journal of operations management*, 2000, v. 18, pp. 549-575.

[http://dx.doi.org/10.1016/S0272-6963\(00\)00029-2](http://dx.doi.org/10.1016/S0272-6963(00)00029-2)

Anjard, Ronald P. “Process mapping: one of the three, new, special quality tools for management, quality and all other professionals”. *Microelectronics and reliability*, 1996, v. 36, n. 2, pp. 223-225.

Cervera-Martínez, Josep. *La transición a las nuevas ISO 9000:2000 y su implantación*. Madrid: Díaz de Santos, 2001. ISBN: 978 84 797 8519 2

Davenport, Thomas. *Process innovation*. Boston: Harvard Business School Press, 1993. ISBN: 978 08 758 4366 7

Duvauferrier, Régis; Badonnel, Stéphanie; Vannoorenbergh, Yves; Philippe, Christian; Hamon, Jean-Louis; Rambeau, Marc; Korvin, Brigitte; Rolland, Y. “Management en processus d’un département d’imagerie médicale. Impact sur le système qualité”. *Journal de radiologie*, 2002, v. 83, n. 1, pp. 13-25.

http://www.em-consulte.com/showarticlefile/121142/pdf_36106.pdf

<http://dx.doi.org/JR-01-2002-83-1-0221-0363-101019-ART5>

Gibb, Forbes; Buchanan, Steven; Shah, Sameer. “An integrated approach to process and service management”. *Intl journal of information management*, 2006, v. 26, pp. 44-58.
<http://dx.doi.org/10.1016/j.ijinfomgt.2005.10.007>

Heinrich, Bernd; Henneberger, Matthias; Leist, Susanne; Zellner, Gregor. “The process map as an instrument to standardize processes: design and application at a financial service provider”. *Information systems and e-business management*, 2009, v. 7, pp. 81-102.

http://www.uni-augsburg.de/exzellenz/kompetenz/kernkompetenzzentrum_fim/Forschung/paper/paper/wi-209.pdf

<http://dx.doi.org/10.1007/s10257-007-0064->

ISO. *ISO 9000 - Quality management principles*.

http://www.iso.org/iso/iso_catalogue/management_and_leadership_standards/quality_management/qmp.htm

López-Rey, Susana. *Implantación de un sistema de calidad. Los diferentes sistemas de calidad existentes en la organización*. 1ª ed. Vigo: Ideaspropias, 2006. ISBN: 978 84 965 7825 8

Pérez-Fernández-de-Velasco, José-Antonio. *Gestión por procesos: cómo utilizar ISO 9001:2000 para mejorar la gestión de la organización*. Madrid: ESIC, 2004. ISBN: 978 84 735 6389 5

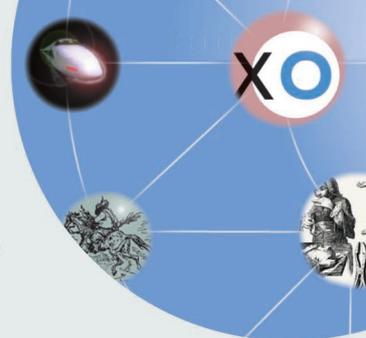
Rey-Peteiro, Domingo. *Gestión por procesos y modelado de procesos*, 2005.
<http://www.gestiopolis.com/recursos4/docs/ger/gestitra.htm>

Ruiz-Canela-López, José. *La gestión por la calidad total en la empresa moderna*. Madrid: Rama, 2003. ISBN: 978 84 789 7592 1

Villanueva, Isabel; Sánchez, Juan; Pastor, Óscar. “Elicitación de requisitos en sistemas de gestión orientados a procesos”. En: *Procs of the 11th Intl workshop on requirements engineering (WER'05)*, pp. 38-49, Porto, Portugal, 2005.

http://wer.inf.puc-rio.br/WERpapers/artigos/artigos_WER05/isabel_villanueva.pdf

doc6 en 30 segundos



Nuestra misión

Aportar soluciones para mejorar la eficiencia de empresas y organizaciones a través de una eficaz gestión de la información y del conocimiento.

La empresa

Creada en 1988. Le aportamos **más de 20 años de experiencia.**

- Consultoría en gestión documental, de la información y del conocimiento
- Sistemas de Gestión Documental: mapas documentales, cuadros de clasificación de archivos, concreción de funcionales, etc.
- Intranets, Extranets. Arquitectura y usabilidad web. Taxonomías
- Servicios profesionales adecuados a cada necesidad
- Soluciones en gestión documental
- Recursos electrónicos de información. Representante en España de OCLC
- Formación personalizada

Los 6 valores de doc6

Los valores por los que se rige la actividad de doc6 definen nuestro código ético, nuestros métodos de trabajo y organización, la cultura de la empresa. Son los siguientes:

- Excelencia en calidad
- Innovación permanente. Creatividad
- Cooperación y trabajo en equipo
- Confidencialidad
- Profesionalidad y honestidad con clientes y proveedores
- Responsabilidad social

Los clientes

El principal activo de la empresa, junto con sus empleados.

Más de 1000 clientes: empresas, laboratorios, administraciones públicas, universidades, hospitales, bufetes de abogados, museos, fundaciones, colegios profesionales.

Nuestros profesionales

Un equipo de **más de 100 profesionales.**

Una eficaz suma de conocimientos: informáticos y documentalistas con experiencia.

El 90% son licenciados superiores o diplomados en informática o documentación.



doc 6 S.A. dispone de la etiqueta **EUREKA de I+D** para el proyecto COGNOS (n. 2734) de metodología para el desarrollo de mapas del conocimiento en las organizaciones.



doc 6 ha obtenido la **Certificación ISO 9001** otorgada por Bureau Veritas Quality International (BVQI) con las acreditaciones ENAC (España) y UKAS (Reino Unido) para sus Centros de Barcelona y

doc 6
consultores en recursos de información

Mallorca, 272, planta 3 - 08037 Barcelona • Tel. 932 154 313 Fax 934 883 621
Orense, 14, 5º A - 28020 Madrid • Tel 915 535 207 Fax 915 346 112

mail@doc6.es - www.doc6.es



SOFTWARE DOCUMENTAL



TEMATRES: SOFTWARE PARA GESTIONAR TESAUROS



Audilio Gonzales-Aguilar, María Ramírez-Posada y Diego Ferreyra



Audilio Gonzales-Aguilar es doctor en derecho y nuevas tecnologías, con calificación magna cum laude, por la *Universidad de Montpellier I*. Es profesor titular de la *Universidad Paul Valéry Montpellier 3*, en Montpellier, Francia, donde es miembro del equipo de investigación *Lerass-Ceric*. Director del master profesional en gestión de la información y de la documentación. Sus áreas de investigación son: análisis de redes sociales aplicadas a contenidos documentarios, hipertextualidad del documento digital y visualización y representación de la información.

Universidad Paul Valéry Montpellier 3
Departamento de Documentación
Route de Mende, 34199 Montpellier Cedex 5
audilio.gonzales@univ-montp3.fr
<http://www.univ-montp3.fr>



María Ramírez-Posada es bióloga marina, candidata a maestría en desarrollo sostenible y medio ambiente (*Universidad de Manizales*, Colombia), y candidata a maestría en gestión de la información y del conocimiento (*Universidad Paul Valéry Montpellier 3*, Francia). Experiencia como investigadora en el área del desarrollo sostenible empresarial y visualización de la sustentabilidad. Experiencia como docente a nivel pre gradual y pos gradual.

Webdocumenta
Grupo de Investigación Eco-ambiental
maria.ramirezposada@webdocumenta.com
<http://201.234.78.173:8080/gruplac/jsp/visualiza/visualizagr.jsp?nro=0000000005056>



Diego Ferreyra es bibliotecario por la *Universidad Nacional de Misiones*, licenciado en bibliotecología y documentación por la *Universidad del Museo Social Argentino* y especialista en teoría del diseño comunicacional por la *Universidad de Buenos Aires*. Es docente adjunto regular a cargo de la cátedra de *Redes y Sistemas de Información* y docente adjunto a cargo de la cátedra de *Recursos de Información en Redes Globales* de la *Facultad de Filosofía y Letras* de la *Universidad de Buenos Aires*. Mantiene los sitios *R020 Bibliotecología* y *Ciencias de la Información* y el servidor de vocabularios *TemaTres*.

R020 Bibliotecología y Ciencias de la Información
<http://www.r020.com.ar>
<http://www.vocabularyserver.com>
diego@r020.com.ar

Resumen

Los tesauros, como lenguaje documental de términos normalizados, han tenido un desarrollo importante para la gestión documental con el uso de aplicaciones de software como el *TemaTres*, que es un servidor de vocabulario controlado. Esta aplicación libre para uso en la web facilita el acceso continuo a los conjuntos documentales de temas específicos para estudio, investigación y toma de decisiones. Se presenta el proceso de instalación, funcionalidades y la interface que muestra la estructura de la aplicación y las vías de explotación de *TemaTres* para la representación visual y la gestión de vocabularios controlados.

Palabras clave

Lenguaje documental, Tesauros, Términos normalizados, Términos controlados, Software, Aplicaciones, Programas, Gestión documental.

Title: *TemaTres*: software to implement thesauri

Abstract

The documentary language thesaurus, as standard terms, have been an important development for document management using software applications such as *TemaTres*, which is a controlled vocabulary server. This free application, for use on

Artículo recibido el 26-04-12
Aceptación definitiva: 06-05-12

the web, provides continuous access to the sets of documents on specific topics for study, research and decision making. It presents the installation, functionality and interface that shows the application structure and ways of exploiting *TemaTres* for visual representation and management of controlled vocabularies.

Keywords

Documentary language, Thesaurus, Software, Applications, Standard terms, Controlled terms, Document management.

Gonzales-Aguilar Audilio; Ramírez-Posada, María; Ferreyra, Diego. *TemaTres: software para gestionar tesauros. El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 319-325.

<http://dx.doi.org/10.3145/epi.2012.may.14>

1. Introducción

Para organizar, gestionar y recuperar la información se usan herramientas tales como tesauros, clasificaciones, ontologías, glosarios y diccionarios que abarcan áreas del conocimiento más o menos específicas. Se trata de lenguajes documentales, sistemas semánticos artificiales que mediante el uso de términos normalizados o controlados permiten describir el contenido de los documentos y evitar posibles ambigüedades.

En este artículo se presenta el software *TemaTres*, que es una aplicación web para gestionar representaciones lingüísticas formales del conocimiento: vocabularios controlados, taxonomías, tesauros, listas de encabezados, etc.

Los tesauros son estructuras conceptuales formadas por unos términos llamados descriptores¹ entre los cuales se establecen relaciones semánticas. Igualmente se indican términos equivalentes² “no descriptores”³, cuyo uso no se recomienda.

El uso de *TemaTres* garantiza la coherencia y la normalización de la información en un sistema distribuido, eliminando la ambigüedad, la polisemia y la sinonimia del vocabulario especializado.

2. Funciones

Ver figura 1.

3. Instalación

TemaTres es un software libre disponible para los sistemas operativos *Windows*, *Mac OS* y *Linux*. Su instalación (en internet o en el localhost) se hace de la siguiente manera:

3.1. Instalación del localhost

<http://rhr058.com/crear-servidor-local>

Para *Windows* se puede instalar WampServer 2.0⁴, y para *Mac OS* el MAMP⁵. Estos programas se instalan automáticamente. Es necesario que el servidor local (localhost) esté

TemaTres

Es un servidor de vocabularios controlados libre, una aplicación web para gestionar y explotar vocabularios controlados, tesauros, taxonomías y otros modelos de representación formal del conocimiento. Requiere PHP, MySQL y un servidor Web HTTP.

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

abierto, y para saberlo se escribe en el navegador “localhost” o “http://127.0.0.1”. Todas las aplicaciones necesarias para la instalación de *TemaTres* están en la gama de software libre.

3.2. Descargar *TemaTres* del sitio oficial

La última versión es la 1.5:

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

3.3. Descomprimir *TemaTres* para crear la carpeta “Tematres 1.5.1”

Se debe renombrar esta carpeta como “tematres”, luego se debe copiar o mover al repertorio de *Windows C:/wamp/www/...* o, en *Mac OS* en la carpeta Aplicaciones → MAMP → httdocs →....

3.4. Lanzar el servidor (ver figura 2) a partir del navegador escribiendo “localhost” o <http://127.0.0.1>

3.5. Crear la base de datos *TemaTres*

En la página de inicio de MAMP o Wampserver hacemos clic en el enlace a phpMyAdmin para crear la base de datos de “*TemaTres*” y aparecerá la pantalla que se muestra en la figura 3.

Escribir el nombre “tematres” y pulsar el botón “Crear”. El resto de opciones se dejan como están, sin modificar.

3.6. Abrir el fichero “db.tematres.php”

(con bloc de notas en PC o TextEdit en *Mac OS*) que se encuentra en:

C:/wamp/www/tematres/vocab/db.tematres.php

o en *Mac OS* en carpeta Aplicaciones → Mamp → httdocs → tematres → vocab → db.tematres.php)

Los datos para configurar y/o completar este archivo son los siguientes:

Sobre R020

Es un sitio independiente sin fines de lucro dedicado a la investigación y difusión en el campo de la bibliotecología y las ciencias de la información. R020 está orientado al espectro profesional que se dedica a elaborar servicios documentales y de información. Es un espacio multidisciplinar y abierto al intercambio con especialistas de todas las disciplinas relativas al campo de la información. Los servicios ofrecidos son desarrollados tecnológicamente por R020 y mantenidos colaborativamente por la comunidad de usuarios. Fue creado y está dirigido por **Diego Ferreyra**.

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

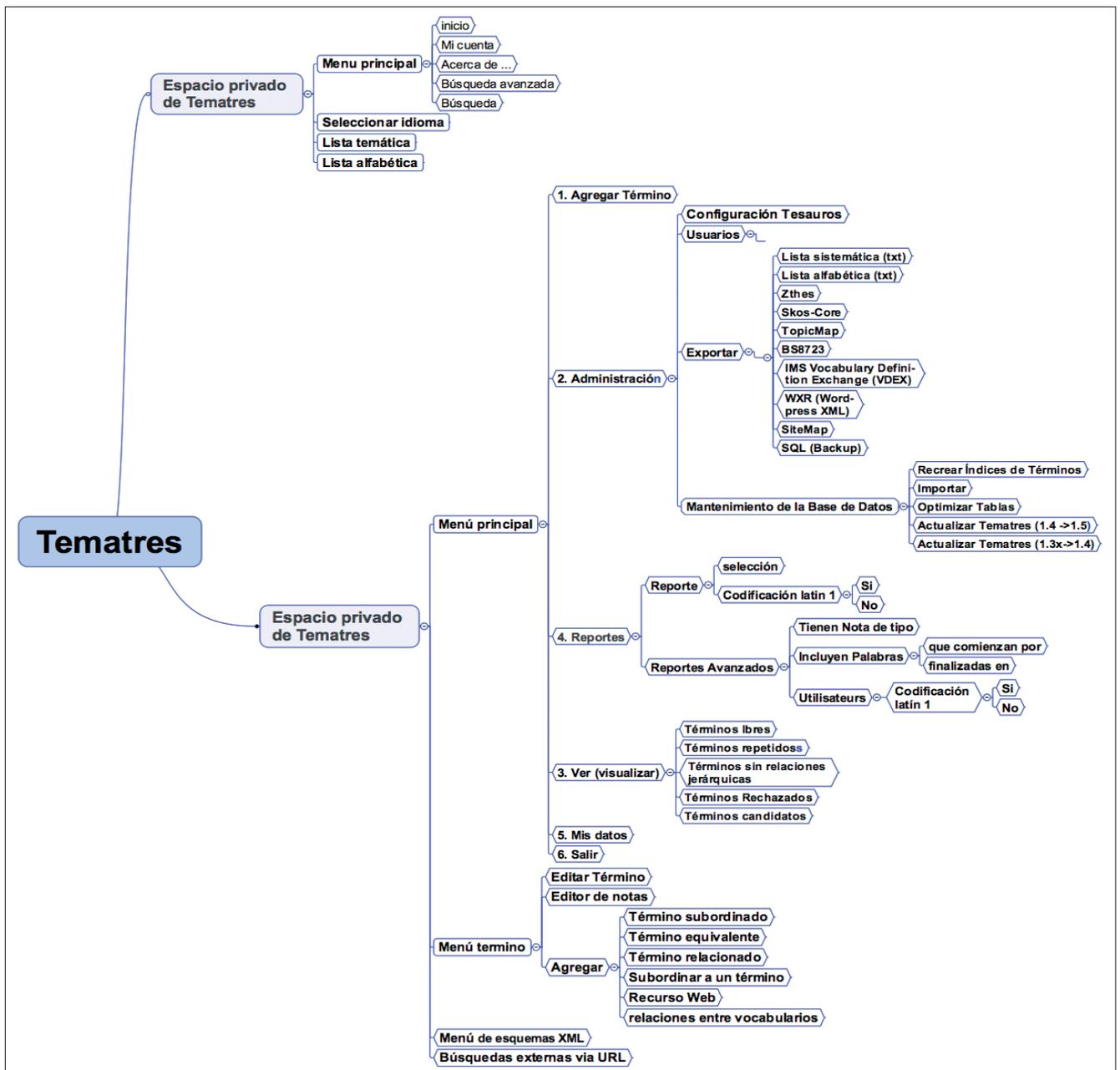


Figura 1. Esquema de las funciones de TemaTres

Selección del tipo de servidor de bases de datos a utilizar. Si se deja en blanco, utilizará MySQL.

```
$DBCFG["DBdriver"] = "";
```

Dirección IP o nombre del servidor de bases de datos, por ejemplo, localhost
 \$DBCFG["Server"] = "localhost"; (en MacOX "localhost:8889")

Nombre de la base de datos
 \$DBCFG["DBName"] = "tematres";

Nombre de usuario para conectarse con la base de datos
 \$DBCFG["DBLogin"] = "root";

Clave de usuario para conectarse con la base de datos
 \$DBCFG["DBPass"] = ""; en MacOX "root")

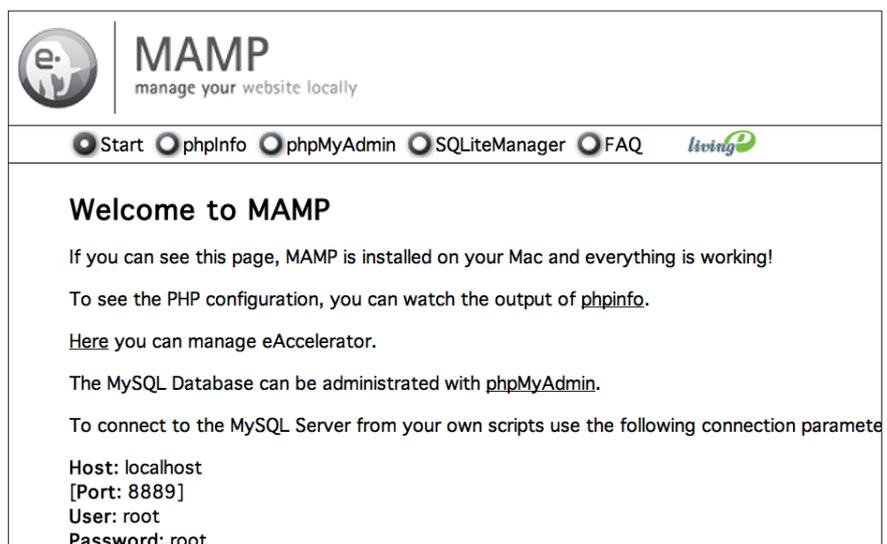


Figura 2. Servidor MAMP para instalación de TemaTres en Mac OX

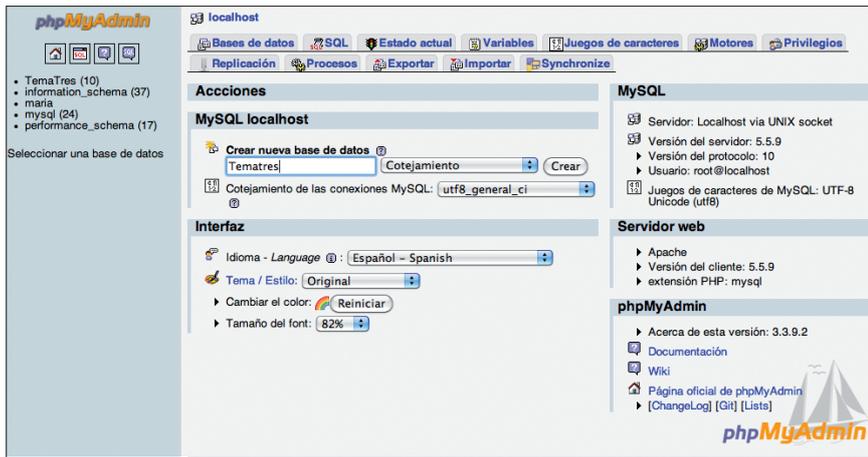


Figura 3. Creación de base de datos *TemaTres* en el servidor phpMyAdmin

Prefijo para las tablas de vocabulario controlado
`$DBCFG["DBprefix"] = "lc_";`

Una vez completada y verificada esta información se guarda el archivo db.tematres.php

3.7. Para instalar *TemaTres* en Windows

Abra el navegador y escriba "http://localhost/tematres" o http://127.0.0.1 tematres. En Mac OX escriba "http://localhost:888/tematres" y siga las instrucciones de instalación, que pueden consultarse en:

<http://r020.com.ar/tematres/wiki/doku.php?id=tematres:install>

Seleccione el idioma y complete los datos sobre el vocabulario y los datos del usuario (figura 4).

Confirme la instalación y acceda a *TemaTres* a partir del enlace "comience a utilizar su vocabulario" (figura 5).

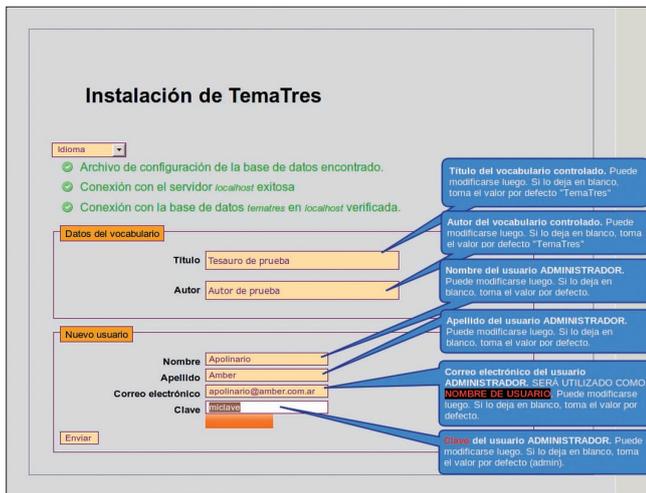


Figura 4. Instalación de *TemaTres*

A continuación se accede a la pantalla inicial de *TemaTres* (ver figura 6).

4. Interfaz

TemaTres posee dos tipos de interfaz: una para navegar, consultar y visualizar un tesaurus (interface pública) y otra para crear, gestionar, administrar y exportar un tesaurus (interface privada o de administración).

4.1. Interfaz pública de navegación

a) El menú principal permite acceder a:

Inicio: navegar como usuario en el sitio de *TemaTres*.

Mi cuenta: acceso a la creación y gestión de tesaurus para un usuario registrado.

Sobre...: datos generales del tesaurus.

Búsqueda avanzada: buscar en el tesaurus por notas, códigos, tipo de término, etc.

TemaTres es un programa web para gestionar representaciones lingüísticas del conocimiento: vocabularios controlados, taxonomías, tesaurus, listas de encabezados...

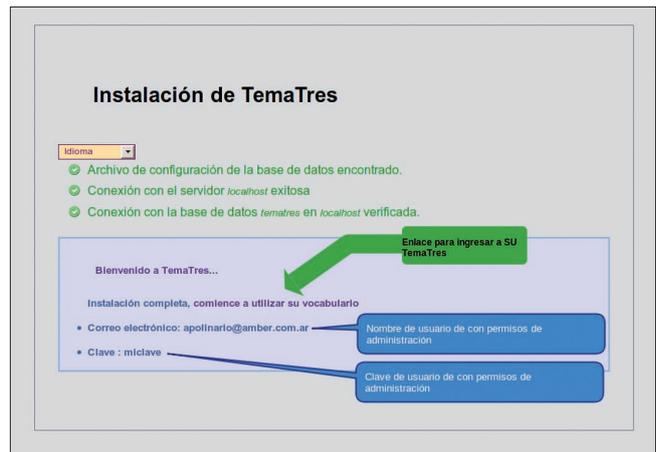


Figura 5. Pantalla de acceso a *TemaTres*

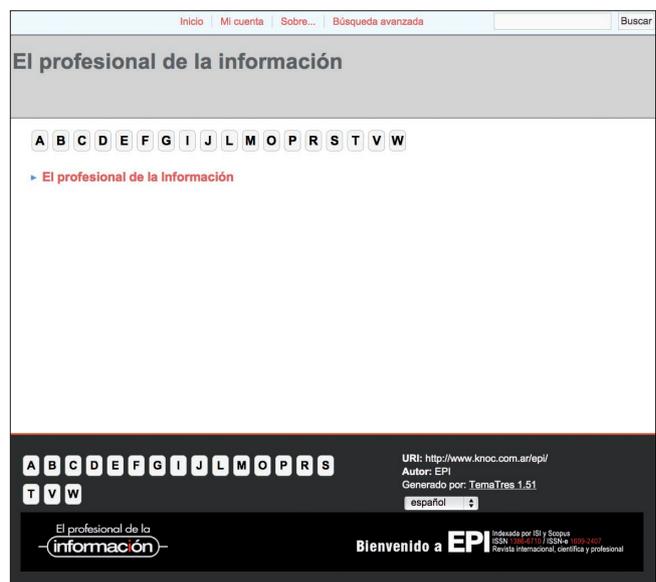


Figura 6. Pantalla inicial de *TemaTres*

http://r020.com.ar/tematres/manual/cmo_agregar_o_crear_un_trmino.html

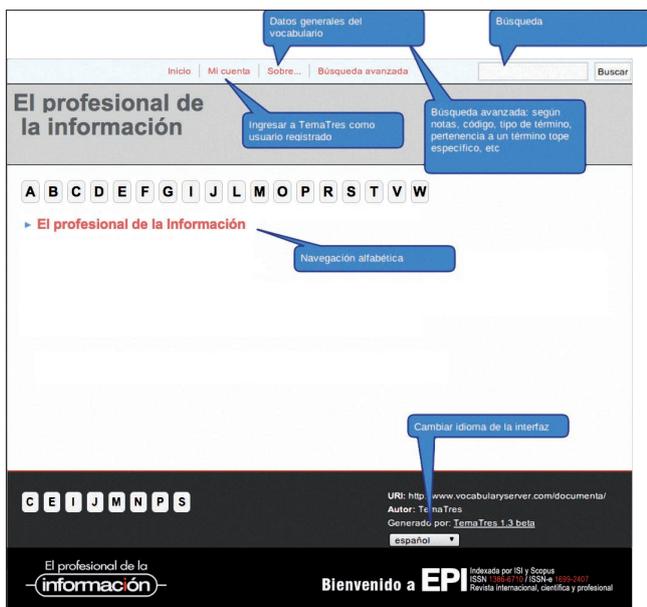


Figura 7. Interfaz pública de navegación de TemaTres

- b) Navegación alfabética por los términos del tesoro.
- c) Navegación sistemática por la estructura jerárquica del tesoro.
- d) Selección de idioma de la interfaz de TemaTres.

4.2. Interfaz privada de creación de un tesoro

Es accesible a partir de la opción del menú “Mi cuenta” para la creación de un tesoro, para agregar términos, administrar los usuarios, los vocabularios controlados y la exportación de tesoros.

- a) Mi cuenta (acceso al modo creación de un tesoro).

Esta opción permite al usuario ser administrador del tesoro. Se deben incluir los datos que muestra la figura 8.

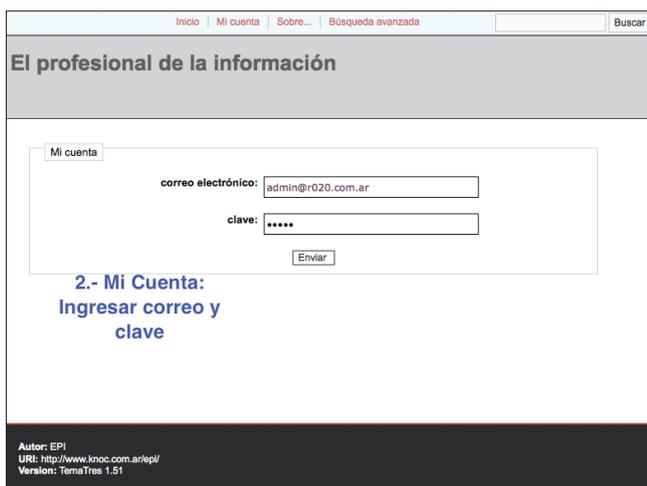


Figura 8. Acceso al modo de creación de un tesoro

De ésta se puede acceder al modo edición de tesoro, que presenta estas opciones:

- 1) Menú principal

Permite: Agregar términos, Administración, Ver, Reportes, Mis datos y Salir.

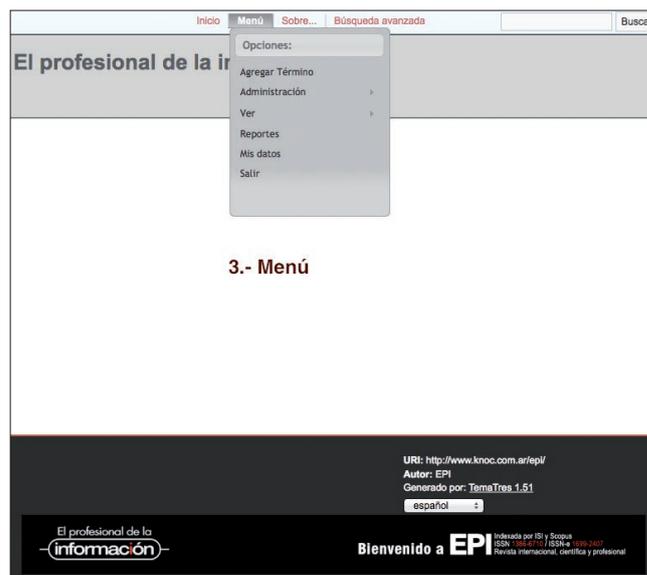


Figura 9. Menú principal de TemaTres

- 2) Menú de términos

Se pueden editar los términos, crear notas (de alcance, históricas, bibliográficas, de catalogación y nota privada). <http://r020.com.ar/tematres/manual/notas.html>



Figura 10. Menú de términos de TemaTres

- 3) Menú de esquemas XML

En estos links se editan los términos en los siguientes esquemas: BS8723-5, DC (Dublin core) MADS, SKOS-Core, VDEX (vocabulary definition exchange), XTM (topics maps) y Zthes:

<http://r020.com.ar/tematres/manual/administracin.html>

- e) Búsquedas externas vía URL

Este grupo de links permite buscar el término específico en Google books, Google image, Google búsqueda exacta y Wikipedia. También se pueden editar los términos en los



Figura 11. Menú de esquemas XML

esquemas BS8723-5, DC (*Dublin core*), MADS, SKOS-Core, VDEX (*vocabulary definition exchange*), XTM (*topics maps*) y Zthes

<http://r020.com.ar/tematres/manual/administracin.html>

5. Ejemplo de explotación de TemaTres

Visual vocabulary

Permite una representación visual del vocabulario controlado. Las representaciones gráficas de conceptos se pueden encontrar en:

<http://vocabularyserver.com/visualvocabulary>

<http://vocab.lternet.edu/visualvocabulary/lter>

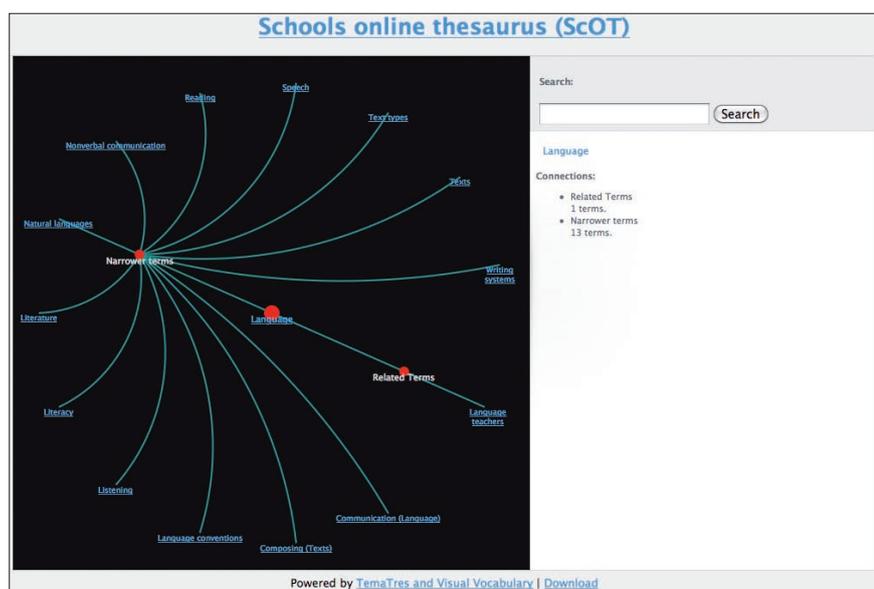


Figura 12. Representación visual del vocabulario controlado

6. Conclusión

Las ventajas para la gestión en TemaTres se clasifican en:

- Ayudas y facilidades de uso y gestión, orientadas a facilitar el proceso de gestión. Ej.: herramientas multiusuario, control activo de consistencia e integridad, comunicación interna, etc.
- Mecanismos activos de control de consistencia e integridad, que operan impidiendo o previniendo errores lógicos (relaciones no admitidas), concurrencias, condiciones lógicas en el paso de estado de un término, etc.
- Herramientas de auditoría y control de calidad, que permiten detectar problemas o aspectos que podrían minar

la productividad del vocabulario, y elaborar índices de calidad e informes de seguimiento sobre el ciclo de vida del vocabulario (vocabularios asociados, términos, editores, etc.).

TemaTres facilita la exportación a otros sistemas.

TemaTres tiene una capacidad de almacenamiento de términos ilimitada.

La instalación del programa aún conlleva una cierta complejidad.

7. Notas

1. Son palabras o expresiones del lenguaje corriente retenidas por el constructor del tesoro para designar los conceptos representativos del documento y las preguntas utilizadas por el indizador.
2. Cuando dos o más términos representan el mismo o casi el mismo concepto, como por ejemplo: sinónimos, nombres comunes de seres vivos y sus equivalentes nombres científicos, variantes ortográficas, acrónimos... Para un índice exacto, claro y consistente, se usa solamente un término para representar un concepto. Controlan la sinonimia, homonimia, antonimia y polisemia.

3. Son sinónimos o casi-sinónimos de los descriptores o términos que designan en el lenguaje de uso conceptos afines a los que cubren los descriptores. Los no descriptores no pueden ser utilizados para la indización del documento o de las preguntas, sino que cada uno de ellos reenvía a uno o dos descriptores para representar los conceptos correspondientes.

4. Descargar desde: <http://www.wampserver.com/en> (para Windows). Para su instalación remitimos a la página oficial: <http://pixelar.me/wampserver-guia-de-descarga-e-instalacion>

5. Descargar desde: <http://www.mamp.info/en/mamp.html> (para Mac OS).

Se puede ver un tutorial en: <http://ayudawordpress.com/como-instalar-wordpress-en-mac>

8. Bibliografía

- Aenor. "Norma UNE-50-106-90". Directrices para el establecimiento y desarrollo de tesauros monolingües: equivalente a ISO 2788-1986. Aenor, 1990.
- Codina, Lluís; Pedraza-Jiménez, Rafael. "Tesauros y ontologías en sistemas de información documental". *El profesional de la información*, 2011, sept.-oct., v. 20, n. 5, pp. 555-563. <http://dx.doi.org/10.3145/epi.2011.sep.10>

García-Marco, Francisco-Javier. "Ontologías y organización del conocimiento: retos y oportunidades para el profesional de la información". *El profesional de la información*, 2007, nov.-dic., v. 16, n. 6, pp. 541-550.
<http://dx.doi.org/10.3145/epi.2007.nov.01>

Martínez, Ana M.; Ristuccia, Cristina A.; Stubbs, Edgardo A.; Valdez, Julia C.; Gamba, Viviana L.; Mendes, Paola V.; Unzurrunzaga, Carolina; Caminotti, María L. "La estructura sistemática del tesoro: Indicadores para evaluar su calidad". En: *Revista española de documentación científica*, 2011, enero-marzo, v. 34, n. 1, pp. 29-43.
<http://dx.doi.org/10.3989/redc.2011.1.765>

Mochón-Bezares, Gonzalo; Sorli-Rojo, Ángela. "Tesauros en acceso abierto en Internet. Un análisis cuantitativo". *Revista española de documentación científica*, 2010, oct.-dic., v. 33, n. 4, pp. 643-663.
<http://dx.doi.org/10.3989/redc.2010.4.763>

Moreiro-González, José-Antonio; Azcárate-Aguilar-Amat, Pilar; Marzal-García-Quismondo, Miguel-Ángel; Tejada-Artigas, Carlos-Miguel; Vergueiro, Waldomiro C. S. "Desarrollo profesional y opinión sobre la formación recibida de los titulados universitarios en información y documentación de las universidades públicas de Madrid (2000-2005)". *El profesional de la información*, 2008, mayo-junio, v. 17, n. 3, pp. 261-272.
<http://dx.doi.org/10.3145/epi.2008.may.02>

Ontalba-Ruipérez, José-Antonio. "Análisis de contenido de la lista de correo electrónico *lweTel* (2001-2007)". *El profesional de la información*, 2009, enero-febrero, v. 18, n. 1, pp. 50-61.
<http://dx.doi.org/10.3145/epi.2009.ene.07>

Sánchez-Cuadrado, Sonia; Morato-Lara, Jorge; Palacios-Madrid, Vicente; Llorens-Morillo, Juan; Moreiro-González, José-Antonio. "De repente, ¿todos hablamos de ontologías?". *El profesional de la información*, 2007, noviembre-diciembre, v. 16, n. 6, pp. 562-568.
<http://dx.doi.org/10.3145/epi.2007.nov.03>

Schäuble, Peter. "Thesaurus based concept spaces", en Sigir 1987: New Orleans, Louisiana, USA. *Procs of the 10th Annual intl ACM Sigir conf. on research and development in information retrieval*. New Orleans, Louisiana, USA, June 3-5, 1987, pp. 254-262.

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

Tramullas, Jesús. "Bibliotecas digitales: una revisión de conceptos y técnicas". Universidad de Zaragoza, 2004.
<http://tramullas.com>

Tudhope, Douglas; Alani, Harith; Jones, Christopher. "Augmenting thesaurus relationships: Possibilities for retrieval", *Journal of digital information*, 2001, n. 3, pp. 123-148.

Zeng, Marcia L. "Z39.19-2005: NISO (National Information Standard Organization). *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. NISO Press, Bethesda, Maryland, USA, 2005.
<http://www.niso.org/standards/resources/Z39-19-2005.pdf>

IraLIS es el registro de autoridades de nombres de autor del repositorio E-LIS

Antes de subir documentos a E-LIS debes iralizarte siguiendo los "Criterios de firma".

The screenshot shows the E-LIS website interface. At the top, it says "E-LIS E-prints in Library and Information Science". Below this, there is a search bar and a navigation menu with options like "Home", "Browse", "Sign up to", and "About us". The main content area displays "Welcome to E-LIS" and "We now have 12302 documents in our archive". Below this, there is a section for "Recent submissions" with a list of articles and their authors.

<http://eprints.rclis.org>

The screenshot shows the IraLIS website interface. At the top, it says "IraLIS International Registry for Authors: Links to Identify Scientists". Below this, there is a navigation menu with options like "Idiomas", "Menú", and "Presentación". The main content area displays "Presentación" and "IraLIS (International Registry of Authors-Links to Identify Scientists) es un sistema de estandarización de las firmas de los autores científicos. Su objetivo es crear un registro de autoridades." Below this, there is a list of criteria for signing.

<http://www.iralis.org>



FOURSQUARE: TU BIBLIOTECA EN EL MOMENTO Y EN EL LUGAR ADECUADOS



Natalia Arroyo-Vázquez



Natalia Arroyo, licenciada en documentación por la *Universidad de Salamanca*, es documentalista en el *Departamento de Análisis y Estudios* de la *Fundación Germán Sánchez Ruipérez*. Sus principales líneas de trabajo son la web social y la web móvil aplicadas a las bibliotecas. Sobre ellos ha publicado varios artículos y comunicaciones e impartido docencia especializada para profesionales. Es autora del libro *Información en el móvil* de la colección *EPI-UOC*, e integrante de los equipos de redacción de *El profesional de la información* y del blog *Biblioblog*, en el que escribe habitualmente.

Fundación Germán Sánchez Ruipérez, Depto. de Análisis y Estudios
Nuestra Señora, 65. 37300 Peñaranda de Bracamonte, Salamanca, España
narroyo@fundaciongsr.es

Resumen

Foursquare es un servicio basado en localización en el que los usuarios comparten información sobre los lugares que visitan, bien añadiendo sus datos, comentándolos o recomendando productos o servicios que se ofrecen en ellos. En este texto se expone brevemente en qué consiste este servicio y se explican algunas pautas para su integración en la estrategia de redes sociales de las bibliotecas.

Palabras clave

Foursquare, Geolocalización, Bibliotecas, Redes sociales.

Title: *Foursquare: your library at the right time and in the right place*

Abstract

Foursquare is a tracker-based service in which users share information about the places they visit by adding their details, discussing or recommending products or services offered to them. This text outlines this service and provides some guidelines for its integration into the social network strategy of libraries.

Keywords

Foursquare, Geolocation, Libraries, Social networks.

Arroyo-Vázquez, Natalia. "*Foursquare: tu biblioteca en el momento y en el lugar adecuados*". *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 326-328.

<http://dx.doi.org/10.3145/epi.2012.may.15>

Lanzado en marzo de 2009, *Foursquare*¹ es un servicio basado en localización en el que la información compartida por sus usuarios es el lugar en el que se encuentran.

Para los más preocupados por la privacidad cabe aclarar que no se trata de decir dónde estamos en cada momento, sino más bien de compartir información sobre los lugares que visitamos (los locales de moda de nuestra ciudad, el plato estrella del restaurante de la esquina o mi espacio favorito al aire libre), información que antes quedaba en las conversaciones con amigos y/o conocidos y que los servicios de la web social están sabiendo recopilar y aprovechar.

El incentivo se obtiene en forma de puntos y de *badges* o insignias que se desbloquean cada vez que el usuario señala un número de sitios que cumplen una condición determi-

nada: por ejemplo, la insignia *bookworm* (gusano o rata de biblioteca) se desbloquea al señalar la tercera biblioteca o librería visitada.

Idóneo para dispositivos móviles, además de un juego colectivo, *Foursquare* constituye una importante base de datos de lugares organizados por categorías —espacios al aire libre, arte y entretenimiento, casa y trabajo, comida, universidad, tiendas y locales de viajes y nocturnos—, de imágenes, opiniones, consejos o recomendaciones sobre los lugares compartidos y de las personas que los visitan. Pero sobre todo es una colección de experiencias compartidas que se pueden recuperar con un dispositivo móvil desde un lugar cercano al que se han producido; es decir, en el momento y el lugar adecuados.

Artículo recibido el 03-05-12

Aceptación definitiva: 12-05-12

Gracias a *Foursquare* es posible descubrir a otras personas que se encuentran o han estado cerca de nosotros, compartir información sobre nuestros sitios favoritos, localizar el restaurante vegetariano más cercano o conocer lo que dicen de nuestro servicio, tienda o empresa.

Por todo ello comienza a situarse en el punto de mira de empresas e instituciones, que encuentran en él una forma de ganar visibilidad y un escaparate para mostrarse y atraer a sus seguidores.

“La *New York Public Library* ha sido pionera en el uso de *Foursquare* desde diciembre de 2010”

Su interés ha crecido con el anuncio, a comienzos de agosto de 2011, de la posibilidad de crear páginas en la plataforma sin restricciones (hasta entonces tenían que ser previamente autorizadas, lo cual no agilizaba precisamente el proceso). Estas páginas, al igual que las de *Facebook*, están pensadas para marcas e instituciones y permiten agrupar todas las ubicaciones relacionadas con la marca, escribir *tips* o recomendaciones específicas del tipo “prueba nuestro pollo a la barbacoa” o lanzar ofertas para nuestros seguidores (“descuento del 10% para quienes se acerquen a probar nuestro menú”, por ejemplo).

Las páginas ofrecen algunas opciones para personalizar el espacio (cambiar la imagen de cabecera y añadir un texto y enlaces a nuestro sitio y perfiles en las redes sociales) y analizar las estadísticas de uso. Entre quienes se han unido está la *Casa Blanca*², que comparte los lugares visitados por el presidente Obama en sus viajes y enlaza a información relacionada con ellos.

Foursquare en la estrategia de medios sociales de las bibliotecas

Las bibliotecas aprovechan la Web social desde hace tiempo al igual que otras empresas e instituciones, pero los servicios basados en localización no suelen contemplarse en sus estrategias de medios sociales, quizás porque no están muy extendidos aún o porque se encuentran en un momento incipiente de su implantación.

Sin embargo, los servicios basados en localización como *Foursquare* también recogen los comentarios de los usuarios sobre marcas, empresas e instituciones y son los que más directamente conectan lo virtual con lo presencial—quienes hacen *check-in* en un lugar es porque se encuentran en él— y son capaces de ofrecer contenidos y ofertas en contexto, proyectando el mensaje “si estás cerca quizás te interesen nuestros productos o servicios”.

La *New York Public Library*³ ha sido pionera en el uso de *Foursquare*, abriendo su propia página en diciembre de 2010, que cuenta con más de 45.000 seguidores. En ella se sugieren cosas que hacer en sus diferentes sedes, ofertas especiales para quienes obtienen el título de *mayor* o alcalde (que se logra cuando has visitado un número determinado de veces un mismo lugar, lo que significa que eres un habitual).

La *New York Public Library* ha creado además un *badge* especial con motivo del centenario de su emblemático edificio de la *Quinta Avenida*, fruto de la colaboración con *Foursquare*.

“*Foursquare* constituye una importante base de datos de lugares organizados por categorías”



<http://img.somosblackberry.com>

Para concluir, pasemos a un plano más práctico: ¿qué pasos se deben seguir para comenzar a utilizar *Foursquare* en una biblioteca? He aquí unas sugerencias:

1. Hazte visible

Señala la ubicación de tu biblioteca en el mapa de *Foursquare* y en otros servicios de geolocalización. En caso de que ya haya sido incluida, reclama el lugar—de esa forma podrás gestionar los datos de tu biblioteca— y completa todos los datos posibles. Así podrán localizarla quienes accedan a *Foursquare* desde un lugar cercano y la estarás dotando de mayor visibilidad.

2. Agrupa tu presencia

Crea la página de tu biblioteca, personalízala con una imagen de cabecera, un texto explicativo y enlaces a tu sitio web y perfiles en redes sociales.

<https://es.foursquare.com/nyp1>

3. Recomienda

Añade en tu página recomendaciones sobre cosas que hacer en la biblioteca, como actividades o nuevas secciones. Se trata de invitar a participar en las experiencias que ofrece la biblioteca.

4. Oferta

Lanza ofertas y regalos para las personas que llegan a la biblioteca o a quienes se convierten en *mayor*: “si eres el *mayor*, pásate por el mostrador y llévate una camiseta conmemorativa de regalo”. De esta manera pondrás cara a tus seguidores virtuales.

5. Escucha y responde

Presta atención a lo que se dice en *Foursquare* de tu biblioteca, responde cuando lo creas conveniente y conoce a las personas que la visitan. Obtendrás información sobre las opiniones y experiencias de tus usuarios y podrás mejorar tu reputación digital.

6. Recuerda a tus usuarios hacer *check-in*

Explica a tus usuarios qué es *Foursquare* y cómo utilizarlo.

Coloca también en la biblioteca carteles que indiquen que está presente en este servicio y que inviten a hacer *check-in* en el servicio. No sabes si tendrá mucho éxito o no, pero al menos será una actividad más y un motivo para que se hable de la biblioteca, y con algún lector seguro que conecta.

7. No abandones la página

Si abres la página debes mantener una continuidad de actividad en ella, de lo contrario mejor que no la abras. Resulta muy decepcionante para los lectores haber hecho el esfuerzo de inscribirse y participar, y que luego nadie responda.

8. Comparte la experiencia

Cuenta la experiencia a otros colegas en los foros profesionales, quizá pueda servirles a la hora de poner en marcha experiencias similares.

Notas

1. <http://www.foursquare.com>
2. <https://es.foursquare.com/whitehouse>
3. <https://es.foursquare.com/nyp1>

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 (99 euros/año), puesto que en esta modalidad no hay que cubrir los gastos de imprenta ni de correo postal.



Rapidez, precisión, flexibilidad

En el siglo XXI, más que nunca, la información es poder. Ayudar a empresas e instituciones a gestionar y rentabilizar su caudal informativo, tanto el que poseen como el que genera diariamente su actividad, es el objetivo y la razón de ser de Baratz.

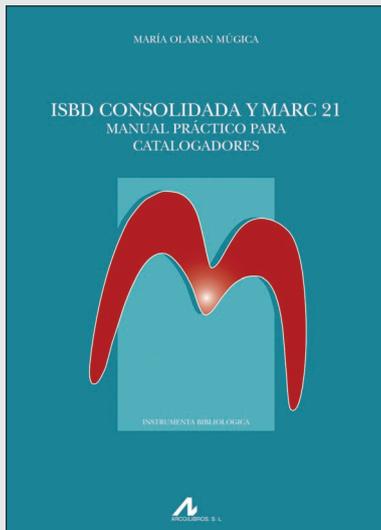
www.baratz.es



91 456 03 60 | informa@baratz.es

Raimundo Fernández Villaverde 28, 1.ª 28003 Madrid

NOVEDADES EDITORIALES



ISBD consolidada y MARC 21: manual práctico para catalogadores

María Olaran

En 2011 se publicó en inglés la edición definitiva de ISBD consolidada. Al no existir una traducción española de la versión definitiva, su difusión es aún limitada. Con este manual, totalmente práctico, se muestran los principales cambios que afectan a la descripción bibliográfica de monografías impresas modernas. También se aplican las Reglas de Catalogación en la resolución y forma de los puntos de acceso, codificando los registros en MARC 21.

Tras una introducción a la situación actual de la práctica descriptiva y normalizadora en el mundo a comienzos de la segunda década del siglo XXI, se presenta la realidad de la catalogación en España. Continúa con una exposición sobre la metodología utilizada, y finaliza con una tabla mostrando los principales cambios entre los registros bibliográficos de este texto y los manuales anteriores. En la parte práctica, se resuelven 111 supuestos.

Tanto por su contenido como por su estructura, esta obra se dirige a un público heterogéneo. En primer lugar, se orienta a todos los profesionales de la información, bibliotecarios y documentalistas, que cargan registros en catálogos o bases de datos. Es también un recurso para la docencia, tanto en el ámbito universitario, como en otro tipo de centros de formación, especialmente aquellos orientados a la preparación de oposiciones. Además, resulta útil para aquellas personas que con conocimientos en la materia, quieran actualizarse en las últimas tendencias normativas de una forma sencilla y didáctica.

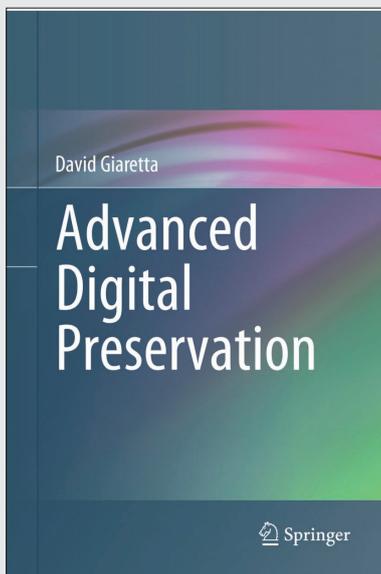
Olaran, María

ISBD consolidada y MARC 21: manual práctico para catalogadores

Madrid: Arco-Libros, 2012, 607 pp.

ISBN 978 84 7635 836 8

Precio: 23,80 €



Advanced digital preservation

David Giaretta

Es creciente la necesidad de abordar la fragilidad de la información digital, de la que nuestra sociedad depende en gran medida para el buen funcionamiento en todos los aspectos de la vida cotidiana. La preservación digital ha sido discutida ya en muchos libros y artículos, por lo que ¿hay todavía necesidad de uno más? Sí, porque en su mayor parte, esas publicaciones se centran en documentos, imágenes y páginas web -objetos que normalmente se prestan a ser fácilmente visualizados por un espectador humano. Sin embargo, hay muchos otros tipos de objetos digitales que tienen que ser preservados, como bases de datos, conjuntos de datos científicos o el mismo software.

David Giaretta, director de la *Alliance for Permanent Access*, y sus colaboradores explican por qué las herramientas y técnicas utilizadas para la conservación de los objetos representables no son suficientes para los otros tipos de objetos digitales, y proporcionan los conceptos, técnicas y herramientas necesarios.

1. Teoría: conceptos y técnicas esenciales para la preservación de la información codificada digitalmente.
2. Práctica: uso y validación de las herramientas y técnicas.
3. Cómo juzgar si el dinero está siendo bien empleado, en términos de eficacia y participación en los costos.

El libro está apoyado por muchas horas de vídeos y presentaciones del proyecto *Caspar* y por un conjunto de software de código abierto.

Giaretta, David

Advanced digital preservation

Springer, 2011, XXII, 510 pp.

ISBN 978 3 642 16808 6

Precio: 99,95 €



Creación de Archivos 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 7.0

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.

DIGIDIR 2.1

Directorio para Archivos, Bibliotecas y Museos con generación automática de estadísticas y sistemas de información geográfica (GIS).

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 (FASE DANUBIO)

Implementación del esquema ESE 3.4

(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
- 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 2.1 y OAIS ISO 14721
- Intercambio de metadatos en METS 1.9 (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.13)/RDA
- Creación de eBooks o libros digitales en formatos: ePub y Mobipocket.
- Adaptación de DIGIBIB a Linked Open Data



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



Nº 6002188



C/ Claudio Coello, 123. Madrid. Tel.: 915 81 20 01. digibis@digibis.com

www.digibis.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

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

Precios 2012

Suscripción normal: 186,54 € + IVA = 194 €

Suscripción personal: (sólo a domicilios particulares) 86,54 € + IVA = 90 €

Gastos de envío fuera de España: Europa: 40 € Américas: 60 €

Suscripción sólo online: 95,19 € + IVA = 112,32 €

Número suelto: 31,74 € + IVA = 33 €

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. "Título del artículo".
Título de la publicación periódica, año, mes, v., n., pp. xx-yy.
Dirección url iniciada en nueva línea sin barra y sin punto final
<http://dx.doi.org/10.xxxx/doi>

Ponencia presentada en un congreso:

Apellido, Nombre; Apellido2, Nombre2. "Título de ponencia".
En: *Nombre del congreso*, año, pp. xx-yy.
Dirección url iniciada en nueva línea sin barra y sin punto final
<http://dx.doi.org/10.xxxx/doi>

Monografías e informes:

Apellido, Nombre; Apellido2, Nombre2. *Título del trabajo*.
Lugar de publicación: editor, fecha, ISBN: espacios, sin guiones
Dirección url iniciada en nueva línea sin barra y sin punto final
<http://dx.doi.org/10.xxxx/doi>

Capítulo de una monografía:

Apellido, Nombre; Apellido2, Nombre2. "Título del capítulo".
En: *Apellido, Nombre; Apellido2, Nombre2. Título de la monografía*. Lugar de publicación: editor, fecha, pp. xx-yy. ISBN: espacios, sin guiones
Dirección url iniciada en nueva línea sin barra y sin punto final

Recurso en línea:

Apellido, Nombre; Apellido2, Nombre2. *Título del recurso*.
Dirección url iniciada en nueva línea sin barra y sin punto final

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>