

Revista española de documentación científica

vol. 39, n. 3 (2016)

Estudios

¿Relegan los informativos audiovisuales en España la investigación en Ciencias Humanas y Sociales? ‘Las otras ciencias’

María Dolores Meneses Fernández

Redes de conectividad entre empresas tecnológicas a través de un análisis métrico longitudinal de menciones de usuario en Twitter

David Azorín-Richarte, Enrique Orduna-Malea, José Antonio Ontalba-Ruipérez

Medición de la presencia de la lengua española en la Internet: métodos y resultados

Daniel Pimienta, Daniel Prado

Impacto de la crisis económica en las bibliotecas universitarias españolas

José Simón-Martín, Alicia Arias-Coello, Clara Simón-Blas

Influencia del programa Campus de Excelencia Internacional en la posición de las universidades españolas en el ranking de Shanghái

Teodoro Luque-Martínez, Luis Doña-Toledo, Domingo Docampo

Organización y puesta en valor del fondo documental del Instituto Español de Entomología (archivo histórico del Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas)

María Cruz Osuna Arias, Juana Molina Nortes, Carolina Martín Albaladejo

Gestión de fondos de archivos con datos enlazados y consultas federadas

Yusniel Hidalgo-Delgado, José A. Senso, Amed Leiva-Mederos, Pedro Hípola

Análisis bibliométrico de las tesis doctorales españolas indexadas con el descriptor “Sector de la educación” (1976/2014)

Olga Moreno-Fernández, Pilar Moreno-Crespo

Crítica de libros

Métricas de la Web Social para Bibliotecas. Nieves González-Fernández-Villavicencio. Barcelona: Editorial UOC, 2014 (El profesional de la información: 4). 282 p. ISBN: 978-84-9116-181-3

Adrián Macías Alegre

SEO: Introducción a la disciplina del posicionamiento en buscadores. Jorge Serrano-Cobos. Editorial UOC, 2015. Colección EPI Scholar: n. 3, 200 p. ISBN: 978-84-9064-956-5

Lluís Codina

ESTUDIOS / RESEARCH STUDIES

¿Relegan los informativos audiovisuales en España la investigación en Ciencias Humanas y Sociales? 'Las otras ciencias'

María Dolores Meneses Fernández*

* Universidad de La Laguna
Correo-e: dmeneses@ull.edu.es

Recibido: 05-08-2015; 2ª versión: 09-11-2015; Aceptado: 15-12-2015.

Cómo citar este artículo/Citation: Meneses Fernández, M. D. (2016). ¿Relegan los informativos audiovisuales en España la investigación en Ciencias Humanas y Sociales? 'Las otras ciencias'. *Revista Española de Documentación Científica*, 39(3): e139. doi: <http://dx.doi.org/10.3989/redc.2016.3.1326>

Resumen: Este estudio explora la idea de ciencia e investigación difundida por los informativos televisivos y radiofónicos en España. Responde el interrogante de si el periodismo audiovisual español discrimina la investigación en Ciencias Humanas y Sociales (CHS) respecto a las llamadas 'ciencias duras'. Para comprobarlo se rastrearon los informativos emitidos durante las ediciones de la Semana de la Ciencias celebradas de 2001 a 2014. Los resultados revelan a partir de qué disciplinas construyen los periodistas su percepción de la ciencia. En ese panorama audiovisual se cuantifica la cobertura de las investigaciones en Humanidades y Sociales. Los datos permiten, además, comparar el comportamiento del periodismo español con el de otros países, insertándolo en el mapeo internacional de la cobertura mediática de ambas ramas del conocimiento.

Palabras clave: Periodismo; televisión; radio; comunicación de la ciencia; Ciencias Humanas; Ciencias Sociales; transferencia de conocimiento.

¿Does Spanish television and radio news relegate research in Human and Social Sciences? 'The 'garbage science'

Abstract: This paper shows how science and research are presented by Spanish television and radio news. The study is guided by the question of whether or not Spanish audiovisual journalists relegate research in the Humanities and Social Sciences vs. the 'hard sciences'. To answer it, news broadcasts by public and private television channels and radio stations about Science Week from 2001 to 2014 were analyzed. The results reveal which disciplines journalists built their perception of science. The audiovisual media coverage of science is analyzed and quantified in order to isolate the specific coverage of Humanities and Social Sciences. The data obtained allow us to learn of the professional behaviour of Spanish journalists and to compare it to that of other countries in terms of how coverage of news concerning Humanities and Social Sciences is handled elsewhere.

Keywords: Journalism; television; radio; communication of Sciences; Humanities; Social Sciences; knowledge transfer.

Copyright: © 2016 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Attribution (CC BY) España 3.0.

1. INTRODUCCIÓN, JUSTIFICACIÓN Y CONTEXTO DEL ESTUDIO

Este estudio analiza la relación ciencias-medios de comunicación desde la perspectiva periodística. Se escogió en este estudio la Semana de la Ciencia porque es una unidad temática y temporal definida, celebrada en España entre la primera y la tercera semana de noviembre. Recibe el apoyo del Ministerio de Economía y Competitividad, de la Fundación Española para la Ciencia y la Tecnología (FECYT), del CSIC, y de numerosas instituciones públicas y privadas participantes. Se instauró con carácter nacional en 2001, habiéndose organizado catorce ediciones que actúan de escaparate para la ciencia y la investigación en las distintas disciplinas académicas.

La observación de los informativos televisivos españoles permite estimar que, a priori, solo determinados temas de las CHS compiten con las llamadas ciencias duras, a pesar de que los cuatro planes nacionales I+D+i aprobados desde 2001 distribuyen el presupuesto público de investigación entre todas las ramas del conocimiento.

La tabla I recoge ejes de actuación de los planes nacionales I+D+i, dominados por las CTIM (Ciencia, Tecnología, Ingenierías, Matemáticas), con presencia constante de las CHS. La consideración de cada rama del conocimiento en dichos planes justifica que los periodistas se remitan a ellos en la búsqueda de temáticas científicas y enfoques informativos. Las numerosas revistas académicas dan visibilidad a esas investigaciones, constituyendo otra fuente periodística potencial de primer orden.

El periodismo de ciencia viene siendo estudiado en diversos países desde la década de los años 70. Trabajos más recientes enfocan con persistencia las dificultades inherentes a la transformación de contenidos de especialidad en otros periodísticos, subrayando las áreas conflictivas en la comunicación entre científicos y periodistas (Peters y otros, 2008; Allgaier y otros, 2013a y b). De manera tangencial, algunos aluden al entendimiento de la ciencia según la tradición anglosajona, que la circunscribe a las ciencias duras (del inglés *hard sciences*, también traducido ciencias difíciles) o ciencias naturales, englobadas en las siglas CTIM. Esta acotación del periodismo científico a las CTIM

Tabla I. Presencia de las Humanidades y Sociales en los Planes Nacionales I+D+i del 2000 al 2015

Plan I+D+i	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
% PIB I+D+i	0,4	0,42	0,44	0,48	0,49	0,51	0,52	0,55	0,59	0,65	0,67	0,65	0,61	0,68	0,61	0,61	
M€/año	3.048	3.522	3.792	4.001	4.039	4.804	5.486	6.269	7.173	7.372	7.380	6.873		5.604	sd	sd	
Áreas prioritizadas de CHS	Ciencias Humanas. Ciencias Sociales. Ciencias Jurídicas. Conservación del patrimonio histórico cultural. Atención sociosanitaria. Sociedad de la información. Ordenación del territorio. Turismo, ocio y deporte. Atención especial a las Humanidades.				Humanidades. Ciencias Sociales y Económicas: patrimonio histórico y cultural, identidades culturales, ética e investigación científica. Lenguaje, pensamiento y sociedad. Comunicación. Lingüística. Nuevas tecnologías en Humanidades. Instituciones, desarrollo y sostenibilidad, cohesión social, económica y territorial. Identidad personal y colectiva. Decisiones públicas y privadas. Organizaciones, gobernanza, economía, sistemas políticos. Cognición, cerebro. Conducta y educación. Competitividad y eficiencia. Tecnologías turísticas.				Humanidades Ciencias Sociales, Económicas y Jurídicas Tecnologías turísticas, deporte y actividad física. Igualdad de oportunidades de mujeres y hombres. Fomento de la cultura científica y tecnológica.				Cambio demográfico y bienestar. Cambios e innovaciones sociales. Economía y sociedad digital. Investigación en Ciencias Sociales y Humanidades.				

Fuente: CICYT (2000, 2007), MEC (2005) e INE (2015).

es debido, como señala H.P. Peters (2013), a que los esfuerzos académicos se han centrado en el binomio ciencias duras-medios. Pero no se aportan argumentaciones epistemológicas, ni otras basadas en el sistema investigador de los Estados, que fundamenten no conceder cobertura periodística adecuada a la importancia de la investigación en CHS.

En el ámbito hispano, las CHS reciben la etiqueta 'letras', distanciándolas de su condición de disciplinas científicas; y en el ámbito anglosajón, una parte de los periodistas las consideran "*garbage science*" (Dunwoody, 1984; Haslam y Bryman, 2003a-b). Esta percepción en clave de ciencias/letras puede estar condicionando la cobertura periodística de las CHS respecto a las CTIM, a pesar de tratarse de una oposición desmontada por diversos autores desde hace décadas (Diamond, 1987; Hedges, 1987; Weiss y Singer, 1988; Fenton y otros, 1998; Hijmans y otros, 2003; Gil, 2004; Borjas y Vilchez, 2009; los citados por Summ y Volpers, [2015] Böhme-Dürr, 1992; Priest, 2014). Presuponer que las CHS son disciplinas de perfil científico bajo (ciencias blandas o fáciles) explica que otros autores utilicen las expresiones saberes de primer y segundo grado (García, 1999), o ciencias y letras; son apelaciones que remiten a la fiabilidad del conocimiento generado por unas y otras (Weigel y Pappas, 1981; Fara, 2014; Ruiz-Castell, 2014).

Sin embargo, el afianzamiento de grupos y proyectos de investigación transdisciplinares reconoce la complejidad de muchos de los fenómenos naturales, sociales y humanos, contribuyendo a difuminar esa división (González y Gómez, 2014). Cabe, por tanto, preguntarse si a la investigación humanística y social se le da una cobertura periodística acorde a la inversión económica que recibe. De ser así, las disciplinas humanas y sociales deberían integrar las temáticas estudiadas por el periodismo de ciencia en España (Meneses, 2004; Francescutti, 2010).

La idea imperante de periodismo de ciencia puede estar reduciendo la versatilidad de los redactores y del periodismo ofrecido a la sociedad (Dunwoody, 1984), dado que, a priori, la investigación científica sobre el ser humano en sus dimensiones individual, social, geográfica y cronológica no termina de atraer suficiente atención en las redacciones de los medios. Un ejemplo es la escasa atención periodística prestada a los resultados de las investigaciones, precisamente, en Comunicación Social (Brantner y Huber, 2013). Por el contrario, cada vez más investigadores sociales superan estas divisiones y desvalorizaciones, vinculando de forma indisoluble las disciplinas humanas y sociales al cambio medioambiental para entender las

transformaciones de la naturaleza y de la sociedad (Hackmann y St. Clair, 2012; UNESCO, 2013).

En esta línea integradora de las ciencias actúan organismos con autoridad científica internacional, como el CSIC de España, el CNRS de Francia, los *Research Councils* de Gran Bretaña y la Fundación Alemana de Investigación. Todos incluyen las CHS en su estructura, diferenciando las categorías principales Humanidades, Ciencias Sociales, Ciencias Naturales, Ciencias de la Vida e Ingeniería. Esas instituciones proyectan en la Semana de la Ciencia la diversidad investigadora, como muestran sus programas de actividades anuales; por ello surge la duda de cuánto y cómo se hacen eco los periodistas de esa pluralidad científica.

2. MARCO CONCEPTUAL

El periodismo es un agente formador en ciencia. Promueve la capacidad personal de elaborar opiniones y criterios propios; vuelve más atento al individuo; lo alerta, fomentándole la capacidad de reacción frente a abusos, manipulaciones e inversiones de la riqueza pública en propósitos ajenos al bienestar social. Además, complementa y dinamiza un sistema educativo formal lento en la incorporación de los avances recientes a los contenidos curriculares.

La noción de periodismo científico aplicada en este estudio incluye las investigaciones en disciplinas experimentales, humanas, naturales, de la salud y sociales, constituya su objeto de estudio la naturaleza, el universo, el ser humano, los grupos humanos o cualquiera de sus producciones tangibles e intangibles. Este concepto responde a la acepción 4 de Pelayo García Sierra (1999), que la define como "Ciencia categorial ampliada (las ciencias positivas culturales: Lingüística, Antropología)", que incluye la Antropología, la Arqueología, la Comunicación Social, la Historia, la Psicología Social o la Sociología, entre otras. Se contraponen a la acepción 3 "Ciencia categorial estricta, ciencia positiva, ciencia en sentido «moderno» (Mecánica, Termodinámica, Biología molecular)", que incluye desde la Astrofísica, la Biología, la Física y la Química hasta la Nanotecnología y las Neurociencias (las CTIM).

Se ha considerado también que la presencia periodística de las distintas ciencias se correlaciona con la capacidad de predecir y explicar de forma categórica hechos. Es una cobertura vinculada a la idea de autoridad, seriedad, rigurosidad y capacidad de influencia en individuos y sociedades; y esa influencia depende de la fiabilidad y seguridad que inspiran los resultados de investigaciones libres de relatividades y subjetividades del observador (Kreimer, 2010).

3. OBJETIVOS

El propósito de este estudio abarca los tres objetivos siguientes:

1. Categorizar los contenidos informativos referidos a ciencias, investigaciones científicas e investigadores/as.
2. Conocer la cobertura temática de las CHS en televisión y radio informativas.
3. Identificar el arquetipo de lo científico en los informativos de las cadenas de televisión españolas.

4. MATERIAL Y MÉTODOS

4.1. Recolección de datos

El estudio se basó en el rastreo de los contenidos sobre ciencia e investigación científica emitidos por las televisiones y radios durante las Semanas de la Ciencia en España. La búsqueda se inició en las Web de los medios televisivos y radiofónicos; sin embargo, el reducido servicio a la carta de algunas cadenas privadas obligó a abrir la búsqueda a toda la Internet.

4.2. Unidades de análisis

Recurrir a la Semana de la Ciencia como hilo conductor facilitó la identificación de las unidades de análisis. Se recopilaban las piezas halladas, independientemente del género periodístico y del programa, priorizando los informativos y las noticias insertas en magazines. Consideramos como unidades diferentes el avance dado al principio del informativo y la noticia desarrollada, sumándose ambos minutos en la duración computada.

4.3. Medios estudiados

Justificamos centrar el estudio en medios audiovisuales porque en los años estudiados la televisión era el medio con más impacto en la sociedad española. La audiencia general de medios de 2001 indica que la penetración de la televisión en todas las franjas de edad era del 89,2%, y la radio alcanzó el 52,4%, frente al 20,4% de Internet. En ese año, el *share* de TVE1 era el mayor, con un 29,7%, seguido por Telecinco (21,2%), Antena 3 (19,5%) y La2 de TVE (5,2%) (AIMC-Estudio General de Medios, 2001; Peters y otros, 2014; Whitelegg y otros, 2009). En 2014, la televisión seguía alcanzando un 88,6% de penetración; la radio, un 61% e Internet, un 60,7%. Por tanto, seleccionamos los dos medios con mayor consumo social (AIMC, 2014). En 2014, Telecinco, Antena 3 y TVE habían pasado a ser, por ese orden, las cadenas con mayor *share*.

En este escenario de consumo mediático, la búsqueda abarcó todas las televisiones y radios españolas. La estructura del audiovisual en España requirió diferenciar los distintos canales de televisión y emisoras de radio. Respecto a RTVE, las siglas TVE aúnan los informativos regionales; las siglas TVE1, los programas emitidos a nivel nacional por La1; y las siglas TVE2, los emitidos por La2.

A diferencia de RTVE, las cadenas privadas, autonómicas y locales ofrecen en sus web oficiales un servicio a la carta limitado. Por ello se consultaron sus canales en YouTube donde alojan los programas informativos. En esta web también se identificaron videonoticias publicadas por instituciones científicas durante la Semana de la Ciencia.

4.4. Programas

De TVE se visionaron los Telediarios de La1 y La2 Noticias. Los informativos territoriales de TVE revisados fueron Informatiu Balear, Informativo de Madrid, Informativo Telerioja, L'informatiu Comunitat Valenciana, Noticias Andalucía, Noticias Aragón, Noticias de Castilla y León, Noticias de Castilla-La Mancha, Noticias de Ceuta, Noticias de Extremadura, Noticias de Melilla, Noticias Murcia, Panorama Regional (Principado de Asturias), Telecanarias, Telecantabria, Telenavarra, Telenorte País Vasco y Telexornal Galicia. También se incluyeron las noticias de las televisiones autonómicas y locales. La programación de RNE abarcó las distintas emisoras, incluida Radio 4 de RNE que emite en Cataluña. El rastreo de los informativos de Baleares, Cataluña, Galicia y Valencia incluyó los contenidos en lenguas cooficiales.

Acotamos el estudio a contenidos informativos, incluidos magazines, porque interesaba conocer la idea de ciencia ofrecida por los periodistas. Justificamos esta delimitación porque el perfil de licenciado/a en Ciencias de la Información o de la Comunicación, y de graduado/a en Periodismo está incorporado a las redacciones informativas, mientras que se carecía de datos sobre los perfiles profesionales de las productoras audiovisuales que proveen a las televisiones de otros contenidos de ciencia.

4.5. Periodo temporal revisado

Centrar el estudio en la Semana de la Ciencia minimizó la influencia del debate sobre el concepto de periodismo científico (véase Summ y Volpers, 2015) en la selección de noticias. Es una referencia cronológica clara en la que investigadores e instituciones disponen de una plataforma pública para exhibir líneas de trabajo y logros. La duración real de dos o más semanas permitió obtener un corpus

de noticias estadísticamente representativo, ya que las fechas de celebración variaron en cada comunidad autónoma, si bien destacó el despliegue en Madrid, sede de RTVE, del CSIC, de la FECYT, de universidades y otras entidades participantes. El periodo cronológico resultante excedió las catorce semanas oficiales, ya que se comenzó a informar antes de inaugurarse el evento y se continuó más allá de la clausura.

Tabla II. Fechas de referencia de las semanas de la ciencia en la Comunidad de Madrid

I Semana de la Ciencia 5-11 noviembre 2001	VIII Semana de la Ciencia 10-23 noviembre 2008
II Semana de la Ciencia 4-17 noviembre 2002	IX Semana de la Ciencia 9-22 noviembre 2009
III Semana de la Ciencia 3-16 noviembre 2003	X Semana de la Ciencia 8-21 noviembre 2010
IV Semana de la Ciencia 10-24 noviembre 2004	XI Semana de la Ciencia 7-20 noviembre 2011
V Semana de la Ciencia 7-20 noviembre 2005	XII Semana de la Ciencia 5-18 noviembre 2012
VI Semana de la Ciencia 6-19 noviembre 2006	XIII Semana de la Ciencia 4-17 noviembre 2013
VII Semana de la Ciencia 5-18 noviembre 2007	XIV Semana de la Ciencia 10-16 de noviembre de 2014

4.6. Organización y tratamiento de los datos

Los datos fueron registrados en la base de datos de un programa estadístico (SPSS, v.22). El análisis de contenido se basó en operativizarlos, clasificándolos en dos tipos de variables. Por un lado, las tres variables independientes contextuales fecha, medio y programa de emisión; y, por otro, las variables dependientes descriptivas nominales, que incluyeron las ciencias y los términos (palabras clave) utilizados por periodistas y entrevistados, las instituciones citadas, la actividad profesional de las fuentes, la concepción (catego-

rial ampliada o estricta) de ciencia manifestada por fuentes y periodistas, la representación de científico/a (hombre, mujer, joven, mayor), el retrato arquetipo de investigador/a (nombre propio) y la duración de la pieza (tabla III).

Las variables dependientes permitieron trazar la imagen de ciencia e investigación proyectada por los medios audiovisuales. Las variables independientes facilitaron detectar las diferencias entre los medios en la cobertura de la investigación en CHS.

4.7. Análisis de datos

Se compararon los medios de comunicación y programas, el número de unidades informativas, las disciplinas científicas citadas, los términos utilizados, los investigadores entrevistados, las instituciones mencionadas y los retratos arquetipo de científico/a. Se aplicó estadística para cuantificar la incidencia de cada variable, calculándose mediante análisis de frecuencia de respuesta múltiple a partir de tablas y gráficas de contingencia. Asimismo, se despejó la correlación entre cobertura de temas de CHS y medios.

5. RESULTADOS

El primer resultado relevante es que, habiéndose instituido la Semana de la Ciencia en 2001, es en 2009 cuando comienzan a aparecer noticias sobre ella. De 2001 a 2008 no se localizó ninguna alusión informativa, aun recurriendo a la búsqueda cruzada de términos afines a nuestro campo semántico. La primera pieza registrada la emitió *Noticias Murcia* el 5-11-2009 y la última el 16-12-2014. Entre 2001 y 2008 las referencias procedían de agencias (*Europa Press*), Canal UNED (televisión educativa) y TV-Web de universidades; en ningún caso de informativos.

Tabla III. Etiquetas de la base de datos estadística correspondientes a las variables independientes (1-3) y dependientes (4-13) que describen cada pieza informativa

1. Fecha de emisión	2. Medio de comunicación	3. Programa
4. Disciplinas citadas ($d_1 - d_n$)	5. Términos vinculados a la idea de ciencia e investigación ($t_1 - t_n$)	
6. Instituciones citadas ($i_1 - i_n$)	7. Fuentes expertas entrevistadas ($f_1 - f_n$)	
8. Profesión de las fuentes entrevistadas		
9. Actitud de la fuente respecto a qué es ciencia:		
categorial estricta	categorial amplia	no manifestada/neutra
10. Actitud del periodista respecto a qué es ciencia:		
categorial estricta	categorial amplia	no manifestada/neutra
11. Representación de científicos/as:		
hombre	mujer	joven sénior
12. Investigador/a citado/a como retrato arquetipo (nombre)		
13. Duración de la pieza		

La búsqueda en los medios audiovisuales nacionales y regionales proporcionó los datos siguientes: se registraron 115 piezas periodísticas; de ellas 26 aludieron a una o más disciplinas de CHS. En las 115 piezas se citaron 89 disciplinas científicas, utilizándose 739 términos científico-tecnológicos que, tras agrupar las equivalencias semánticas, quedaron organizados en 446. Informaron sobre la Semana de la Ciencia 37 programas en 21 cadenas y canales, 4 emisoras, más las Web de RTVE, Telecinco y Antena 3 TV, y un canal Youtube.

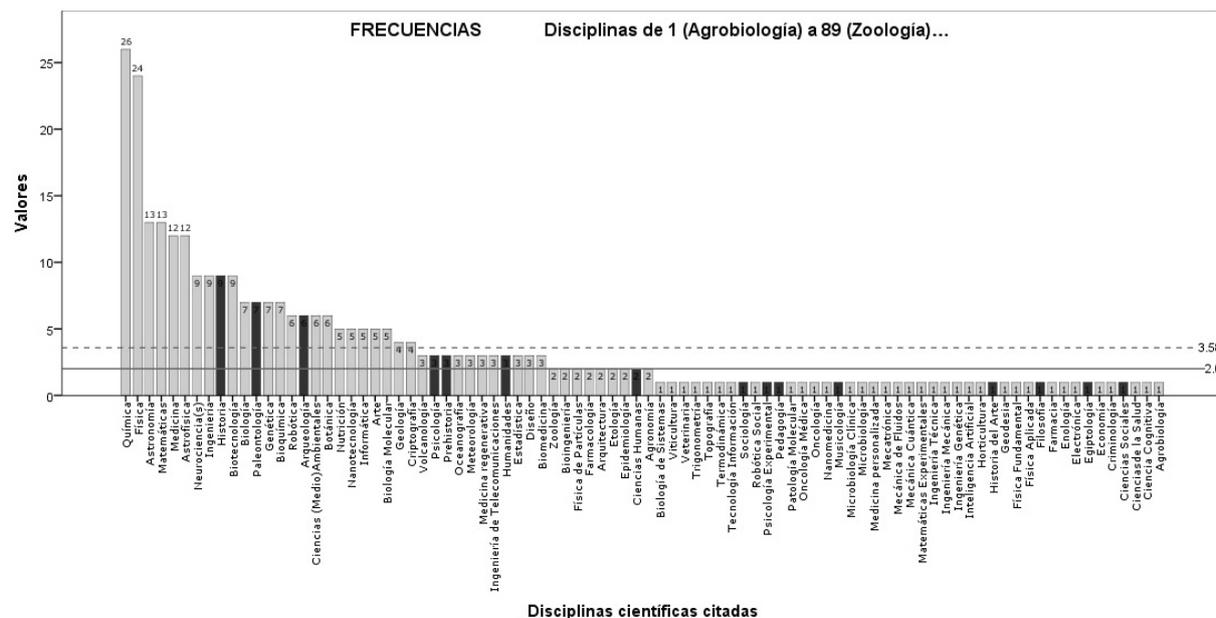
La figura 1 recoge las 89 disciplinas mencionadas y las frecuencias de cada una. El dominio de la Química, Física, Astronomía, Matemáticas, Medicina y Astrofísica es evidente. La primera disciplina no CTIM es la Historia, seguida de Paleontología (vinculada a Prehistoria) y Arqueología; más alejadas, Psicología, Prehistoria y Humanidades. La presencia reducida de la Psicología contrasta con sus métodos experimentales y cuantitativos, y con ser una de las disciplinas sociales más tratadas por periodistas en otros países (Schmierbach, 2005; Cassidy, 2008, 2014). La Sociología y la Filosofía son testimoniales, como la Egiptología, una disciplina a priori de las que más presencia se esperaba, en función de la espectacularidad de los descubrimientos publicados en los medios.

La figura 1 muestra también que el estudio de la cultura material y restos paleontológicos por la Arqueología y la Paleontología sí recibe periodísti-

camente la consideración de ciencia. No ocurre así con la investigación de la narrativa inmaterial no observable directamente ni sujeta a experimentación. Si bien el análisis de contenidos confirma la consideración de la Historia como ciencia, no se cita como tal, como lo ejemplifica la noticia del *Informativo de Madrid* (TVE) del 8-11-2011, dedicada a las actividades en laboratorios y museos de historiadores y arqueólogos. A pesar del estatus científico implícito, se observa que, al tratar la Historia, suele ser de forma subsidiaria de las CTIM, por ejemplo para hablar de la Astronomía en la época del Imperio Romano o, en *Noticias* de TV del Principado de Asturias del 8-11-2014, para comentar "la ciencia [leyes físico-químicas] que empleaban los hombres del paleolítico".

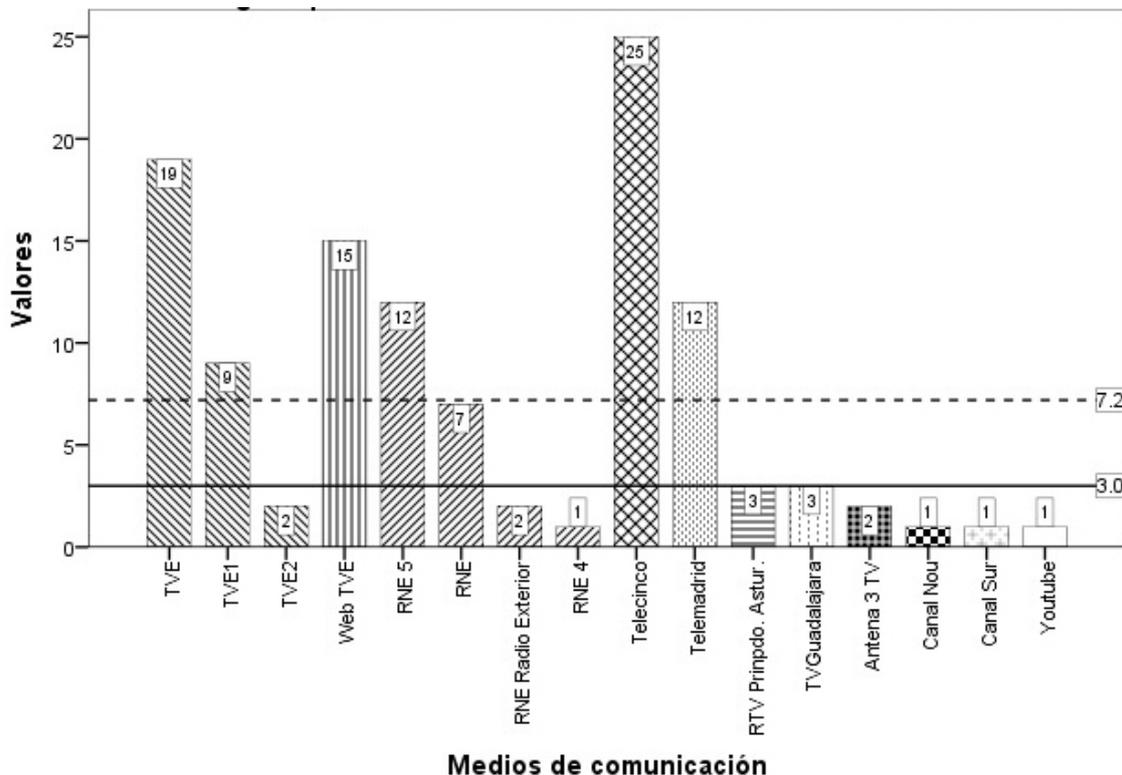
La figura 2 muestra que los términos relacionados con las Ciencias Humanas y Sociales se limitan a 11, y a un total de 57 frecuencias de uso (N = 739). Se trata de una presencia reducida respecto al total de términos (T = 446), en su mayoría de las CTIM. Estos datos corroboran la presencia escasa de la investigación en CHS en la agenda temática de los informativos. La frecuencia mayor corresponde a patrimonio histórico arqueológico (16) seguida de museos (13). Por tanto, los términos que destacan de Humanidades son los asociados a la investigación de restos materiales histórico-arqueológicos, objetos de estudio observables y tangibles para los periodistas. Las Ciencias Sociales están casi ausentes.

Figura 1. Disciplinas ordenadas de más a menos citadas*



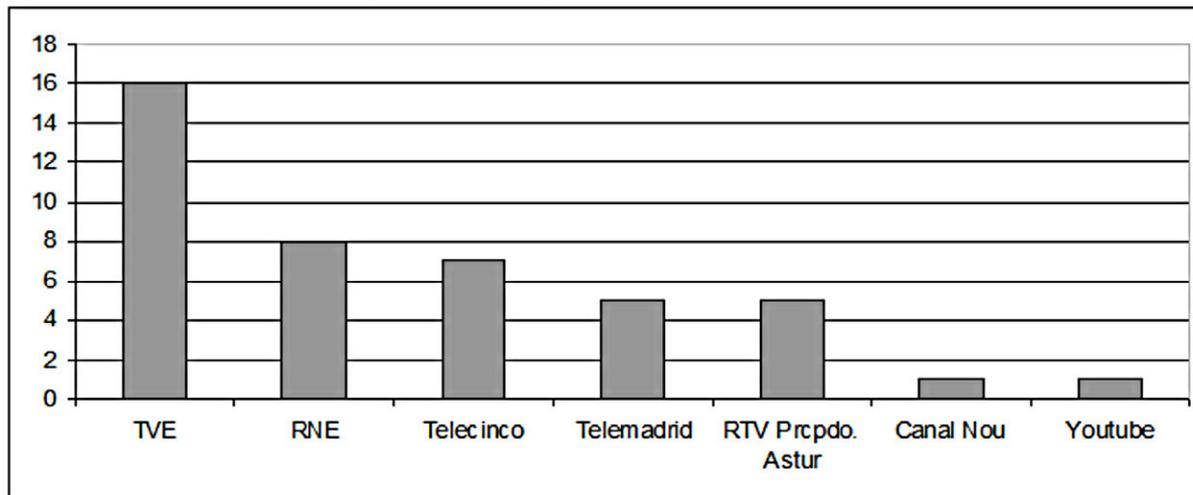
* Las barras negras corresponden a disciplinas de CHS. Línea horizontal continua: M_e ; línea horizontal discontinua: \bar{X} .

Figura 3. Medios de comunicación que publicaron las noticias referidas a la Semana de la Ciencia*



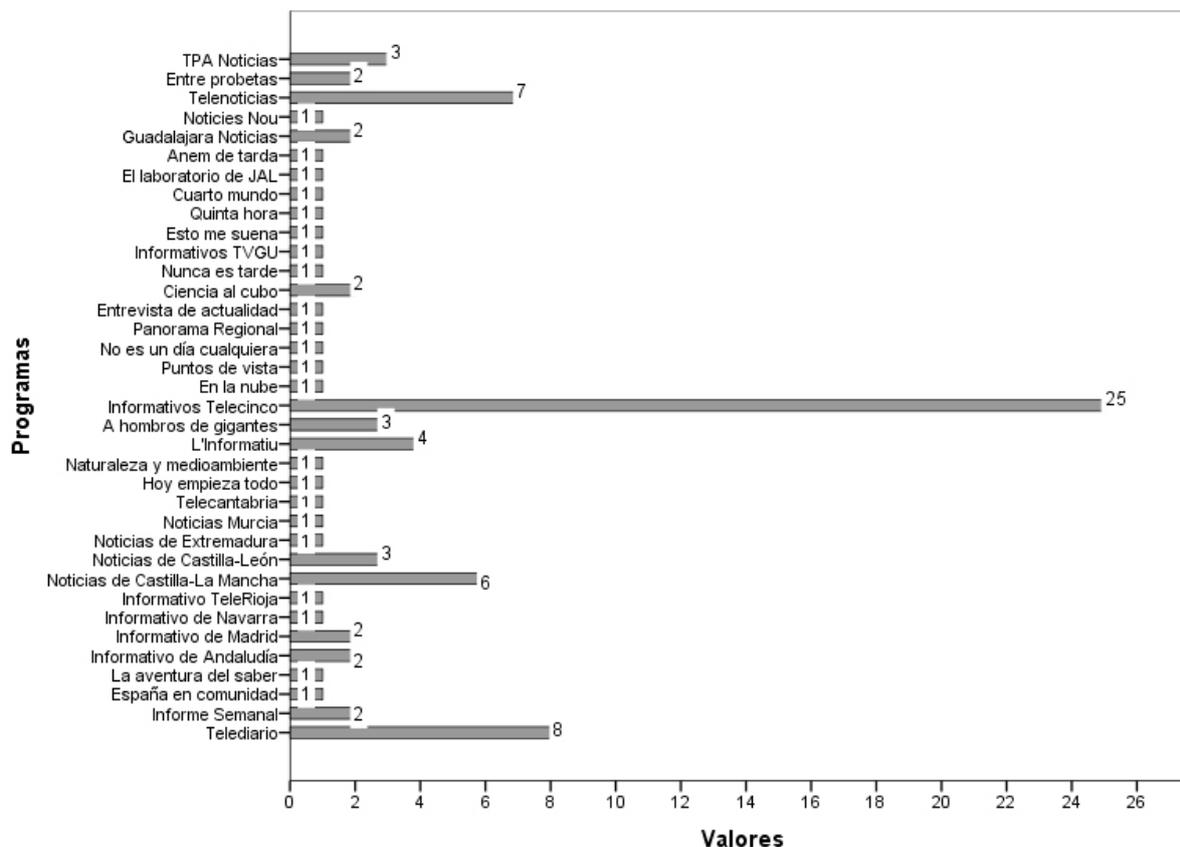
* Línea continua: M_e . Línea discontinua: \bar{X} .

Figura 4. Relación entre investigación en CHS y medios



dades autónomas. De las televisiones autonómicas, Telemadrid sigue la tónica de las televisiones públicas nacionales, y, al contrario, sorprende la escasa presencia e incluso ausencia de otras comunidades autónomas con actividad científica conocida.

La figura 4 recoge los 6 medios -de 29- que incluyeron las CHS al informar de investigación. TVE y RNE son los más destacados. Telecinco baja de forma sustancial, lo que indica que despuntó en el cómputo total por las noticias sobre las CTIM.

Figura 5. Programas que han emitido contenidos de ciencia e investigación durante las Semanas de la Ciencia

Las figuras 3, 4 y 5 muestran las comunidades autónomas que mayor presencia otorgaron a ciencia e investigación, según los contenidos de los programas televisivos y radiofónicos analizados. Sin discriminar por disciplinas, sobresalen Madrid (12), Castilla-La Mancha (7), Valencia (5), Asturias (4), Castilla-León (3), Andalucía (2) y cierran Cantabria, Cataluña, Extremadura, La Rioja y Murcia (1). Por tanto se deduce un desequilibrio entre la actividad investigadora de ciertas regiones y su escaso impacto en los informativos nacionales y territoriales. Por otro lado, si se observan los temas de ciencia informados vemos que si una comunidad autónoma destaca en determinadas disciplinas y cuenta con centros de investigación y museos temáticos, las noticias locales se centran en esas disciplinas al disponerse de fuentes expertas cercanas. Lo ejemplifican Asturias y Cantabria con Paleontología y Prehistoria; Valencia y las tecnologías aplicadas a la gastronomía; y Madrid con Robótica, Nanotecnología, Historia y Arte.

El escenario trazado hasta aquí es de dominio de la ciencia circunscrita a las CTIM, con escasa presencia de las CHS y limitada a pocas disciplinas (figura 2).

Estos resultados contrastan con estudios recientes de los Países Bajos que muestran que el 56% de la cobertura de la ciencia se relaciona con las Ciencias Sociales y el 8% con las Humanidades (Hijmans y otros, 2003); y de Croacia, con datos que van en la misma línea: el 19% de Ciencias Sociales y el 21% de Ciencias Humanas (Šuljok y Brajdić Vuković, 2013, citados por Summ y Volpers, 2015).

No obstante, la comparativa de los tiempos informativos dedicados a CHS y CTIM matiza el desequilibrio. Tras desagregar disciplinas y minutado se percibe la atomización del tiempo de las CTIM entre numerosas disciplinas, con 3:46:49 horas (59,4%) de las 6:21:44 analizadas; mientras que las CHS ocupan 2:34:55 horas (40,6%), con piezas de duración superiores a la noticia estándar.

La presencia superior, diversificada y continuada de las CTIM ratifica la dominancia de la acepción categorial estricta de ciencia, ejemplificada en una noticia de RTV del Principado de Asturias del 11-11-2014, en la que una alumna de Bachiller responde a la pregunta del periodista que "está bien que todos, incluso los de Humanidades, vengamos a estas actividades" [de la Semana de la Ciencia].

De las 60 instituciones científicas mencionadas sobresalen la FECYT (5), el CSIC (4), la Agencia Espacial Europea (4) y el Jardín Botánico de Madrid (4). Por otra parte, el Museo Naval de Madrid (3), el Centro de Ciencias Humanas y Sociales del CSIC (2), el Archivo Histórico de Castilla-León (1), el Museo Sorolla de Madrid (1), el Museo Tito Bustillo (1), la *Smithsonian Institution* (1), la Sociedad Prehistórica de Cantabria (1) y el Museo de Altamira (1) ratifican la Historia, la Arqueología y la Paleontología como las disciplinas de CHS más destacadas.

El arquetipo del investigador reproducido viene establecido por los 15 nombres propios citados, de los cuales Jane Goodall es la única mujer; son Santiago Ramón y Cajal, Severo Ochoa, Charles Darwin, Isaac Peral, Juan de la Cierva o Peter Higgs. Pocos investigadores actuales y ninguno de CHS.

Solo 8 fuentes entrevistadas (7%) y 20 periodistas (17,4%) hablaron de la ciencia según la acepción categorial ampliada. De las 42 fuentes entrevistadas, 33 eran hombres y 9 mujeres. Sus actividades profesionales las describen como investigadores o técnicos de museos.

6. DISCUSIÓN

La comparativa con los resultados obtenidos por autores que observaron la relación CHS-periodistas revela diferencias sustanciales. S. Dunwoody (1984), H.P. Peters y otros (2012) señalan que los investigadores de CHS son mucho más propensos que los de las ciencias duras a contactar frecuentemente con los periodistas. Tal predisposición se debe, en parte, al interés de los medios por estas disciplinas, a pesar de que constituyen una comunidad de investigadores menor que la de, por ejemplo, Biología y Neurociencia. J. Olmos y otros (2014a, 2014b) llegan a conclusiones similares en lo concerniente a la predisposición de los investigadores en CHS a interactuar con la sociedad, en procesos comunicativos distintos al mediático.

En países punteros en I+D+i donde se llevaron a cabo los estudios citados (EEUU, Japón, Alemania, Reino Unido y Francia), la regla Ingelfinger es eficaz en CTIM (salvo Matemáticas), y mucho menos en CHS (salvo Psicología). Esto se traduce en que la disposición de los investigadores de CHS a proporcionar a los periodistas información sobre sus investigaciones en curso es mayor que en las CTIM, ya que en estas últimas se es reacio a compartir estudios no publicados por riesgo a que pierdan valor. Esta condición de las revistas científicas es menos obstaculizadora en Humanidades y Sociales, lo que concuerda con el deseo de los investigadores en CHS de que los conocimientos

generados en sus investigaciones formen parte de la cultura general de todo el mundo. Esa actitud los aleja del elitismo de los investigadores de las CTIM y del modelo de déficit.

En el caso de EEUU, M. Schmierbach (2005) verificó que la diferenciación entre CHS y CTIM condiciona la actitud de los periodistas. Tras encuestar a 56 redactores de otros tantos diarios estadounidenses, observó que se ven influidos en la cobertura de noticias sobre Ciencias Sociales, a las que relegan frente al carácter cuantitativo y más preciso de otras investigaciones. Comprobó incluso que se otorga a lo cuantitativo más interés periodístico y predisposición a publicarlo que al mismo estudio cualitativo con los mismos resultados, dejando al margen en la decisión el valor noticioso.

Este comportamiento periodístico, en el que encajan los informativos españoles, se explica por el hecho de que los periodistas se adhieren a las actitudes predominantes respecto a lo que es aceptable o no como noticias científicas. Los medios de comunicación tienen la capacidad de establecer la agenda temática pública y fijar cómo han de entenderse e interpretarse dichos temas, en coherencia o no con las comunidades de expertos. Los *gatekeepers* participan en esa influencia mediática, por lo que es plausible que editores y redactores opten por no ir a contracorriente al configurar la actualidad informativa; bien al contrario, se adaptan a ella al decidir qué es noticia (Dittus, 2005; Pena, 2006); y, por extensión, qué es noticia de ciencia.

El comportamiento periodístico descrito plantea dudas sobre la escasa presencia de las Humanidades y Sociales en la agenda informativa. Si el modelo de déficit o modelo de educación pública plantea que solo los científicos son capaces de entender la complejidad de los temas científicos (Pouliot, 2009), cabría suponer que esa complejidad se minimiza en la cobertura periodística de la investigación en CHS, dado que los estudios de Periodismo se incluyen en ellas. Sin embargo, el estándar de investigación científica imperante y las propias investigaciones sobre periodismo científico dificultan que la cobertura periodística de las CHS se aproxime a la de las CTIM (Peters y otros, 2012). Este desequilibrio lleva cada vez a más autores a admitir la necesidad de incluir las CHS en el concepto de periodismo científico (Haslam y Bryman, 2003b; Peters y otros, 2012).

En comparación con los países citados, cabe cuestionarse si la desproporción ente CHS y CTIM en los medios audiovisuales españoles se debe a la consideración de las distintas disciplinas en la tradición académica local, a la imagen dominante de investigación científica o a la escasa proactividad

de los periodistas. Se explicaría así que una sola noticia de las analizadas incida en la importancia de la investigación humana y social.

El hecho de que pocos periodistas apliquen la acepción categorial ampliada de ciencia concuerda con que solo un periodista haya trascendido la agenda oficial para informar de la importancia de la investigación científica en CHS. Este comportamiento periodístico proyecta una idea de ciencia focalizada en enfermedades, planetas, estrellas, laboratorios, batas blancas y probetas que, en ocasiones determinadas, dan paso a historiadores, antropólogos, psicólogos, politólogos, economistas o sociólogos.

Es sabido que la relación periodistas-científicos viene condicionada por la transformación de mensajes científicos en otros periodísticos y por el dominio temático de los redactores; no obstante, ambos condicionantes no deberían afectar a la cobertura de las CHS, ya que los estudios de Periodismo instruyen en teoría social, Humanidades y Sociales. Si se añade un contexto favorable al contar con científicos predispuestos a interactuar con los periodistas, los resultados de este estudio llevarían a entender la escasa presencia de las CHS en los informativos por el posible cuestionamiento de su estatus científico. Si estudios previos indican que la construcción periodística de la ciencia en España es superficial (Moreno, 2010; Moreno y Gil, 2014), este trabajo muestra que, en el caso de las noticias de CHS, lo es aun más.

7. CONCLUSIONES

El periodismo audiovisual español reproduce un concepto restringido de ciencia e investigación, ya que la mayoría de las disciplinas de las que se informa corresponde a la idea de ciencias duras: Ciencia, Tecnología, Ingeniería, Matemáticas.

La categorización de la agenda periodística científica revela que solo unas pocas CHS irrumpen en la agenda informativa, y lo hacen con un tratamiento asemejado al de las CTIM al tratarse de disciplinas humanísticas y sociales con métodos basados en trabajos de campo, análisis de materiales, experimentación y reproducción. Por tanto, los periodistas consideran que la información de ciencia tiene en las CTIM, y no en las CHS, su agenda temática natural.

La excepción la marcan investigaciones en Humanidades centradas en restos patrimoniales y en procesos taxonómicos. Los factores que favorecen esa presencia de las CHS en los informativos son los que las relacionan con objetos de estudio materiales observables a los que se puede aplicar méto-

dos experimentales. Incluyen Historia, Paleontología, Arqueología y Prehistoria, y, en menor medida Pedagogía, Psicología experimental y Sociología.

Si bien las CHS irrumpen en la agenda informativa cuando estudian materiales tangibles, las investigaciones con objetos de estudio intangibles tienen menor presencia. Paradójicamente, algunos de los objetos de estudio de las CTIM con mayor presencia informativa encierran gran carga de abstracción (la estructura de la materia, ciertas cuestiones astronómicas), si bien son representables visualmente, algo importante en televisión.

Los tecnicismos utilizados inciden en lo afirmado al revelar que la sociedad recibe una idea de ciencia e investigación asociada a enfermedades, estrellas, galaxias, planetas, universo, energías, moléculas, átomos, protones, neutrones, aparatos complejos, laboratorios y probetas. Una visión afín a las ciencias naturales y experimentales, a las tecnologías, las ingenierías y las matemáticas.

Estamos ante una idea de ciencia dominada por la diversidad, continuidad y sentido de utilidad de las CTIM. Solo el cómputo del tiempo informativo favorece a las CHS.

En consecuencia, los medios ofrecen una reconstrucción periodística de la ciencia que excluye una parte sustancial de la investigación avalada por los Planes Nacionales de I+D+i. Esa escasa cobertura informativa de la investigación en CHS choca con el sistema investigador español y con la dinámica periodística de países desarrollados.

Como particularidad se detecta influencia territorial en la cobertura de las investigaciones en CHS, ya que en las comunidades autónomas con patrimonio histórico, artístico, arqueológico y paleontológico notable, divulgado en museos y apoyado por una política científica favorable, esas temáticas se convierten en centro de atención periodística. Esto indica que, si bien en periodismo prima la noción categorial estricta de ciencia, en contextos que facilitan a los periodistas temas y fuentes de CHS, estas disciplinas entran en la información científica.

La omisión o desconocimiento de la estructura del sistema científico-investigador español por los periodistas lleva a reflexionar sobre la formación en ciencia, epistemología y teoría social de los estudiantes de Periodismo. Junto con esto, la polarización del sistema educativo en ciencias/letras contribuye a relegar la investigación humanística y social.

Los desequilibrios en la comunicación social de la investigación son contraproducentes para la sociedad, para el sistema I+D+i y para el periodismo científico. Esta imagen parcial ofrecida por el medio más influyente en España, la televisión,

fomenta una valoración pública incompleta de la investigación nacional, supeditándola a un prestigio social y académico hoy anacrónico. El periodismo reproduce esa imagen en lugar de reajustarla a la realidad actual, informando de la influencia que tiene la investigación en las distintas disciplinas en el desarrollo de los individuos y las sociedades.

Dado que los temas tratados en las noticias se refieren a investigaciones españolas, la escasa presencia de las CHS se interpreta, además de por las razones expuestas, porque los trabajos desarrollados en estas ramas de conocimiento no son suficientemente difundidos por sus responsables más allá de las publicaciones académicas.

Lo descrito arriba muestra que los periodistas no contribuyen a que las disciplinas humanas y sociales se normalicen en la agenda informativa sobre investigación científica. De hecho, cuando se informa de investigaciones en CHS se hace subordinándolas a las CTIM o porque los periodistas aprovechan iniciativas de investigadores de Humanidades y Sociales, no por iniciativa propia.

Los expertos entrevistados tampoco superan la escisión entre Ciencias de la Naturaleza y Experimentales, y Ciencias Humanas y Sociales. Y el arquetipo de científico representado, hombre y mayor, se asocia con las CTIM; en ningún caso con las CHS.

El tratamiento periodístico de las investigaciones en CHS descrito no contempla su capacidad de influir en la toma de decisiones por individuos, grupos humanos y gobernantes para la mejora social. Una única noticia vincula la investigación en CHS a la toma de decisiones políticas.

Lo anterior indica que el factor condicionante no es tanto el uso de la definición estrecha o amplia de periodismo de ciencia (Summ y Volpers, 2015), sino la aplicación por los periodistas de la acepción estricta o ampliada de ciencia.

El estudio ha revelado que, de darse cualquier protagonismo de las CHS fuera de la Semana de la Ciencia, correspondería a investigaciones extranjeras de las que se hacen eco las televisiones y radios españolas.

8. REFERENCIAS

- Allgaier, J.; Dunwoody, S.; Brossard, D.; Lo, Y.-Y.; Peters, H.P. (2013a). Medialized Science? Neuroscientists' reflections on their role as journalistic sources. *Journalism Practice*, vol. 00 (00), 1-7.
- Allgaier, J.; Dunwoody, S.; Brossard, D.; Lo, Y.-Y.; Peters, H.P. (2013b). Journalism and Social Media as Means of Observing the Contexts of Science. *BioScience*, vol. 63 (4), 284-287. <http://dx.doi.org/10.1525/bio.2013.63.4.8>
- Borjas Gil, M.I.; Vilchez Paa, C.F. (2009). Ciencias «duras» vs. ciencias «blandas». *Revista Electrónica de Humanidades, Educación y Comunicación Social*, vol. 4 (7), 195-209.
- Brantner, C.; Huber, B., (2013). How visible is Communication Studies? Press Coverage of the Discipline in Three German-Language Quality Newspapers. *Studies in Communication/Media*, vol. 2 (2), 247-264. <http://dx.doi.org/10.5771/2192-4007-2013-2-247>
- Cassidy, A. (2008). Communicating the Social Sciences. En: Bucchi, M. and Trench, B. (editores) *Handbook of Public Communication of Science and Technology*. Londres, Routledge, Ch. 16, p. 225-236.
- Cassidy, A. (2014). Communicating the Social Sciences: A Specific Challenge? En: Bucchi, M. and Trench, B. (editores) *Handbook of Public Communication of Science and Technology* (2.ª ed.) Londres, Routledge, Ch. 14, p. 186-197.
- Diamond, J. (1987). Soft sciences are often harder than hard sciences. *Discover*, vol. August, p. 34-39.
- Dittus, R.B. (2005). La opinión pública y los imaginarios sociales: hacia una redefinición de la espiral del silencio. *Athenea Digital*, vol. 7, 61-76.
- Dunwoody, S. (1984). Mass media coverage of the Social Sciences: Some new answers to old questions. *Ecquid Novi: African Journalism Studies*, vol. 5 (2), 83-92.
- Fara, P. (2014). ¿Qué es la ciencia? Confusiones de una historiadora. *Métode Science Studies Journal*, vol. 84, 45-49.
- Fenton, N.; Bryman, A.; Deacon, D. (1998). *Mediating social science*. Londres: Sage Publications.
- Francescutti, L.P. (2010). *La información científica en los telediarios españoles*. Barcelona, Cuadernos de la Fundación Dr. Antonio Esteve nº 21.
- García Sierra, P. (1999). [169] Ciencia (acepciones de). *Diccionario filosófico. Manual de materialismo filosófico, una introducción analítica*. Edición digital Proyecto Filosofía, <http://www.filosofia.org/filomat/df169.htm> [19-04-2015].
- Gil Antón, M. (2004). ¿Ciencias duras y ciencias blandas? Una falsa dicotomía. *Contaduría y Administración*, vol. 213, 151-164.
- González Alcaide, G.; Gómez Ferri, J. (2014). La colaboración científica: principales líneas de investigación y retos de futuro. *Revista Española de Documentación Científica*, 37 (4): e062. <http://dx.doi.org/10.3989/redc.2014.4.1186>

- Hackmann, H.; St. Clair, A.L. (2012). *Transformative Cornerstones of Social Science Research for Global Change*. Paris, Report of the International Social Science Council.
- Haslam, C.; Bryman, A. (2003a). Introduction. En: Cheryl H. and Bryman, A. (editores) *Social Scientists Meet the Media*. Londres y Nueva York, Routledge-Taylor & Francis (1994), p. 1-15.
- Haslam, C.; Bryman, A. (2003b). Social scientists and the media: an overview. En: Cheryl H.; Bryman, A. (editores) *Social Scientists Meet the Media*. Londres y Nueva York, Routledge-Taylor & Francis (1994), p. 186-211.
- Hedges, L.V. (1987). How Hard Is Hard Science, How Soft Is Soft Science? The Empirical Cumulativeness of Research. *American Psychologist*, vol. 42 (2), 443-455. <http://dx.doi.org/10.1037/0003-066X.42.5.443>
- Hijmans, E.; Pleijter, A.; Wester, F. (2003). Covering Scientific Research in Dutch Newspapers. *Science Communication*, vol. 25 (2), 153-176. <http://dx.doi.org/10.1177/1075547003259559>
- Kreimer, P. (2010). ¿Dos culturas o múltiples culturas? Ciencias duras, ciencias blandas y "science studies". *Medicina*, vol. 70 (5), 475-478.
- Meneses Fernández, M.D. (2004). *Aspectos de periodismo especializado: ciencia, cultura y canariedad en la prensa insular canaria, años 1975 y 1982*. La Laguna, Servicio de Publicaciones de la Universidad de La Laguna.
- Moreno Castro, C. (2010). La construcción periodística de la ciencia a través de los medios de comunicación social: hacia una taxonomía de la difusión del conocimiento científico. *Artefactos*, vol. 3 (1), 109-130.
- Moreno Castro, C.; Gil Pérez, A. (2014). ¿Periodismo diletante o ciencia mediática? La metamorfosis del artículo científico en noticia de prensa. En: Barrio Alonso, C. y Cáceres Gómez, S. (editores) *Fronteras de la ciencia. Dilemas*. Madrid, Biblioteca Nueva-OEI, p. 48-58.
- Olmos-Peñuela, J.; Castro-Martínez, E. (2014a). ¿Cómo interaccionan los investigadores de Humanidades y Ciencias Sociales del CSIC con otros agentes sociales fuera de los cauces institucionales?. *Revista Española de Documentación Científica*, 37(4): e072. <http://dx.doi.org/10.3989/redc.2014.4.1165>
- Olmos-Peñuela, J.; Castro-Martínez, E.; Fernández-Esquinas, M. (2014b). Diferencias entre áreas científicas en las prácticas de divulgación de la investigación: un estudio empírico en el CSIC. *Revista Española de Documentación Científica*, 37(2):e040. <http://dx.doi.org/10.3989/redc.2014.2.1096>
- Pena de Oliveira, F. (2006). *Teoría del Periodismo*. Sevilla, Comunicación Social, Ediciones y Publicaciones, 241 p.
- Peters, H.P.; Brossard, D.; de Cheveigné, S.; Dunwoody, S.; Kallfass, M.; Miller, S.; Tsuchida, S. (2008). Interactions with the Mass Media. *Science*, vol. 321, 204-205. <http://dx.doi.org/10.1126/science.1157780>
- Peters, H.P.; Spangenberg, A.; Lo, Y.-Y. (2012). Variations of scientist-journalist interactions across academic fields: Results of a survey of 1600 German researchers from the humanities, social sciences and hard sciences. En: Bucchi, M.; Trench, B. (editores) *12th International Conference Public Communication of Science Quality, Honesty and Beauty in Science and Technology Communication* PCST, Florence, Italy, 18-20 April 2012), Vincenza, Observa Science in Society, p. 257-263.
- Peters, H.P. (2013). Gap between science and media revisited: Scientists as public communicators. *PNAS*, vol. 110 (3), 14102-14109. <http://dx.doi.org/10.1073/pnas.1212745110>
- Peters, H.P.; Dunwoody, S.; Allgaier, J.; Lo, Y.-Y.; Brossard, D. (2014). *Public communication of science 2.0. EMBO reports*, vol. July, 1-4. <http://dx.doi.org/10.15252/embr.201438979>
- Pouliot, C. (2009). Using the Deficit Model, Public Debate Model and Co-production of Knowledge Models to Interpret Points of View of Students Concerning Citizens' Participation in Socioscientific Issues. *International Journal of Environmental & Science Education*, vol. 4 (1), 49-73.
- Priest, S. (2014). ¿Qué tienen de científico las Ciencias Sociales? La complejidad de medir el comportamiento humano. *Métode Science Studies Journal*, vol. 84, 57-63.
- Ruiz-Castell, P. (2014). La ciencia: percepción e inestabilidad. ¿Puede haber un único relato? *Métode Science Studies Journal*, vol. 84, 65-71.
- Schmierbach, M. (2005). Method Matters: The Influence of Methodology on Journalists' Assessments of Social Science Research. *Science Communication*, vol. 26 (3), 269-287. <http://dx.doi.org/10.1177/1075547004273025>
- Summ, A.; Volpers, A.-M. (2015). What's science? Where's science? Science journalism in German print media. *Public Understanding of Science*, 1-16. <http://dx.doi.org/10.1177/0963662515583419>
- Unesco (2013). *Rapport mondial sur les sciences sociales 2013. Changements environnementaux globaux*. Conseil International des Sciences Sociales, París, Éditions Unesco.
- Weigel, R.H.; Pappas, J.P. (1981). Social science and the press: A case study and its implications. *American Psychologist*, vol. 36 (5), 480-487. <http://dx.doi.org/10.1037/0003-066X.36.5.480>
- Weiss, C.H.; Singer, E. (1988). *Reporting of Social Science in the National Media*. Nueva York: Russell Sage Foundation, 299 p.

Whitelegg, E.; Holliman, R.; Allgaier, J.; Scalon, E.; Hodson, B. (2009). Invisible witnesses? How scientists, technologists, engineers and mathematicians are represented on UK television. *'Changing research landscapes to make the most of human potential 10 years of EU activities in Women and Science and Beyond'*, EU Commission, Prague, 14-15 May.

Documentos

AIMC (2001). EGM: Año Móvil febrero-noviembre 2001. Madrid, AIMC.

AIMC (2014). EGM: Año Móvil febrero-noviembre 2014. Madrid, AIMC.

CICYT-Comisión Interministerial de Ciencia y Tecnología (2000). *Plan Nacional de Investigación, Desarrollo e Innovación Tecnológica 2000-2003*. Aprobado por el Consejo de Ministros el 12 de noviembre de 1999. Presidencia del Gobierno, Oficina de Ciencia y Tecnología. Madrid, Ministerio de la Presidencia Secretaría de Estado de la Comunicación.

INE (2015). Estadística sobre Actividades de I+D, INEbase y Ciencia y Tecnología. Instituto Nacional de Estadística. Disponible en: <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft14%2Fp057&file=inebase&L=0> / <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft14%2Fp057%2Fa2013%2F&file=pcaxis&L=0> [consultada el 7 de julio de 2015].

MEC-Ministerio de Educación y Ciencia (2005). *El Plan Nacional I+D+i en cifras*. Secretaría General de Política Científica y Tecnológica, Subdirección General de Coordinación del Plan Nacional de I+D+i. Madrid, Secretaría General Técnica, Subdirección General de Información y Publicaciones.

CICYT-Comisión Interministerial de Ciencia y Tecnología (2007). *Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2008-2011*. Madrid, Fundación Española para la Ciencia y la Tecnología (FECYT).

ESTUDIOS / RESEARCH STUDIES

Redes de conectividad entre empresas tecnológicas a través de un análisis métrico longitudinal de menciones de usuario en Twitter

David Azorín-Richarte*, Enrique Orduna-Malea**, José-Antonio Ontalba-Ruipérez**

* SEO&SEM Valencia. Valencia (España). Correo-e: daazric@gmail.com.

** Universitat Politècnica de València (UPV). Valencia (España).

Correos-e: daazric@gmail.com, enorma@upv.es, joonrui@upv.es

Recibido: 19-06-2015; 2ª versión: 10-12-2015; Aceptado: 15-12-2015.

Cómo citar este artículo / Citation: Azorín-Richarte, D.; Orduna-Malea, E.; Ontalba-Ruipérez, J. A. (2016). Redes de conectividad entre empresas tecnológicas a través de un análisis métrico longitudinal de menciones de usuario en Twitter. *Revista Española de Documentación Científica*, 39(3): e140. doi: <http://dx.doi.org/10.3989/redc.2016.3.1316>

Resumen: El objetivo principal de este trabajo es identificar y describir (mediante un análisis cibernético de menciones de usuario) la intensidad y evolución de las relaciones establecidas entre compañías de un determinado sector industrial (tecnología) a través de sus correspondientes cuentas corporativas de *Twitter*, con el propósito de comprobar el valor que un análisis métrico de estas características puede tener a la hora de determinar la conectividad entre dichas compañías. Para ello se han contabilizado las menciones, tanto directas (MT) como *ReTweets* (RT), entre las cuentas de *Twitter* de una muestra de 50 compañías internacionales durante un período de seis meses. Los resultados indican que el grado de interacción entre las 50 empresas tecnológicas es débil (tanto si contamos el número de conexiones establecidas como si cuantificamos la intensidad de estas conexiones), estable, concentrado en unas pocas relaciones específicas y de carácter marcadamente asimétrico. Se concluye que, dada la baja interactividad detectada, las cuentas corporativas de las empresas tecnológicas en *Twitter* no son suficientes para analizar desde un punto de vista métrico la conectividad web establecida entre éstas, aunque sí útiles para conocer las políticas de comunicación oficiales entre ellas.

Palabras clave: Menciones web; indicadores de red; análisis longitudinal; *Twitter*; *Topsy*; *Google Finance*; compañías tecnológicas; Estados Unidos.

Connectivity networks between technological companies through a longitudinal metric analysis of user mentions in Twitter

Abstract: The main objective of this work is to identify and describe (through a cybermetric analysis of user mentions) the intensity and evolution of the relationships between companies in a particular industry (technology) through their corresponding corporate Twitter accounts, with the purpose of checking the value that a metric analysis like this may offer for determining the web connectivity between these companies. For this purpose, we have quantified the number of mentions, both direct (MT) and *ReTweets* (RT), between the Twitter accounts of a sample of 50 international companies during a period of six months. The results indicate that the degree of interaction among the 50 technology companies is weak (whether we count the number of connections established or quantify the strength of these connections), stable, concentrated in a few specific relations, and sharply asymmetrical. Given the low interactivity found, we conclude that the corporate Twitter accounts of technology companies are not sufficient for analyzing the web connectivity established between them from a metric viewpoint, although they are useful for understanding the official communication policies between them.

Keywords: Web mentions; network metrics; longitudinal analysis; *Twitter*; *Topsy*; *Google finance*; technology companies; United States.

Copyright: © 2016 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Attribution (CC BY) España 3.0.

1. INTRODUCCIÓN

Desde la aparición de *Twitter* en el año 2006 de las manos de Evan Williams, Biz Stone y Jack Dorsey (Carlson, 2011), el número de usuarios de esta red social ha aumentado de manera ininterrumpida hasta llegar a los más de 640 millones de usuarios registrados actualmente (marzo de 2015), de los que el 44.7% (289 millones) son usuarios activos mensuales, con un flujo de aproximadamente 58 millones de *Tweets* diarios¹. Estos datos han llevado a que la sede web de esta plataforma (<*twitter.com*>) ocupe el octavo lugar mundial en *traffic rank*, según los datos de *Alexa*². Esta expansión de *Twitter* ha generado un inmenso depósito global de *Tweets* (más de 600 mil millones en la actualidad), cuyo análisis supone un auténtico desafío, especialmente para las ciencias de la información. Prueba de ello es la iniciativa de la *Library of Congress* de crear un futuro archivo de *Tweets* para fomentar el análisis científico de sus contenidos³.

El interés de *Twitter* como fuente de análisis de contenidos se plasma en el creciente número de investigaciones basadas exclusivamente en la recolección y análisis de *Tweets*. En ese sentido, Williams y otros (2013) analizan un corpus de 500 artículos (publicados entre 2007 y 2011) en los que *Twitter* constituye la fuente principal de datos. Los resultados indican que el análisis de contenido de los *Tweets* (61%) y el estudio de usuarios (21%) conforman los principales objetos de estudio. Posteriormente, Zimmer y Proferes (2014) realizan un análisis sistemático de otro corpus formado por 380 textos (tanto artículos como tesis doctorales) publicados entre 2007 y 2012 en distintas áreas de conocimiento. Los autores clasifican los textos en función de la técnica de análisis (identifican hasta un total de nueve), concluyendo que el método más empleado es el análisis de contenidos (61.25%), seguido del análisis de redes (20.94%) y del análisis de sentimiento (16.49%), mientras que las disciplinas científicas que han abordado estos estudios son fundamentalmente las Ciencias de la computación (37.95%), Ciencias de la información (21.46%) y Ciencias de la comunicación (13.61%), aunque en áreas como la Economía y Negocios (juntas suponen el 7.85%) se constata un interés creciente.

Uno de los factores que más afecta -de manera global- a las distintas técnicas de análisis en *Twitter* es el tipo de usuario³, pues de ello dependen entre otros aspectos el grado de adopción del canal social, el tipo de contenidos que se van a generar o el mayor o menor grado de relación que se va a establecer con otros usuarios. Según un estudio realizado en marzo de 2014 por *Brandchats*, basado en el análisis de 569.100 menciones en

Twitter, existen doce tipos de usuarios (Huevo; Merodeador; Concursante; *ReTweet*; Explorador; Robot; *Celebrity*; #sigueme; *Networker*; Medios; Empresarial; y Fiel). De todos ellos, el tipo "empresarial", que corresponde con las cuentas asociadas a empresas y negocios, destaca especialmente por aportar un volumen de menciones relativamente bajo (2% de las menciones del estudio reseñado anteriormente), pero por generar un impacto importante en el resto de usuarios, especialmente en los catalogados como "fieles".

Twitter y la empresa

Soat (2010), tras analizar una muestra de 400 empresas, identifica los siguientes objetivos principales en el uso corporativo de *Twitter*: generar transmisión boca en boca (*mouth-to-mouth*), incrementar la fidelidad de los usuarios, mejorar el conocimiento de productos o marcas, proporcionar nuevas ideas a la empresa y mejorar la calidad del servicio al cliente. No obstante, la consecución de estos objetivos no resulta sencilla. Kietzmann y otros (2011) sugieren la existencia de cuatro aspectos básicos que deberían guiar a las compañías en la correcta utilización de estos canales:

- Cognize*: comprender verdaderamente sus funcionalidades (por ejemplo, para identificar influenciadores, analizar la competencia, etc.);
- Congruity*: construir estrategias adecuadas y coherentes a las características de cada canal;
- Curate*: gestionar las interacciones (y los contenidos) que se comparten, lo que implica decidir cuándo se debe intervenir y quién debe hacerlo;
- Chase*: buscar los flujos de información que puedan afectar a la posición actual o futura de la empresa en el mercado.

Dadas las características del *microblogging* en general, y de *Twitter* en particular, Jansen y otros (2009) sugieren que estos canales sean considerados por las compañías en sus estrategias comerciales. Este hecho se refleja en la realización de distintos estudios orientados al uso comercial de *Twitter* desde las áreas de la comunicación corporativa (Stieglitz y Krüger, 2011; Carim y Warwick, 2013) y marketing (Bulearca y Bulearca, 2010), donde la comprensión de las distintas posibles interacciones entre clientes y compañías es fundamental.

Desde el punto de vista de los clientes, los *Tweets* hacia (o sobre) empresas y marcas pueden ser categorizados en cuatro grandes bloques, en función de su propósito general (Jansen y otros, 2009): sentimiento (expresión positiva o negativa acerca

de una marca); búsqueda de información (consulta para conocer información acerca de una marca); suministro de información (ofrecer información sobre una marca); y comentario (uso de una marca en un *Tweet*, sin ser el objetivo principal de éste).

Desde el punto de vista de las compañías, se diferencian cuatro canales utilizados a la hora de difundir contenidos a través de *Twitter* (Paniagua y Sapena, 2014): responsabilidad social corporativa, marketing, *networking* corporativo y preferencias reveladas de los clientes, existiendo amplias diferencias en las estrategias seguidas en cada canal. Por ejemplo, Swani y otros (2013), tras analizar 3.982 *Tweets* enviados por un total de 277 empresas situadas en *Fortune 500*, descubren diferencias significativas en las estrategias de marketing entre servicios y productos, que se amplían cuando los destinatarios de los *Tweets* son consumidores o empresas. De hecho, las compañías utilizan *Twitter* no sólo para promocionar productos, marcas o servicios entre los posibles clientes sino para difundir información a los propios accionistas (Romero y otros, 2011; Rybalko y Seltzer, 2010) o a otras empresas.

Sin embargo, el diferente grado de adopción de *Twitter* entre las compañías dificulta las tareas de análisis e incide en el tipo de contenidos difundidos, así como en el tipo e intensidad de las relaciones con otros usuarios. Wamba y Carter (2013), tras realizar una encuesta a 453 pequeñas y medianas empresas en cuatro países (Estados Unidos, Reino Unido, Australia e India), descubren que factores como la localización geográfica, el grado de innovación de la empresa o incluso la edad del *manager* de la empresa impactan en el grado de adopción. Otros factores, como las políticas de uso de canales sociales o el sector industrial, son igualmente claves para comprender el uso de *Twitter* por parte de las compañías.

La publicación y compartición de ciertos tipos de contenidos por parte de las empresas en *Twitter* generan un impacto no sólo en los clientes o en la competencia sino en el propio mercado. Bollen y otros (2011), tras analizar más de nueve millones de *Tweets* publicados por más de dos millones de usuarios en 2008, observan que la información extraída de los *Tweets* mediante análisis de sentimiento permite mejorar sustancialmente los modelos de predicción de cambios en la cotización de las empresas en el índice *Dow Jones Industrial Average*. Paniagua y Sapena (2014) constatan igualmente que los canales de las empresas en *Facebook* y *Twitter* tienen un impacto significativo en los precios de las acciones de las empresas que cotizan en bolsa (*Ibex 35* y *Nasdaq*), aunque la cantidad de seguidores que se precisa para ello exige un umbral mínimo (que en *Twitter* se estima entre 4.141 y 4.316 usuarios).

La repercusión de ciertos *Tweets* con noticias de o sobre empresas (especialmente en aquéllas que cotizan en bolsa), unido a la posibilidad de que puedan publicarse noticias erróneas o falsas, ha llevado a los expertos a discutir si ciertas noticias deberían o no ser difundidas en los canales sociales, ante la posibilidad de manipulación del mercado⁵. Por ejemplo, en Estados Unidos, tal y como indica el experto Enrique Dans⁶, la comunicación de noticias susceptibles de afectar a la cotización de una compañía en el mercado debía llevarse a cabo únicamente a través de canales específicos, tal y como exigía la *U.S. Securities and Exchange Commission* (SEC). En 2008 se actualizó la normativa para permitir el uso de sitios web corporativos oficiales, y se volvió a actualizar en 2013 para incluir también el uso de cuentas oficiales de medios sociales⁷. Como consecuencia, las compañías han desarrollado distintas políticas de comunicación en las redes sociales, tanto en sus distintas cuentas oficiales como especialmente en las cuentas pertenecientes a sus empleados, hecho que influye claramente en el tipo de presencia e impacto de estas empresas en *Twitter*.

Aparte de las posibles políticas de difusión de contenidos, el sector industrial influye igualmente en el grado de adopción de *Twitter* y, por tanto, de su utilización. Case y King (2010) estudian el grado de implementación y uso de redes sociales en las compañías listadas en *Fortune 200* (edición 2009) y descubren que sólo el 64.5% de las empresas tiene cuenta oficial de *Twitter*. Este porcentaje baja al 54% si solamente se consideran las 50 primeras empresas (Case y King, 2011). El 79% de las compañías con una cuenta activa utilizan *Twitter* principalmente para la distribución de noticias y en menor medida para el marketing, promociones, servicios de usuario o recursos humanos, aunque se detectan diferencias por sector. En concreto, las compañías de los sectores computación/TIC presentan grados de adopción y uso superiores al resto (95% de las 19 compañías de este sector incluidas en *Fortune 200*), porcentaje igualmente alto para las empresas de telecomunicaciones (88% de las 8 empresas incluidas en este ranking), que contrastan con el 42% del sector de la banca (n=12) o el 64% del sector de productos de salud (n=22).

El mayor uso de *Twitter* por las empresas de carácter tecnológico no es de extrañar. Gentle (2009) describe los usos más comunes de *Twitter* por estas empresas, entre los que destaca la distribución de PDFs con la actualización de manuales de instrucción, el aviso a los usuarios sobre la actualización de aplicaciones o el seguimiento de las consultas de usuarios. Los *Tweets* tienen igualmente potencial como medios para difundir consejos, trucos

y enlaces de ayuda online. Shu (2015) tipifica los *Tweets* difundidos por empresas tecnológicas hacia los usuarios en cinco grandes categorías (comunicación corporativa; marketing; comunicación de emergencias; comunicación pedagógica; y comunicación técnica), que ejemplifican la adecuación de *Twitter* en este sector.

Conectividad entre empresas

El grado de adopción y los distintos usos de *Twitter* por parte de las compañías determinan finalmente las redes sociales susceptibles de ser creadas entre usuarios, que pueden distinguirse: a) en función del tipo de relación (*Follower/Followed*, mención de *hashtag* o una mención de usuario); y b) de la fuerza de esta relación (cantidad y calidad de seguidores e interacciones) (Huberman y otros, 2008).

El estudio de las relaciones establecidas entre usuarios (principalmente mediante técnicas provenientes tanto del análisis de redes sociales como de la cibermetría) supone una importante fuente de información para conocer las relaciones informales que hayan podido ser establecidas (o que lo serán con cierta probabilidad) entre empresas. Esto permitiría, entre otras cosas, conocer la mayor o menor influencia general de una empresa dentro de su sector en un canal de comunicación masivo como es *Twitter*, la emergencia o decadencia de ciertos usuarios en el tiempo, el tipo de relaciones (más o menos intensas; más o menos superficiales; más o menos amistosas) de una empresa con su competencia directa o el papel mediador de alguna compañía que pueda servir de vínculo entre diversos grupos establecidos. Siendo todos estos datos de gran valor para la toma de decisiones estratégicas (especialmente en el desarrollo de políticas de comunicación corporativa, gestión de crisis, detección de usuarios clave con quienes conectar para generar tráfico sus sitios web, etc.), dentro de una compañía.

Si nos centramos en las relaciones establecidas a través de las menciones de usuario, las razones por las que una compañía pueda querer mencionar a otra en un *Tweet* pueden ser diversas, tanto neutras (participar en una conversación donde pueda aparecer mencionada otra empresa), como positivas (estrechar lazos, crear comunidad, ayudar a difundir ciertos productos, noticias o eventos, especialmente en los que participen conjuntamente, fomentar debates o conversaciones de valor añadido a la comunidad, etc.) o directamente negativas (criticar un producto, servicio, comentario, etc.).

Finalmente, la forma de mencionar podrá indicar y modular la naturaleza de la relación; así, un *Tweet* directo (publicación de un *tweet* original con

mención directa a otro usuario) reflejará una relación activa y manifiesta a generar una conversación, mientras que un *ReTweet* (acción por la que se publica un *Tweet* ya creado previamente por otro usuario) reflejará una relación más pasiva y vinculada a la mera redifusión de contenidos, bien creados directamente por otra compañía o en los que aparecen simplemente mencionados.

Dado que *Twitter* es una red de interconexiones (tanto de usuarios como de temas), las características estructurales son fundamentales para conocer la posición e influencia de cada nodo, por lo que el uso de técnicas de Análisis de Redes Sociales (SNA) resulta fundamental (Barabási, 2014). La aplicación de estas técnicas a los estudios métricos de la web (Barabasi y Albert, 1999; Adamic y Huberman, 2001) y a la cibermetría son comunes (Ortega y Aguillo, 2011; Ortega y Aguillo, 2013). Además, el estudio de la evolución de las conexiones establecidas en una determinada red a través de técnicas SNA puede ayudar a conocer asimismo los procesos de creación, difusión y destrucción de agrupaciones, así como la realización de estudios de causalidad con ciertos fenómenos sociales.

En función de la estructura formada por las relaciones entre usuarios, se han identificado hasta seis tipos de redes en *Twitter* (Smith y otros, 2014): divididas (multitudes polarizadas); unificadas (multitudes apretadas); fragmentadas (agrupación de marcas); agrupadas (agrupación de comunidades); *In-hub & Spoke* (redes de difusión); y *Out-hub & Spoke* (redes de apoyo).

Estudios cibernéticos aplicados a la relación web entre compañías

Con anterioridad a la existencia de *Twitter*, las relaciones establecidas entre las empresas en un entorno online se habían realizado mediante técnicas cibernéticas, especialmente a través del análisis de enlaces (Thelwall, 2004; Orduña-Malea y Aguillo, 2014). Aunque esta disciplina se orienta principalmente al análisis de recursos de investigación, a partir de la publicación de los trabajos pioneros de Vaughan (2004) los indicadores web comenzaron a aplicarse de forma más sistemática en entornos empresariales, demostrando así cómo el análisis de enlaces proporcionaba información comercial útil.

A partir de entonces, se desarrolló una limitada pero importante línea de investigación centrada en el análisis de los sitios webs de compañías comerciales (Romero-Frias, 2011), en la que destaca por una parte un área centrada en el estudio de posibles correlaciones entre los indicadores web y ciertas variables financieras y, por otra parte, un

área enfocada en el estudio de las propiedades de ciertos indicadores a la hora de aportar nuevas perspectivas acerca de las relaciones entre empresas de sectores tanto homogéneos como heterogéneos (Vaughan y Wu, 2004), donde destacan los trabajos realizados a través de co-enlaces (Vaughan y You, 2006) y co-menciones (Vaughan y You, 2010).

De forma paralela se desarrollaron numerosos trabajos basados en el testeo de ciertas plataformas que permitían la obtención de ciertos indicadores (visitas, búsquedas y enlaces principalmente) aplicados al rendimiento web de las empresas, como por ejemplo ciertas fuentes de tráfico web como *Alexa*, *Compete* y *Google Trends* (Vaughan y Yang, 2013), motores de búsqueda comerciales como *Altavista*, *MSN*, *Google* o *Yahoo!* (Vaughan y Wu, 2004; Vaughan y You, 2006) o motores de búsqueda específicos como *Google News* o *Google Blogs* (Vaughan y Romero-Frias, 2012), y más recientemente *Topsy* (Vaughan, en prensa).

Unidades de análisis en las relaciones web

La aparición de *Twitter* (entre otras plataformas sociales) permitió, por tanto, expandir la capacidad de analizar las relaciones web entre compañías que proporcionaba la cibermetría clásica (basadas en menciones web entre sus sitios web) gracias a la cuantificación de las menciones explícitas entre las correspondientes cuentas de usuario creadas en esta plataforma de *microblogging*.

Ambos entornos (la Web en general y *Twitter* en particular) presentan, sin embargo, un mismo problema relacionado con la elección de la unidad de análisis para cuantificar las menciones, debido a la presencia desagregada y dispersa de las empresas en la Web (Orduña-Malea y otros, 2015a).

En el caso de una compañía internacional, es común que ésta tenga sedes en diferentes países, productos y servicios concretos, marcas, información para los inversores, uno o varios blogs (corporativos o para productos específicos), etc. Cada uno de estos elementos puede disponer de un sitio web completamente independiente (denominado Additional Web Domain o AWD). Aunque la mayoría de estos AWDs poseen una visibilidad mínima, en algunos casos pueden llegar a tener un impacto web mucho mayor que el sitio web corporativo (pensemos en *Movistar* y *Telefónica*, por ejemplo).

Esta dispersión se replica en *Twitter*, donde una compañía puede disponer de multitud de cuentas creadas, tanto generales como para determinados productos (especialmente en las compañías tecnológicas), actividades, eventos, delegaciones, car-

gos, etc., con la problemática añadida de que las políticas de creación y difusión de contenidos pueden ser bien diferentes en cada perfil. Por ello, la elección de las unidades de análisis (en el caso de no poder considerar todas las cuentas) es crucial para poder contextualizar las relaciones explícitas entre empresas.

Entre todas las posibles cuentas, la corporativa tiene una significación especial para la compañía, pues es la elegida como canal de comunicación oficial. Por tanto, los contenidos que se generen y las relaciones que se establezcan desde esta cuenta con otras compañías tendrán a priori una significación estratégica especial, aunque las cuentas de un producto concreto o de un directivo mediático puedan disponer de un mayor número de seguidores (y/u otras conexiones más intensas).

A pesar del interés que supone conocer y caracterizar las relaciones web existentes entre empresas, y más en una plataforma de difusión masiva de contenidos como *Twitter*, este tipo de análisis es abordado usualmente desde los principios del marketing y la comunicación corporativa, y donde los métodos utilizados suelen quedar ocultos por secreto industrial. Por ello, la aplicación y testeo de metodologías procedentes de la cibermetría y de las ciencias de la información en este ámbito resultan, por su escasez, necesarias.

Objetivos

El objetivo principal de este trabajo es identificar y describir (mediante un análisis ciber métrico de menciones de usuario) la intensidad y evolución de las relaciones establecidas entre compañías de un determinado sector industrial (tecnología) a través de sus correspondientes cuentas corporativas de *Twitter*, con el propósito de comprobar el valor que un análisis métrico de estas características puede tener a la hora de determinar la conectividad entre dichas compañías.

Con el fin de caracterizar la intensidad y evolución de las relaciones entre compañías, se plantean las siguientes tareas específicas:

- Medir el grado de interacción (alto o bajo) entre las empresas de la muestra analizada a partir del número de menciones (en adelante, MT) y *ReTweets* (en adelante, RT) que se envían entre ellas.
- Cuantificar la evolución de las interacciones a lo largo del tiempo.
- Comprobar la existencia de simetría o asimetría entre esas interacciones, teniendo en cuenta que las menciones son unidireccionales.

- Verificar la existencia de *clusters* de empresas, y su posible evolución a lo largo del tiempo.

2. METODOLOGÍA

A continuación se detalla el procedimiento seguido para cumplir con los objetivos planteados anteriormente.

Selección y obtención de la muestra

Dado el mayor grado de adopción de *Twitter* por parte de las empresas tecnológicas, se plantea analizar una muestra exclusiva de compañías incluidas en este sector industrial. Para ello, se seleccionaron las 50 empresas tecnológicas con mayores ingresos netos anuales (*Annual Net Income*), según los datos disponibles en el momento de obtención de la muestra (octubre de 2014), a partir del listado que ofrece *Google Finance* para las empresas de este sector⁸.

Google Finance es un producto de información financiera lanzado en marzo de 2006⁹, que cubre cerca de 24 mercados bursátiles en 15 países. Entre sus principales funcionalidades se encuentra la aportación de datos económicos y financieros de empresas por sectores industriales a través de cerca de 50 variables financieras. Aunque existen otros productos similares (*Yahoo Finance*, *Morningstar*, etc.), se eligió *Google Finance* simplemente por su facilidad de uso para obtener el listado de compañías, aunque este listado se podría obtener de cualquier otra fuente similar, pues los ingresos netos anuales oficiales no dependen de la fuente, sino de la información proporcionada por las propias empresas.

Para cada una de estas compañías se localizó manualmente la existencia de una cuenta oficial de *Twitter*. Con el propósito de minimizar los posibles sesgos geográficos en la adopción de esta red social, se contemplaron únicamente las cuentas oficiales correspondientes a las delegaciones de Estados Unidos, país con el mayor porcentaje de usuarios activos en *Twitter* (24.3%)¹⁰, a fecha de 2013 (últimos datos disponibles en el momento de la obtención de la muestra). Igualmente, se contemplaron únicamente las cuentas oficiales corporativas, dejando fuera de este estudio a las diferentes cuentas de productos específicos, servicios, etc.

Del listado inicial de 50 empresas tecnológicas, se descartaron tres compañías (*Taiwan Semiconductor*, *Baidu* y *NetEase*), que incumplían el criterio de disponer de una cuenta oficial en una delegación estadounidense. Para suplir esta carencia,

se escogieron las tres empresas siguientes en el ranking ofrecido por *Google Finance*. El caso de *Apple* merece una atención especial; esta empresa queda situada en primera posición en función de los ingresos anuales en el momento de realización del trabajo, pero no disponía de una cuenta oficial corporativa. Dada la importancia de esta empresa en el sector tecnológico mundial se decidió, a modo de prueba de control, incluir una cuenta no oficial pero con un alto grado de seguidores (@AppleOfficiall), con el fin de determinar su grado de interactividad con el resto de cuentas oficiales de otras compañías. En la Tabla I se muestran las empresas que conforman la muestra final, junto a su cuenta en *Twitter*, número de seguidores (a fecha de octubre de 2014; extraídos de *Twitter*), y número de *Tweets* totales publicados mensualmente (de abril a agosto de 2014).

Selección y obtención de indicadores

Para cada una de las 50 cuentas de *Twitter* recopiladas, se procedió a contabilizar la cantidad de menciones, tanto MT como RT, entre todas ellas. Para ello se utilizó un método indirecto (Del-Fresno-García, 2014) a través del buscador especializado *Topsy*¹¹, que permite no sólo acceder al archivo completo de *Tweets* de *Twitter*, sino además realizar búsquedas avanzadas mediante algunos comandos. El uso de *Topsy* en estudios métricos de la Web ya ha sido utilizado y probado con anterioridad (Vaughan, en prensa; Orduña-Malea y otros, 2015b).

En este caso, para obtener las menciones MT y RT entre las diferentes cuentas se ha utilizado el comando "FROM:", con el que se obtiene el número de *Tweets* (en cualquier idioma) difundidos por una cuenta 1 en la que se mencionara a una cuenta 2, de la siguiente forma: <@empresa2 FROM:@empresa1>. En la figura 1 se ejemplifica este proceso mediante la consulta <@dell FROM:@intel>, que devuelve el número de *Tweets*, en un determinado período de tiempo, en los que la empresa *Intel* ha mencionado (mediante MT o RT) a la empresa *Dell*, a través de sus respectivas cuentas oficiales de *Twitter*.

Dado que el propósito es conocer tanto la intensidad como la evolución del grado de interconectividad entre las cuentas de *Twitter*, se diseñó un proceso de medida longitudinal basado en la ejecución de seis tomas mensuales (desde marzo hasta agosto de 2014). En cada toma se contabilizó el número de menciones desde cada cuenta al resto (un total de 2.450 posibles combinaciones), distinguiendo si la mención era de tipo MT o RT.

Tabla I. Muestra de empresas tecnológicas y sus cuentas de *Twitter*

EMPRESA	TWITTER	FOLLOWERS	NÚMERO DE TWEETS PUBLICADOS					
			ABRIL	MAYO	JUNIO	JULIO	AGOSTO	TOTAL
<i>Apple</i>	@AppleOfficiall	26.021	7	31	17	35	53	143
<i>Microsoft</i>	@microsoft	4.388.419	38	57	56	114	71	336
<i>IBM</i>	@IBM	94.648	70	66	48	49	47	280
<i>Google</i>	@GoogleAtWork	376.872	-	-	-	-	-	-
<i>Oracle</i>	@Oracle	226.133	103	60	58	65	44	330
<i>Cisco</i>	@Cisco	346.392	82	106	51	45	48	332
<i>Intel</i>	@intel	2.918.623	69	86	91	61	77	384
<i>Qualcomm</i>	@qualcomm	102.517	26	25	27	25	29	132
<i>Hewlett-Packard</i>	@hp	620.935	228	239	202	254	210	1133
<i>SAP AG</i>	@SAP4SME	5.403	-	-	-	-	-	-
<i>EMC</i>	@EMCcorp	56.670	87	182	115	139	137	660
<i>eBay</i>	@eBay	333.178	336	236	218	264	324	1378
<i>Dell</i>	@Dell	305.020	150	284	313	335	319	1401
<i>Canon</i>	@CanonUSA	13.345	26	37	26	26	39	154
<i>Texas Instruments</i>	@TXInstruments	33.406	190	195	192	195	148	920
<i>Corning</i>	@corning	7.595	34	34	46	35	33	182
<i>Seagate</i>	@SEAGATE	45.602	178	164	182	180	199	903
<i>Infosys</i>	@infosys	86.360	82	57	91	78	62	370
<i>Facebook</i>	@facebook	13.600.365	2	1	6	6	3	18
<i>ASML</i>	@ASMLcompany	4.657	24	21	14	19	4	82
<i>Yahoo!</i>	@Yahoo	1.178.132	488	506	472	529	618	2613
<i>Micron</i>	@MicronTech	6.010	1	1	3	0	4	9
<i>Xerox</i>	@XeroxCorp	113.024	229	189	204	202	165	989
<i>Motorola Solutions</i>	@MotoSolutions	4.940	67	86	122	110	157	542
<i>Cognizant</i>	@Cognizant	77.396	286	318	286	154	292	1336
<i>America Online</i>	@AOL	110.313	501	520	1056	4059	4360	10496
<i>SanDisk</i>	@SanDisk	29.364	82	120	130	70	82	484
<i>Western Digital Corp.</i>	@westerndigital	53.359	42	66	62	62	103	335
<i>Wipro</i>	@Wipro	204.795	53	44	44	64	44	249
<i>CA Technologies</i>	@CAInc	82.728	187	262	204	205	200	1058
<i>Ericsson</i>	@ericsson	15.375	78	64	88	98	83	411
<i>Intuit</i>	@Intuit	54.142	69	86	91	61	77	384
<i>Symantec</i>	@symantec	59.841	244	226	234	286	285	1275
<i>Vmware</i>	@VMware	112.020	66	188	110	81	127	572
<i>Analog Devices</i>	@adi_news	12.339	21	41	45	25	26	158
<i>Check Point Software Tech.</i>	@checkpointsw	22.546	25	44	24	24	28	145
<i>Amphenol</i>	@AmphenolLtd	573	28	16	26	25	23	118
<i>Nvidia</i>	@nvidia	422.664	35	33	41	72	103	284
<i>Avago Technologies</i>	@Avagotech	1.057	8	5	8	7	8	36
<i>KLA-Tencor</i>	@KLATencor_Info	650	22	20	18	24	24	108
<i>Garmin</i>	@Garmin	56.817	44	31	53	46	41	215
<i>Tyco International</i>	@Tyco_Suzuki	16.087	0	0	0	0	0	0

<i>Leidos Holdings</i>	@leidosinc	702	33	28	26	19	13	119
<i>NetApp</i>	@netapp	102.271	89	90	80	86	85	430
<i>Computer Sciences Corp.</i>	@CSC	4.926	59	46	70	49	44	268
<i>Xilinx</i>	@XilinxInc	9.725	65	63	40	41	45	254
<i>Harris Corporation</i>	@harriscorp	3.052	9	11	26	25	24	95
<i>Maxim Integrated Products</i>	@maxim_ic	7.245	24	25	24	24	24	121
<i>Avnet</i>	@Avnet	2.307	9	7	30	30	29	105
<i>Altera</i>	@alteracorp	8.564	8	19	8	5	8	48

Fuente: conjunto de empresas: *Google finance*; *Followers* y *Tweets* mensuales: *Twitter*.

Las cuentas utilizadas para *Google* y *SAP AG* han dejado de estar activas, por lo que no se ha podido calcular el número de *Tweets* en el período analizado.

Figura 1. Consulta selectiva de menciones entre dos cuentas de *Twitter* a través del buscador *Topsy*

The screenshot shows the Topsy search interface. The search bar contains the query "@dell FROM:@intel". The results are sorted by relevance. The results list several tweets from Intel (@intel) mentioning Dell XPS Ultrabooks, including promotional offers and product reviews.

Los datos obtenidos de cada toma mensual se transformaron manualmente en un fichero .net que fue exportado a la aplicación *Gephi* (v.082), desde donde se extrajeron los datos de centralidad mensuales de la red correspondiente, tanto a nivel de nodo (grado, *closeness*, *betweenness*, coeficiente de *clustering*, *eigenvector*), como a nivel de red (grado medio, diámetro, densidad, longitud media del camino, coeficiente medio de *clustering*). En la tabla II se incluye la definición de cada uno de estos indicadores.

En cada toma de datos se recopiló adicionalmente el número de menciones totales acumuladas, de forma que la diferencia de menciones entre meses proporcionara el número de menciones relativas al último mes (de abril a agosto). Con el fin de contextualizar estos datos, se procedió igualmente a recopilar la producción mensual total de *Tweets* emitida por cada cuenta (igualmente de abril hasta agosto de 2014), que se extrajeron de forma retrospectiva (a fecha de noviembre de 2015) directamente desde la búsqueda avanzada de *Twitter*.

Con estos datos se calculó la tasa de publicación para cada cuenta (número de *Tweets* totales emitidos entre abril y agosto por una determinada cuenta, dividido por el número de *Tweets* recibidos por esa cuenta desde el resto de cuentas durante ese mismo período).

Finalmente, todos los datos obtenidos fueron exportados a una hoja de cálculo para ser tratados estadísticamente. El proceso de análisis fue realizado durante los meses de septiembre y octubre de 2014.

3. RESULTADOS

En el momento de la primera toma de datos (marzo 2014), el número de *Tweets* totales acumulados en los que, desde alguna de las 50 empresas se menciona a cualquiera de las restantes, es de 2.043. Esta cifra asciende hasta los 2.381 en la última toma (agosto 2014), registrándose por tanto un total de 338 *Tweets* durante los seis meses de medida. La evolución del número de *Tweets* enviados puede observarse en la figura 2. Por otro lado, el 85% del número de *Tweets* totales recopilados (2.017) constituyen *MT* mientras que los *RT* suponen solo el 15% de los *Tweets* analizados (364).

La compañía que más *Tweets* acumulados ha enviado al resto de compañías es *Hewlett-Packard* (292 *Tweets*), seguida de *NetApp* (283), aunque si contabilizamos únicamente los meses de la muestra, esta última es la más activa (57

Tweets, por sólo 21 de *Hewlett-Packard*), seguida de *CSC* (47). En el caso de los *Tweets* recibidos acumulados, *Microsoft* se sitúa en primer lugar (323 *Tweets* recibidos), seguida de *VMware* (286) e *Intel* (280), aunque durante los seis meses de medida la más activa es *VMware* (51 *Tweets*), seguida de *Intel* (47).

En cuanto a las compañías con menor interactividad, se identifican doce empresas que no han enviado ningún *Tweet* al resto durante el período de análisis (seis de las cuales sí habían remitido alguno con anterioridad). Por otro lado, se identifican veinte empresas sin ningún *Tweet* recibido en el período (cuatro de las cuales no han recibido nunca un *Tweet* del resto de empresas consideradas). La cuenta no oficial de *Apple* y la de *Tyco* son las únicas que no han enviado ni recibido ningún *Tweet* ni en el período ni en el total histórico. En cualquier caso, pese a algunas ligeras diferencias entre los *Tweets* totales acumulados (a fecha de agosto) y los *Tweets* generados (enviados o recibidos) sólo durante los meses de medida (rango), la correlación entre estas dos variables es muy alta. En el caso de los *Tweets* enviados es de r (Pearson) = 0.82, y en el caso de los *Tweets* recibidos asciende a r (Pearson) = 0.92.

En todo caso, el número de *Tweets* recibidos puede venir influido tanto por la cantidad de *followers* que una empresa pueda tener como por la cantidad de mensajes emitidos. En el caso del número de *followers*, a mayor número de seguidores, la proba-

Tabla II. Indicadores estructurales (a nivel de nodo y de red)

Nivel de nodo	Alcance
Grado	Número de nodos adyacentes a un nodo "x"
Closeness centrality	Frecuencia con la que un nodo "x" aparece en el camino más corto entre pares de nodos en una red.
Betweenness centrality	Distancia media entre un nodo "x" al resto de nodos de la Red
Clustering coefficient	Grado en el que los nodos del vecindario de un nodo "x" están conectados entre sí
Eigenvector centrality	Medida de la importancia de un nodo en la red; se calcula de forma recursiva asignando marcadores relativos a cada nodo en función de sus conexiones, y donde la conexión a un nodo altamente puntuado contribuye más a la propia puntuación de un nodo
Nivel de Red	Alcance
Grado medio	Promedio de los Grados de todos los nodos pertenecientes a una Red
Coefficiente medio de clustering	Promedio de los coeficientes de clustering correspondientes a cada nodo de la Red
Longitud media del camino	Distancia promedia entre todos los pares de nodos de la Red
Diametro	Distancia máxima existente entre un par de nodos
Densidad	Lo cerca que se encuentra una red de estar completa, es decir, de poseer todas las aristas posibles (en cuyo caso, densidad es igual a "1")

Figura 2. Evolución del número de *Tweets* y tipología (Menciones y *ReTweets*)

bilidad de que un determinado mensaje se difunda más (y por tanto una empresa sea más mencionada) puede aumentar. En este caso, se ha calculado la correlación de Spearman (pues la distribución de *followers* es asimétrica) entre el número de *followers* de cada empresa y el número de *Tweets* recibidos durante los meses de medida ($r= 0.67$; $\alpha= 0.01$) y los totales acumulados ($r= 0.64$; $\alpha= 0.01$), verificándose una correlación moderada y significativa.

Respecto a la cantidad de mensajes totales emitidos, a mayor cantidad de *Tweets* enviados aumentan igualmente las posibilidades de recibir una mención. En la Tabla III se muestran las 10 cuentas

con la mejor tasa de publicación (liderada por Facebook, que recibe 1 *Tweet* por cada 1.3 emitidos) y con la peor (siendo Yahoo! la empresa peor situada, con una tasa de 2613). Por otro lado, se detectan 18 empresas que han emitido *Tweets* pero que no han recibido ninguno, por lo que no se puede calcular la tasa de publicación. En la Tabla III se muestra el balance de publicaciones (*Tweets* emitidos menos *Tweets* recibidos) para contextualizar la falta de *Tweets* recibidos. Por ejemplo, AOL emitió 10.496 *Tweets*, y no recibió ninguno del resto de empresas, mientras que Checkpoint tampoco recibió ningún *Tweet*, pero solamente emitió 145.

Tabla III. Tasa de *Tweets* emitidos en función de *Tweets* recibidos

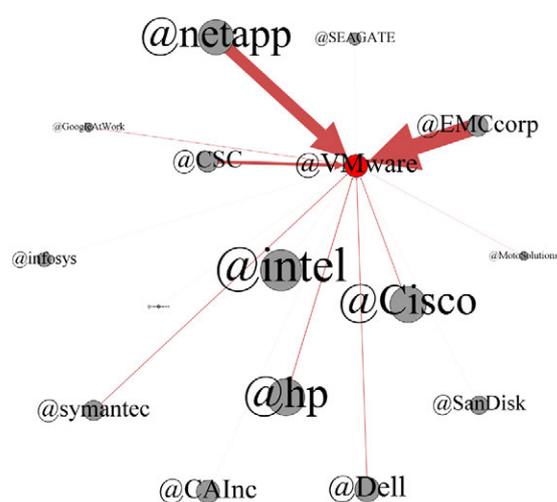
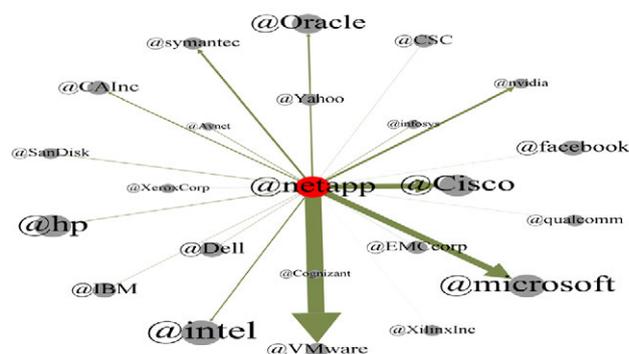
MEJOR TASA DE PUBLICACIÓN		PEOR TASA DE PUBLICACIÓN		SIN TASA DE PUBLICACIÓN	
CUENTA	FACTOR	CUENTA	FACTOR	CUENTA	BALANCE
@facebook	1.3	@corning	182	@AOL	-10496
@MicronTech	3	@symantec	182.1	@Cognizant	-1336
@microsoft	7.8	@Dell	200.1	@Intuit	-782
@intel	8.2	@CSC	268	@MotoSolutions	-542
@Cisco	9.5	@ericsson	411	@infosys	-370
@IBM	10	@eBay	459.3	@westerndigital	-335
@VMware	11.2	@TXInstruments	460	@XilinxInc	-254
@nvidia	14.2	@XeroxCorp	494.5	@Garmin	-215
@Avnet	17.5	@SEAGATE	903	@checkpointsw	-145
@Oracle	20.6	@Yahoo	2613	@AppleOfficiall	-143

En la Tabla IV se muestran las empresas con una mayor actividad tanto en el envío como en la recepción de *Tweets* durante los meses de la muestra (rango). Se ofrecen igualmente, a modo de ilustración, los datos de *Tweets* acumulados (tanto enviados como recibidos) por mes. Adicionalmente, se muestra la red formada para la compañía que más *Tweets* genera (*NetApp*) y más *Tweets* recibe (*VMware*) durante el período de medida.

Las diferencias entre los rankings de empresas por *Tweets* enviados y recibidos indican que las compañías tienen comportamientos diferentes a la hora de enviar o recibir *Tweets*. Por ejemplo, durante los meses de medida, *NetApp* envía hasta 57 *Tweets* al resto de empresas, mientras que sólo recibe 4; algo similar le ocurre a *CSC* (envía 47 *Tweets* y recibe solamente 1). Por el contrario, *VMware* envía solamente 5 *Tweets* pero recibe 51, e *Intel* envía igualmente 5, y recibe hasta 47 *Tweets*.

Tabla IV. Compañías con mayor número de *Tweets* enviados (superior) y recibidos (inferior)

TWEETS ENVIADOS							
EMPRESA	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOSTO	RANGO
<i>NetApp</i>	226	245	256	266	270	283	57
<i>CSC</i>	129	136	138	151	165	176	47
<i>EMC</i>	136	137	154	164	169	180	44
<i>Hewlett-Packard</i>	271	278	282	287	291	292	21
<i>CA</i>	94	97	105	105	107	113	19
<i>Cognizant</i>	20	21	21	26	26	34	14
<i>Symantec</i>	24	24	24	24	34	37	13
<i>KLA-Tencor</i>	12	22	22	22	23	25	13
<i>Dell</i>	49	50	53	57	57	61	12
<i>Microsoft</i>	98	104	107	108	109	109	11
TWEETS RECIBIDOS							
EMPRESA	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOSTO	RANGO
<i>Vmware</i>	235	238	253	267	273	286	51
<i>Intel</i>	233	251	257	264	275	280	47
<i>Microsoft</i>	280	291	299	310	318	323	43
<i>Cisco</i>	95	98	109	115	122	130	35
<i>IBM</i>	65	68	70	78	85	93	28
<i>Nvidia</i>	58	69	76	77	77	78	20
<i>Hewlett-Packard</i>	144	153	156	158	159	162	18
<i>Oracle</i>	118	120	121	124	127	134	16
<i>Facebook</i>	137	136	140	142	144	150	13
<i>EMC</i>	53	55	59	61	61	63	10



Los datos relativos a esta asimetría, a la hora de enviar y recibir *Tweets*, se pueden observar en la Tabla V. En ella se muestran las empresas con una mayor asimetría positiva (más *Tweets* recibidos que enviados) y negativa (más *Tweets* enviados que recibidos), considerando tanto los *Tweets* generados durante los meses de medida como el total de *Tweets* acumulados.

En cualquier caso, estos datos brutos de *Tweets* deben contextualizarse, pues una compañía puede recibir muchos *Tweets* y que la mayoría de ellos provenga de unas pocas fuentes (revelando por tanto conexiones fuertes con ciertas empresas, pero una nula visibilidad respecto al resto), mientras que otras compañías pueden recibir un menor número total de *Tweets* y que la procedencia de éstos esté más repartida. Por ejemplo, de los 286 *Tweets* recibidos por *VMware*, el 45.8% (131) proceden de *EMC*. Por el contrario, de los 292 *Tweets* enviados desde la cuenta de *Hewlett-Packard*, 108 se dirigen a *Intel* (37%) y 102 a *Microsoft* (35%).

En la Tabla VI se muestran las conexiones más fuertes entre las compañías de la muestra al principio (marzo) y al final (agosto) del análisis, lo que permite constatar un crecimiento lento de las co-

nexiones. Destacan especialmente las menciones realizadas desde *EMC* a *VMware*, que crecen desde los 96 *Tweets* en marzo a los 131 en agosto.

Esta distribución desigual de *Tweets* en unos pocos pares de empresas justifica la necesidad de contabilizar el número de empresas a las que se menciona en un *Tweet* (Grado de salida; *outDegree*) y el número de empresas desde las que se recibe un *Tweet* (Grado de entrada; *inDegree*). *Intel* es la empresa que recibe *Tweets* desde un mayor número de compañías (30), seguida de *Microsoft* (29) y *Facebook* (24), mientras que *Netapp* es la empresa que envía *Tweets* a un mayor número de empresas (22), seguida de *Cisco* (20) y *HP* (19). En la Tabla VII se ofrece el Top 10 de las empresas con un mayor grado de entrada y salida.

Con el propósito de visualizar más apropiadamente los vínculos establecidos entre las empresas se ha elaborado una red de relaciones que se muestra en la Tabla VIII, tanto para el mes de marzo (Tabla VIII; izquierda) como agosto (Tabla VIII; derecha). Igualmente se ofrecen los indicadores de red para los nodos más importantes: grado (grado de entrada más grado de salida), *closeness*, *betweenness*, coeficiente de *clustering* y *eigenvector*.

Tabla V. Asimetría en el envío y recepción de *Tweets* (TW): menciones (MT) + *ReTweets* (RT)

EMPRESA	TOTAL ACUMULADO			EMPRESA	TOTAL ACUMULADO		
	TW ENVIADOS	TW RECIBIDOS	ASIMETRÍA POSITIVA		TW ENVIADOS	TW RECIBIDOS	ASIMETRÍA NEGATIVA
<i>VMware</i>	36	286	250	<i>NetApp</i>	283	48	-235
<i>Microsoft</i>	109	323	214	<i>CSC</i>	176	11	-165
<i>Intel Corporation</i>	100	280	180	<i>Hewlett-Packard</i>	292	162	-130
<i>Facebook</i>	2	150	148	<i>EMC</i>	180	63	-117
<i>IBM</i>	3	93	90	<i>CA</i>	113	21	-92
<i>NVIDIA</i>	18	78	60	<i>Analog Devices</i>	124	36	-88
<i>Oracle</i>	79	134	55	<i>Seagate</i>	74	5	-69
<i>Cisco SysTEMS</i>	84	130	46	<i>Wipro</i>	50	10	-40
<i>Xilinx</i>	70	116	46	<i>Infosys</i>	42	12	-30
<i>Dell</i>	61	105	44	<i>Micron Technology</i>	31	5	-26

EMPRESA	TOTAL EN EL PERÍODO			EMPRESA	TOTAL EN EL PERÍODO		
	TW ENVIADOS	TW RECIBIDOS	ASIMETRÍA POSITIVA		TW ENVIADOS	TW RECIBIDOS	ASIMETRÍA NEGATIVA
<i>VMware</i>	5	51	46	<i>NetApp</i>	57	4	-53
<i>Intel Corporation</i>	5	47	42	<i>CSC</i>	47	1	-46
<i>Microsoft</i>	11	43	32	<i>EMC</i>	44	10	-34
<i>Cisco SysTEMS</i>	6	35	29	<i>Cognizant</i>	14	0	-14
<i>IBM</i>	0	28	28	<i>KLA-Tencor</i>	13	0	-13
<i>NVIDIA</i>	1	20	19	<i>CA</i>	19	7	-12
<i>Facebook</i>	0	13	13	<i>Symantec</i>	13	7	-6
<i>Oracle</i>	6	16	10	<i>Intuit</i>	5	0	-5
<i>Avnet</i>	2	6	4	<i>Infosys</i>	5	0	-5
<i>SanDisk</i>	2	4	2	<i>Seagate</i>	6	1	-5

Tabla VI. Intensidad en las relaciones entre compañías tecnológicas en *Twitter* (marzo y agosto 2014)

MARZO 2014			AGOSTO 2014		
SOURCE	TARGET	HITS	SOURCE	TARGET	HITS
@hp	@intel	100	@EMCcorp	@VMware	131
@adi_news	@XilinxInc	100	@hp	@intel	108
@EMCcorp	@VMware	96	@hp	@microsoft	102
@hp	@microsoft	95	@adi_news	@XilinxInc	100
@netapp	@VMware	85	@netapp	@VMware	92
@intel	@hp	52	@intel	@hp	53
@microsoft	@hp	40	@microsoft	@hp	49
@netapp	@microsoft	37	@netapp	@microsoft	49
@hp	@nvidia	32	@netapp	@Cisco	44
@CAInc	@facebook	30	@hp	@nvidia	34

Tabla VII. Empresas con un mayor grado de entrada (*InDegree*) y salida (*OutDegree*) (agosto 2014)

AGOSTO 2014			
COMPAÑÍA	<i>InDegree</i>	COMPAÑÍA	<i>OutDegree</i>
@intel	30	@netapp	22
@microsoft	29	@Cisco	20
@facebook	24	@hp	19
@IBM	22	@CAInc	19
@hp	20	@CSC	17
@Oracle	20	@Oracle	14
@Dell	18	@symantec	14
@Cisco	18	@intel	13
@VMware	15	@EMCcorp	12
@netapp	15	@XilinxInc	11

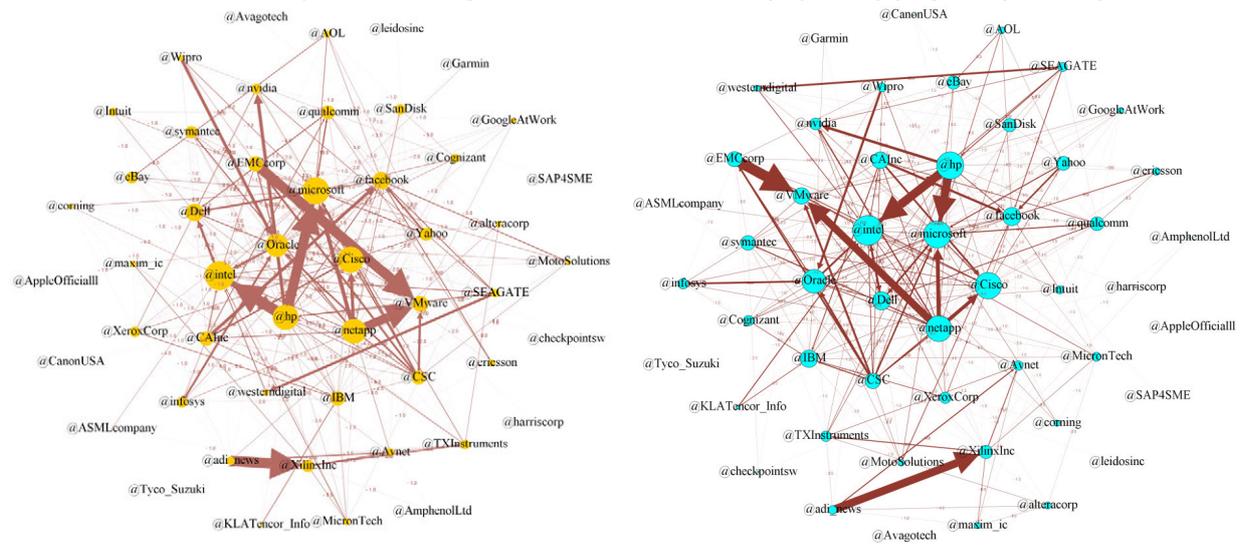
Como se puede observar en la Tabla VIII, la estructura general de la red permanece muy estable durante los meses de medida. El nodo más importante de la red, según su *Eigenvector*, es *Microsoft*, seguida de *Facebook*, *Intel*, *HP* y *Dell*. Estas mismas compañías no sólo permanecen como los nodos más importantes todos los meses, sino que incrementan ligeramente su poder en la Red. Cabe destacar asimismo el papel de *Facebook* que, pese a tener un valor de grado claramente inferior al resto de nodos importantes (25, muy por debajo de los 42 de *Intel*, por ejemplo), logra el segundo puesto en *eigenvector* gracias a disponer de un valor muy alto de grado de entrada (24) y *closeness* (2.98), y muy bajo de *betweenness* (30.36).

La evolución de los indicadores de red se muestra en la Tabla IX. Se puede observar cómo, con el paso del tiempo (al permanecer el número de

nodos estable), el grado medio crece ligeramente (de 7.18 en marzo a 7.7 en agosto), al igual que el diámetro (de 5 a 6). La densidad (de 0.15 a 0.16) y el coeficiente medio de *clustering* (de 0.36 a 0.37) permanecen prácticamente estables, mientras que la longitud media del camino se reduce lógicamente, aunque de forma muy discreta (de 2.25 a 2.22).

En cuanto a los indicadores a nivel de nodo, los valores permanecen lógicamente muy estables durante todos los meses de medida. La correlación entre los valores logrados por las empresas en marzo y en agosto es, por tanto, elevadísima (cercana a 1) para todos los indicadores. En la Tabla IX (inferior derecha) se muestra, a modo de ejemplo, los valores de *inDegree* y *outDegree* en marzo y agosto para las dos empresas con mayor *eigenvector* (*Microsoft* y *Facebook*), donde se observa la estabilidad de los indicadores.

Tabla VIII. Redes de empresas tecnológicas en Twitter en marzo (izquierda) y agosto (derecha) en 2014



Empresa	Degree	Closeness	Betweenness	Clustering coefficient	Eigenvector
@microsoft	39	2.34	135.91	0.24	1.00
@facebook	25	2.98	30.36	0.34	0.95
@intel	42	1.95	211.97	0.23	0.92
@hp	37	1.73	251.35	0.31	0.80
@dell	26	2.25	33.52	0.39	0.76

Empresa	Degree	Closeness	Betweenness	Clustering coefficient	Eigenvector
@microsoft	39	2.32	116.77	0.26	1.00
@facebook	26	2.82	26.65	0.34	0.96
@intel	43	1.95	224.25	0.23	0.93
@hp	39	1.73	215.11	0.31	0.83
@dell	26	2.25	29.57	0.40	0.76

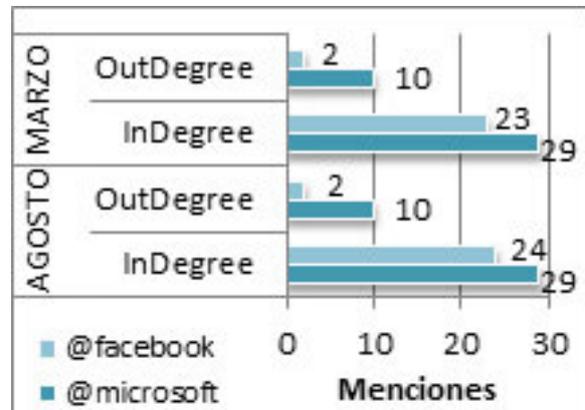
Tabla IX. Evolución y correlación de los indicadores de la Red (marzo y agosto, 2014)

INDICADOR (nivel de red)	MARZO 2014	AGOSTO 2014	VARIACIÓN (%)
Average degree	7.18	7.7	6.75
Diameter	5	6	16.67
Density	0.15	0.16	6.37
Average clustering coefficient	0.36	0.37	4.80
Average path length	2.25	2.22	-1.40

INDICADOR (nivel de nodo)	R*
InDegree	**0.997
OutDegree	**0.996
Degree	**0.997
Closeness	**0.959
Betweenness	**0.995
Clustering coefficient	**0.963
Eigenvector	**0.999

* Correlación entre el ranking de empresas por indicador en marzo y agosto (2014)

** Nivel de significancia $\alpha=0.01$



4. DISCUSIÓN

Los resultados muestran una evolución lenta y asimétrica de las relaciones entre empresas tecnológicas a través de las menciones de *Twitter* a las cuentas corporativas. Aun así, deben discutirse diversos factores que pueden afectar a la interpretación correcta de los resultados.

El hecho de contabilizar *Tweets* selectivos (de una cuenta "x" a una cuenta "y"), unido a la interacción obtenida entre las empresas, ha implicado el manejo de cantidades relativamente bajas de *Tweets* mensuales por consulta (el promedio mensual de *Tweets* recibidos por las 50 cuentas es de 1.4 *Tweets*), hecho que concuerda con el estudio de *Brandchats*, mencionado anteriormente⁴. No se han localizado estudios similares con los que poder comparar directamente los resultados obtenidos, pese a ello se estima que los niveles de interacción obtenidos son discretos.

Respecto a la obtención de los datos, se ha utilizado un método indirecto (*Topsy*) para la captura de los *Tweets* en lugar de la propia API de *Twitter*. En ese sentido, Zimmer y Proferes (2014) ya detectan un descenso del uso de la API en la literatura científica, así como un aumento de trabajos en los que se utilizan datos proporcionados por terceros entre los que destaca *Topsy*, entre otras cuestiones por ofrecer un acceso retrospectivo desde el primer *Tweet* enviado en 2006 hasta la actualidad. Tanto la cobertura completa como los comandos selectivos de búsqueda de *Topsy* aseguran, sin necesidad de acceder a la API de *Twitter*, la localización de todos los *Tweets* que, desde cada empresa de la muestra, se han enviado al resto (tanto durante el último mes como en el histórico de *Tweets*).

Por otra parte, este estudio se ha centrado (con la excepción de *Apple*) en las cuentas oficiales corporativas de las empresas y no en el resto de cuentas (productos, servicios, etc.), desde las que seguramente se pueden haber establecido otro tipo de conexiones entre las compañías analizadas (compañía 1 menciona producto compañía 2; producto compañía 2 menciona producto compañía 3, etc.).

No obstante, el uso de las cuentas oficiales corporativas determina la identificación de relaciones empresariales oficiales. Ciertamente, las menciones entre estas cuentas podrían reflejar relaciones políticamente correctas o excesivamente superficiales, mientras que las cuentas de productos o desarrollos podrían mostrar conexiones más técnicas, por ejemplo. El estudio de las cuentas de los directivos personales podría igualmente aportar un nuevo enfoque al estudio, aunque suelen ser cuentas con poca actividad (de muchos *followers* y po-

cos *Tweets* publicados), y las relaciones mostradas pueden adolecer de los mismos problemas o limitaciones que las cuentas corporativas. Por ejemplo, desde la cuenta personal de Marissa Mayer¹², actual CEO (*Chief Executive Officer*) de *Yahoo!*, no se mencionan las cuentas de *Facebook* ni de *Microsoft* (ni las personales de sus respectivos directivos). En cualquier caso, las cuentas corporativas suponen solamente una primera base para la realización de futuros estudios métricos en los que se consideren la totalidad de cuentas disponibles (categorizadas), que permitan obtener un cuadro más amplio de los vínculos reales existentes.

La oficialidad de la empresa es igualmente importante. Se ha constatado cómo la cuenta @AppleOfficiall no ha sido mencionada por ninguna de las 49 empresas restantes en ningún momento, pese al incremento importante en popularidad (seguidores) que esta cuenta ha experimentado, y los contenidos claramente enfocados a la promoción de productos de la marca *Apple*.

Independientemente del tipo de cuenta utilizada (corporativa o de otro tipo), se deben señalar, entre otras, las siguientes limitaciones:

- El hecho de medir menciones de usuario puede excluir a otro tipo de interacciones en las que no se haya mencionado explícitamente un usuario, como puede ser un *hashtag*, mención textual sin marca alguna (de usuario o *hashtag*) o simplemente un URL embebido que redirija a una página de una compañía cualquiera.
- Algunas empresas pueden cancelar o modificar las cuentas de *Twitter*. Por ejemplo, a fecha de diciembre de 2015, algunas de las cuentas utilizadas en el estudio han dejado de funcionar, como es el caso de @GoogleAtWork (modificada por @GoogleforWork) o @SAP4SME (recientemente adquirida por una usuaria particular), lo que limita los estudios retrospectivos pues el buscador de *Twitter* no proporciona datos de cuentas canceladas o activas bajo otro administrador.
- El número de *Tweets* publicados y de *followers* puede variar de forma muy acusada en poco tiempo, debido a prácticas agresivas de marketing y difusión comercial. Por ello, los datos de la Tabla 1 deben contextualizarse en el entorno del período de tiempo en el que fueron tomados. En la figura 3 se muestra a modo de ejemplo el crecimiento en el número de *followers* de @IBM durante sólo un mes (de noviembre a diciembre de 2015), tiempo en el que obtienen 5.479 nuevos seguidores.

- El trabajo se restringe a un canal determinado (*Twitter*), que limita de forma importante el orden de magnitud de las interacciones existentes entre las compañías. No sólo habría que considerar otras plataformas sociales sino el análisis de enlaces para comprender mejor las relaciones web entre las compañías. A modo de ejemplo, en este trabajo se constatan 19 menciones de usuario de *Dell* a *Microsoft*, cuando un análisis de menciones de URL en *Google*¹³ nos indica aproximadamente algo más de 60.000 resultados.

La muestra de empresas (50) se considera por otra parte suficiente para responder a los objetivos planteados en la investigación, aunque su ampliación en el futuro permitiría describir con mayor granularidad el sector. Es importante indicar que todas las métricas de red dependen del número de nodos incluidos (que en este caso se mantiene constante en 50). La inclusión de nuevos nodos generaría cambios en las métricas a nivel de red (grado medio, coeficiente medio de *clustering*, densidad y diámetro de la red).

Por otro lado, los problemas relativos a la influencia del lugar geográfico en el grado de adopción de *Twitter* impiden la comparación entre empresas a nivel internacional de manera directa. En este estudio solamente se han seleccionado las cuentas oficiales de *Twitter* relativas a la sede de la empresa en un determinado país (Estados Unidos). Sin embargo, las redes obtenidas a partir de las cuentas de esas mismas empresas en otros países

podrían aportar relaciones diferentes, que deben tenerse en cuenta a la hora de interpretar los vínculos establecidos.

Finalmente, la selección de empresas tecnológicas (con un alto grado de adopción de *Twitter* y un amplio y diverso uso de este canal) condiciona las relaciones intensas entre ciertas compañías. La ampliación de este estudio a empresas de otros sectores industriales podría ayudar a investigar la posible relación entre el grado de adopción de *Twitter* y la intensidad de las relaciones por sector.

5. CONCLUSIONES

La principal conclusión que se puede extraer de este estudio es que el grado de interacción entre las 50 empresas tecnológicas con mayores ingresos netos en 2013, a partir de las menciones entre sus cuentas corporativas oficiales de *Twitter*, es débil (tanto si contamos el número de conexiones establecidas como si cuantificamos la intensidad de estas conexiones), estable, concentrado en unas pocas relaciones específicas y de carácter marcadamente asimétrico.

Grado interacción y evolución

A fecha de marzo de 2014, de las 2.450 posibles combinaciones entre pares de compañías, sólo en el 9.14% (224) de éstas se ha detectado al menos un *Tweet* de una empresa a otra (reflejando una densidad baja). Esta cifra asciende al 15.7%

Figura 3. Evolución del número de *followers* para la cuenta oficial de IBM (7 noviembre – 7 de diciembre de 2015)



Fuente: Twitter Counter - <http://twittercounter.com/IBM>

(385) en agosto del mismo año, siendo en todo caso una evolución positiva, aunque lenta, especialmente si tenemos en cuenta que el número de nodos ($n=50$) se mantiene estable durante todos los meses de medida.

Este hecho ha determinado que los distintos análisis seccionales mensuales hayan proporcionado una fotografía similar de las relaciones entre los canales sociales de las empresas. Los indicadores de red obtenidos a partir de las menciones web en *Twitter* han demostrado por otra parte ser estables y robustos ante pequeñas variaciones en las relaciones detectadas entre las compañías.

En cuanto a la intensidad de las relaciones, a lo largo de cinco meses (de abril a agosto de 2014) solamente se han cuantificado 338 *Tweets* entre las 50 empresas de la muestra, una cifra que se considera discreta, dada la cantidad de *Tweets* mensuales publicados por las empresas. Estos datos suponen, por otra parte, el 14.2% del total de *Tweets* históricos contabilizados entre este grupo de empresas (2.381), una cifra que revela una no excesiva concentración de interacciones durante el período de análisis.

Respecto al tipo de interacciones, los datos indican claramente un predominio en el uso de menciones directas (85%) respecto a los *ReTweets* (15%), que reflejan una preferencia por el contacto directo más que por la mera redifusión de contenidos.

Igualmente se detecta una moderada influencia del número de *followers* en la cantidad de *Tweets* recibidos por el conjunto de empresas, aunque dada la alta variabilidad de este indicador, este factor se debería analizar de forma más sistemática para conocer con más exactitud la influencia que éstos suponen a la hora de captar menciones del resto de empresas del sector.

La cantidad de *Tweets* mensuales publicados no parece tener, por otro lado, una influencia tan directa en las interacciones logradas; en este caso la orientación de los contenidos difundidos parece ser más determinante. Por ejemplo, la cuenta de AOL (orientada a la comunicación de contenidos generalistas) produce muchos *Tweets* pero con impacto nulo en el resto de empresas; sin embargo Netapp (más orientada a la difusión de contenidos dirigidos a usuarios específicos) publica muchos menos *Tweets* pero logra una mayor interacción del resto de compañías.

Asimetría

Se ha constatado una alta asimetría en las interacciones entre las empresas. Mientras unas compañías actúan como concentradores (proporcionan muchas menciones al resto, pero reciben

pocas menciones; *Netapp*, *CSC* o *Hewlett-Packard* serían ejemplos de esta categoría), otras se comportan como autoridades (reciben muchas menciones, pero apenas proporcionan; ejemplos serían *VMware*, *Microsoft*, *Intel* o *Facebook*). Aunque no se detectan *clusters* de gran intensidad, se destaca la interacción que *Microsoft* e *Intel* tienen con *HP*.

Clusters

Las empresas con una mayor centralidad de red corresponden a grandes compañías clásicas de la informática (*Microsoft*, *Oracle*, *Cisco*, *Intel* o *HP*), que forman el núcleo central de la red, mientras que empresas orientadas a la Web (como *Yahoo!* o *Google*) aparecen ligeramente más desconectadas y periféricas, debido a que reciben muy pocos *Tweets* del resto de compañías (con la excepción de *Facebook*). En cualquier caso, el bajo volumen de *Tweets* tanto mensuales como históricos explica la alta estabilidad de los indicadores de red (tanto a nivel de nodo como de la estructura de red completa).

Se concluye, por tanto, que las cuentas corporativas de las empresas tecnológicas en *Twitter* no son suficientes para analizar desde un punto de vista informétrico las interacciones y relaciones establecidas entre las compañías en este canal. Es posible que la competencia entre estas empresas y las distintas políticas de comunicación web (reflejadas tanto en el tipo de contenidos generados como en el número de seguidores) influyan en la baja densidad de interacciones constatadas, más si tenemos en cuenta tanto el alto dinamismo de *Twitter* como el carácter tecnológico de las empresas analizadas (que las hace más propensas a usar este canal de comunicación). El estudio de otros sectores industriales ayudaría, asimismo, a relativizar el grado de interacción obtenido.

La ampliación de este estudio al resto de cuentas oficiales (especialmente las relativas a productos) se estima necesaria, pues ello podría hacer aflorar interrelaciones de alta intensidad, no visibles en las cuentas corporativas (marcando así la existencia de canales específicos de comunicación entre compañías), o bien corroborar una carencia más coyuntural de interrelaciones entre empresas en el canal de comunicación web (*Twitter*).

En cualquier caso, tanto la existencia como la ausencia de interrelaciones (y la forma en la que éstas se dan) aportan información valiosa acerca de las políticas y prácticas de comunicación web entre empresas, útiles tanto para las propias compañías (toma de decisiones estratégicas, actividades de *benchmarking*, etc.) como para otros observadores. En ese sentido, el uso

de métodos e indicadores web aplicados en este estudio, con una fuerte base metodológica basada en la cibermetría, han demostrado ser de gran ayuda y valor en la cuantificación y descripción de estas relaciones.

Para finalizar, y dada la estabilidad de las relaciones entre empresas analizadas, se estima necesario el futuro diseño y aplicación de indicadores de red más sensibles a variaciones particulares para ser utilizados en estudios longitudinales, especialmente a la hora de detectar precozmente modificaciones en las relaciones existentes y posibles tendencias en la evolución de una red cerrada de estas características.

7. NOTAS

1. <http://www.statisticbrain.com/twitter-statistics> [fecha de consulta: 12-03-2015]
2. <http://www.alexa.com/topsites> [fecha de consulta: 12-03-2015]
3. http://www.loc.gov/today/pr/2013/files/twitter_report_2013jan.pdf [fecha de consulta: 12-03-2015]
4. <http://www.brandchats.com/los-12-tipos-de-usuarios-en-twitter-segun-brandchats> [fecha de consulta: 12-03-2015]
5. <http://learnbonds.com/is-tesla-motors-inc-tesla-ceo-elon-musk-violating-sec-rules/117158> [fecha de consulta: 12-03-2015]
6. <http://www.enriquedans.com/2015/04/redes-sociales-e-informacion-bursatil.html> [fecha de consulta: 12-03-2015]
7. <http://www.sec.gov/News/PressRelease/Detail/PressRelease/1365171513574#.VSeojvmsXX8> [fecha de consulta: 12-03-2015]

8. REFERENCIAS

- Adamic, I. A.; Huberman, B. A. (2001). The web's hidden order. *Communications of the ACM*, vol. 44(9), 55-59. <http://dx.doi.org/10.1145/383694.383707>
- Barabási, A.L. (2014). *Linked: how everything is connected to everything else and what it means for business, Science, and every day life*. New York: Basic Books.
- Barabasi, A.-L.; Albert, R. (1999). Emergence of scaling in random networks. *Science*, vol. 286 (5439), 509-512. <http://dx.doi.org/10.1126/science.286.5439.509>

6. AGRADECIMIENTOS

Este trabajo ha sido realizado en el marco del Proyecto de investigación con referencia APOSTD/2013/002, financiado por la Conselleria de Educación, Cultura y Deporte de la Comunidad Valenciana (España).

Agradecemos a los revisores sus importantes indicaciones y sugerencias, que han mejorado notablemente el manuscrito final.

ACKNOWLEDGEMENTS

This work has been carried out in the framework of the research project with reference APOSTD/2013/002, funded by the Regional Ministry of Education, Culture and Sport (Generalitat Valenciana, Spain).

The authors would like to thank the reviewers for their valuable suggestions, which have significantly improved the final manuscript.

8. <https://www.google.com/finance?catid=TRBC%3A57&ei=fuQFVMjzJeX6wAOTnIHICA> [fecha de consulta: 12-03-2015]
9. <http://googlefinanceblog.blogspot.com.es/2011/03/happy-birthday-google-finance.html> [fecha de consulta: 12-03-2015]
10. <http://www.statista.com/chart/1642/regional-breakdown-of-twitter-users> [fecha de consulta: 12-03-2015]
11. <http://topsy.com> [fecha de consulta: 12-03-2015]
12. <https://twitter.com/marissamayer> [fecha de consulta: 12-03-2015]
13. Por ejemplo, mediante la siguiente consulta: <"microsoft.com" site:dell.com -inurl:microsoft.com>

- Bollen, J.; Mao, H.; Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, vol. 2(1), 1-8. <http://dx.doi.org/10.1016/j.jocs.2010.12.007>
- Bulearca, M.; Bulearca, S. (2010). Twitter: a viable marketing tool for SMEs. *Global business and management research: an international journal*, vol. 2(4), 296-309.
- Carim, L.; Warwick, C. (2013). Use of social media for corporate communications by research-funding

- organisations in the UK. *Public relations review*, vol. 39(5), 521-525. <http://dx.doi.org/10.1016/j.pubrev.2013.08.006>
- Carlson, N. (2011). The real history of Twitter. *Business insider*. <http://www.businessinsider.com/how-twitter-was-founded-2011-4> [fecha de consulta: 12-03-2015].
- Case, C. J.; King, D. L. (2010). Cutting edge communication: microblogging at the fortune 200, twitter implementation and usage. *Issues in information systems*, vol. 11(1), 216-223.
- Case, C. J.; King, D. L. (2011). Twitter Usage in the Fortune 50: A Marketing Opportunity?. *Journal of marketing development and competitiveness*, vol. 5(3), 94-103.
- Del-Fresno-García, M. (2014). Haciendo visible lo invisible: visualización de la estructura de las relaciones en red en Twitter por medio del análisis de redes sociales. *El profesional de la información*, vol. 23(3), 246-252. <http://dx.doi.org/10.3145/epi.2014.may.04>
- Gentle, A. (2009). *Conversation and community: the social web for documentation*. Laguna Hills; XML Press, p. 342.
- Huberman, B. A.; Romero, D. M.; Wu, F. (2008). Social networks that matter: Twitter under the microscope. *First Monday*, v. 14(1). <http://dx.doi.org/10.5210/fm.v14i1.2317>
- Jansen, B. J.; Zhang, M.; Sobel, K.; Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, vol. 60(11), 2169-2188. <http://dx.doi.org/10.1002/asi.21149>
- Kietzmann, J. H.; Hermkens, K.; McCarthy, I. P.; Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business horizons*, vol. 54(3), 241-251. <http://dx.doi.org/10.1016/j.bushor.2011.01.005>
- Orduña-Malea, E.; Aguillo, I.F. (2014). *Cibermetría: midiendo el espacio red*. Barcelona: UOC.
- Orduña-Malea, E.; Delgado López-Cózar, E.; Serrano-Cobos, J.; Romero, N. L. (2015a). Disclosing the network structure of private companies on the web: the case of Spanish IBEX 35 share index. *Online Information Review*, 39(3), 360-382. <http://dx.doi.org/10.1108/OIR-11-2014-0282>
- Orduña Malea, E.; Torres Salinas, D.; Delgado López Cózar, E. (2015b). Hyperlinks embedded in twitter as a proxy for total external in links to international university websites. *Journal of the Association for Information Science and Technology*, 66(7), 1447-1462. <http://dx.doi.org/10.1002/asi.23291>
- Ortega, J. L.; Aguillo, I. F. (2011). Social Network Tools for the Assessment of the University Web Performance. En: B. White et al (ed.). *Social Media Tools and Platforms in Learning Environments* (pp. 185-201). Berlin Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-642-20392-3_11
- Ortega, J. L.; Aguillo, I. F. (2013). Network visualisation as a way to the web usage analysis. *Aslib Proceedings*, vol. 65(1), 40-53. <http://dx.doi.org/10.1108/00012531311297177>
- Paniagua, J.; Sapena, J. (2014). Business performance and social media: Love or hate?. *Business horizons*, vol. 57(6), 719-728. <http://dx.doi.org/10.1016/j.bushor.2014.07.005>
- Romero-Frías, E. (2011). Googling Companies-a Webometric Approach to Business Studies. *Leading Issues in Business Research Methods*, vol. 1, 93-106.
- Romero, D. M.; Galuba, W.; Asur, S.; Huberman, B. A. (2011). Influence and passivity in social media. En: Gunopulos, D.; Hofman, T.; Malerba, D.; Vazirgiannis, M. (editores). *Lecture notes in computer science*. Springer. Berlin; Heidelberg, Alemania. <http://dx.doi.org/10.1145/1963192.1963250> / http://dx.doi.org/10.1007/978-3-642-23808-6_2
- Rybalko, S.; Seltzer, T. (2010). Dialogic communication in 140 characters or less: How Fortune 500 companies engage stakeholders using Twitter. *Public relations review*, vol. 36(4), 336-341. <http://dx.doi.org/10.1016/j.pubrev.2010.08.004>
- Shu, Z. (2015). *How do technical companies use microblogging?*. Twente; University of Twente, p. 300.
- Smith, M. A.; Rainie, L.; Shneiderman, B.; Himelboim, I. (2014). Mapping twitter topic networks: From polarized crowds to community clusters. *Pew research internet project*. <http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/> [fecha de consulta: 12-03-2015].
- Soat, J. (2010). 7 Questions Key to Social Networking Success. *Informationweek.com*. <http://www.informationweek.com/software/social/7-questions-key-to-social-networking-success/d/d-id/1086153> [fecha de consulta: 12-03-2015].
- Stieglitz, S.; Krüger, N. (2011). Analysis of Sentiments in Corporate Twitter Communication –A Case Study on an Issue of Toyota. *22nd Australasian conference on information systems - ACIS 2011 Proceedings*, pp. 1-11 [paper 29]. Sydney, Australia: ACIS. <http://aisel.aisnet.org/acis2011/29> [fecha de consulta: 12-03-2015]
- Swani, K.; Milne, G.; Cromer, C.; Brown, B. P. (2013). Fortune 500 Companies' Use of Twitter Communications: A Comparison between Product and Service Tweets. *International journal of integrated marketing communications*, vol. 5(2), pp. 47-56.
- Thelwall, M. (2004). *Link analysis: An information science approach*. Amsterdam: Elsevier.

- Vaughan, L. (2004). Exploring website features for business information. *Scientometrics*, vol. 61(3), 467-477. <http://dx.doi.org/10.1023/B:SCIE.0000045122.93018.2a>
- Vaughan, L. (en prensa). Uncovering information from social media hyperlinks: An investigation of Twitter. *Journal of the Association for Information Science and Technology*.
- Vaughan, L.; Romero-Frías, E. (2012). Exploring web keyword analysis as an alternative to link analysis: a multi-industry case. *Scientometrics*, vol. 93(1), 217-232. <http://dx.doi.org/10.1007/s11192-012-0640-x>
- Vaughan, L.; Wu, G. (2004). Links to commercial web sites as a source of business information. *Scientometrics*, vol. 60(3), 487-96. <http://dx.doi.org/10.1023/B:SCIE.0000034389.14825.bc>
- Vaughan, L.; Yang, R. (2013). Web traffic and organization performance measures: Relationships and data sources examined. *Journal of Informetrics*, vol. 7(3), 699-711. <http://dx.doi.org/10.1016/j.joi.2013.04.005>
- Vaughan, L.; You, J. (2006). Comparing business competition positions based on web co-link data: the global market vs. the Chinese market. *Scientometrics*, vol. 68(3), 611-28. <http://dx.doi.org/10.1007/s11192-006-0133-x>
- Vaughan, L.; You, J. (2010). Word co-occurrences on Webpages as a measure of the relatedness of organizations: a new Webometrics concept. *Journal of Informetrics*, vol. 4(4), 483-491. <http://dx.doi.org/10.1016/j.joi.2010.04.005>
- Wamba, S. F.; Carter, L. (2013). Twitter adoption and use by SMEs: An empirical study. *Proceedings 46th Hawaii International Conference on System Sciences (HICSS)*, pp. 2042-2049. Honolulu, Estados Unidos: University of Hawaii. <http://dx.doi.org/10.1109/hicss.2013.577>
- Williams, S. A.; Terras, M. M.; Warwick, C. (2013). What do people study when they study Twitter? Classifying Twitter related academic papers. *Journal of Documentation*, vol. 69(3), 384-410. <http://dx.doi.org/10.1108/JD-03-2012-0027>
- Zimmer, M.; J. Proferes, N. (2014). A topology of Twitter research: Disciplines, methods, and ethics. *Aslib Journal of Information Management*, vol. 66(3), 250-261. <http://dx.doi.org/10.1108/AJIM-09-2013-0083>

ESTUDIOS / RESEARCH STUDIES

Medición de la presencia de la lengua española en la Internet: métodos y resultados

Daniel Pimienta*, Daniel Prado**

*Presidente de la Fundación Redes y Desarrollo (Funredes) y Secretario Ejecutivo de la Red mundial por la diversidad lingüística (Maaya)

**Consultor y miembro del Comité Ejecutivo de la Red mundial por la diversidad lingüística (Maaya)

Correos-e: pimienta@funredes.org, dhprado@gmail.com

Recibido: 13-08-2015; 2ª versión: 21-12-2015; Aceptado: 22-01-2016.

Cómo citar este artículo/Citation: Pimienta, D.; Prado, D. (2016). Medición de la presencia de la lengua española en la Internet: métodos y resultados. *Revista Española de Documentación Científica*, 39(3): e141. doi: <http://dx.doi.org/10.3989/redc.2016.3.1328>

Resumen: En esta comunicación se resumen estudios recientes que exploran y analizan la situación de la lengua española en la Internet e infiere lecciones metodológicas que podrían ser útiles para estudios similares para otros idiomas. El método ha sido utilizado por primera vez para obtener resultados acerca de la situación de la lengua francesa en la Internet, en un estudio financiado por la *Organisation internationale de la Francophonie*. Al tomarse en cuenta una selección de espacios y aplicaciones de mucho uso, la lengua española podría ser considerada entre la tercera y la cuarta lengua dentro de los usos de la Internet (cuando es la tercera en número de internautas) ocupando los primeros lugares en materia de P2P, videos y blogs, apreciándose un rol menor en materia de libros, páginas web y aplicaciones. Es así como el español en la Internet parece actuar más dinámicamente como vector de comunicación que como vector de información.

Palabras clave: Cibermetría de lenguas; diversidad lingüística; medición de lenguas; multilingüismo en el ciberespacio; español en la Internet; presencia de lenguas en la Internet.

Measuring the presence of the Spanish language on the Internet: methods and results

Abstract: This paper summarizes recent studies that explore and analyze the situation of the Spanish language on the Internet and infers methodological lessons that could be useful for similar studies for other languages. The method was first used to obtain results for the state of French on the Internet in a study funded by the *Organisation internationale de la Francophonie*. By taking into account a selection of web spaces and commonly used applications, Spanish could be ranked between third and fourth place in terms of Internet usage (whereas it is third in terms of Internet users), occupying the top positions for P2P, videos and blogs, but having a minor role in terms of books, webpages and applications. Thus Spanish on the Internet appears to act more dynamically as a vector for communication than for information.

Keywords: Cybermetric language; Linguistic diversity; measuring languages; multilingualism in cyberspace; Spanish on the Internet; presence of languages on the Internet.

Copyright: © 2016 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Attribution (CC BY) España 3.0.

1. INTRODUCCIÓN

La Red Mundial por la diversidad lingüística, Maaya (Maaya, 2015), ya sea directamente o a través de algunos de sus miembros, como Funredes (Funredes, 2014), el *Language Observatory Project* (LOP, 2012) o Unión Latina (Unión Latina, 2015), ha llevado a cabo una serie de estudios para analizar la situación de las lenguas en la Internet desde 1988 hasta el presente.

Por un lado, Funredes y la Unión Latina desarrollaron inicialmente una metodología de medición específica para un grupo de lenguas¹, permitiendo realizar una serie de campañas de medición en diferentes espacios de la Internet, entre 1988 y 2008 (Estudio lenguas y ciberespacio, 2012 y Observatorio de lenguas y culturas en la Internet, 2012).

Por otro lado, el *Language Observatory Project* (LOP, 2012) estudió, en la segunda parte del mismo periodo, el comportamiento de las lenguas en la mayoría de los dominios de Internet de nivel superior (conocidos como ccTLD) de países de Asia y África, con la intención de medir la situación particular de las lenguas minoritarias.

Mientras el *Language Observatory Project* basaba sus estudios en el rastreo sistemático de las páginas Web del dominio elegido mediante un algoritmo de reconocimiento de idiomas, Funredes y la Unión Latina utilizaban los dispositivos de conteo de número de ocurrencias de los motores de búsqueda, con un muestreo de palabras suficientemente específicas que pudiesen permitir comparación entre ciertos idiomas.

Pero en 2008 Funredes / Unión Latina dejaron de utilizar dicha metodología como consecuencia del cambio de comportamiento de los motores de búsqueda (porcentajes bajos de indexación, cifras diferentes en función del usuario, preferencia por los sitios que más beneficios económicos aportan al motor, ninguna confiabilidad en los conteos propuestos, etc.) y el *Language Observatory Project* dejó de realizar su medición sistemática, impidiendo así contar con medios serios de supervisión de la evolución de las lenguas en la Internet, dejando lugar a la circulación de cifras provenientes de fuentes poco fiables, de carácter promocional o comercial (Pimienta y otros, 2009).

Para superar esta situación, Maaya diseñó un proyecto de investigación muy ambicioso, Dilinet (Dilinet, 2014), aun en espera de financiación, sin la cual no se puede pretender una investigación profunda del cada vez más *insondable* ciberespacio. En paralelo, editó un compendio de artículos con grandes nombres mundiales especializados en la diversidad lingüística en Internet, algunos de cu-

yos artículos han facilitado y orientado la investigación aquí presentada (Net.Lang, 2012).

En espera del éxito del proyecto Dilinet, se exponen en este trabajo estudios cuyos métodos representan una alternativa, quizás menos ambiciosa, pero válida para alcanzar un nivel aceptable de estimación sobre la situación de las lenguas estudiadas en las zonas más visibles de la Internet, y emprender así una sistematización en la observación de la situación de las lenguas.

El método descrito aquí se ha utilizado por primera vez para la lengua francesa (OIF, 2014), solicitado por la Organización Internacional de la Francofonía (*Organisation internationale de la Francophonie* – www.francophonie.org) y podría servir de inspiración para estudios sobre otros idiomas con un gran número de hablantes, como el portugués, el ruso, el árabe o el alemán. Los resultados mostrados aquí han sido los obtenidos para la lengua española.

2. ANTECEDENTES Y ENFOQUE

Aunque diversos indicadores podrían evaluar de forma precisa la presencia de la lengua española (en educación, en ciencia, en organismos internacionales, en materia de traducción de idiomas, en materia de edición, etc. [Unión Latina, 2010]), quedan dudas al hablar de la presencia española en el ciberespacio, a pesar de que los medios de comunicación le atribuyen, sin mayor reflexión, un hipotético tercer lugar que ocuparía la lengua española en la Internet.

La voluntad, aparentemente simple, de medición de la "presencia" de una lengua en la Internet choca con un malentendido permanente, que es la consecuencia de la escasez de datos sobre el tema, causa de discrepancias en las cifras facilitadas por diversas fuentes.

Dos indicadores diferentes suelen confundirse en los resultados publicados:

- El porcentaje estimado *de usuarios de la Internet* en una lengua determinada;
- El porcentaje estimado *de contenidos en la Internet* para un determinado idioma.

La medición del número de usuarios de la Internet hispanohablantes o la del número de páginas web en español son fundamentalmente diferentes, y reflejan diferentes realidades que merecen una mirada también diferente: la primera medida está relacionada con lo que llamamos *brecha digital* (es decir, el acceso físico a la Internet) y el segundo, con *la brecha de contenidos* (es decir la presencia de información en el idioma estudiado), una brecha mucho menos conocida pero más decisiva.

Es así, cuando se encuentran en los periódicos o en algunos informes las cifras de la "presencia del español", se debe entender y diferenciar si se refiere a la lengua del internauta o a la de los contenidos. Por lo tanto, la afirmación de que el español es la tercera lengua de la Internet, información ampliamente difundida por los medios de comunicación, sólo tiene sentido si se especifica que la tercera población de usuarios de Internet es la hispanohablante. De ninguna manera eso implica que el español es el tercero en términos de contenidos como veremos adelante: nuestro estudio establece que, en términos de páginas web, el español sería la quinta lengua y si se ponderan debidamente todos los factores de presencia relacionados con la Internet (el acceso, los contenidos y los múltiples usos posibles) llegaríamos a una posición mediana entre 3 y 4.

Desde luego, este nuevo estudio introduce una nueva variable que no había sido estudiada en los trabajos anteriores (más centrados en contar páginas web), la de medir los usos en función de aplicaciones.

Medición de usuarios de Internet por idioma

Los datos sobre el número de internautas por lengua provienen de la fuente más consultada en este ámbito: InternetWorldStats (InternetWorldStats, 2013), fuente que, si bien está lejos de satisfacer las expectativas de estadísticas rigurosas, al menos tiene el mérito de ser la única que existe (para medir los usuarios por lenguas) y de actualizarse periódicamente. Su metodología consiste en determinar los principales idiomas que se utilizan en cada país y cruzar esta información con los datos de la Unión Internacional de Telecomunicaciones sobre el número total de usuarios de Internet en cada país. Es de notar que las cifras dejaron de ser actualizadas entre mayo del 2011 y finales del año 2013, tomadas como referencia para este estudio.

Sin embargo, los datos de la Unión Internacional de Telecomunicaciones son producidos por los gobiernos, lo que no es necesariamente un criterio de fiabilidad y de homogeneidad, puesto que algunos países tienden a inflar las cifras dadas a la Unión Internacional de Telecomunicaciones para demostrar el éxito de sus esfuerzos para luchar contra la brecha digital (UIT, 2010). Por otra parte, no hay ninguna mención de la metodología utilizada por InternetWorldStats para contar el número de usuarios de Internet para las lenguas estudiadas. Al parecer, el único criterio es el del (o los) idioma(s) oficial(es) de cada país. InternetWorldStats también utiliza, para cada país, diversas fuentes de mercadeo, sin metodología común entre los paí-

ses. Por último, este estudio se limita a los diez principales idiomas de los usuarios de Internet, en contraste con la compañía que proporcionaba cifras comparables antes de 2007 (GlobalStats, desaparecida hoy día), consultable solo mediante el *Wayback Machine* del sitio <https://archive.org>.

Medición de contenidos por idioma

En materia de medición de la lengua de los contenidos a nivel mundial, hay, y ha habido, profusión de publicaciones (por lo general, originadas por empresas de mercadotecnia) con cifras diversas cuyo método ha sido raramente revelado, lo que imposibilita validar los resultados. Sólo un estudio puntual realizado en los albores del siglo mantenía un rigor cierto y debidamente documentado, de la sociedad Xerox (Grefenstette y Nioche, 2001).

Más recientemente, las "Encuestas de Tecnología W3Techs" (W3Techs, 2014), probablemente la fuente de información más fiable y completa sobre los usos de Internet, tiene la ventaja de ofrecer datos sobre la lengua de los contenidos². W3Techs calcula sus estadísticas a partir de los elementos suministrados por Alexa (Alexa, 2014), una empresa capaz de proporcionar estadísticas sobre la frecuentación de sitios o páginas web a través de una barra de herramientas que los internautas aceptan instalar en su navegador y que reporta a Alexa las etapas de su navegación. Se trata de un muestreo compuesto por un número de usuarios selectos. Alexa presenta así los 25 millones de sitios más visitados sorteados por el número de visitas. A sabiendas de que la web cuenta (en marzo 2015) con casi 877 millones de sitios Internet (Netcraft, 2015), de los cuales se considera que casi 177 millones son sitios realmente en funcionamiento, esa muestra solo representa la punta de una pirámide. De esos 25 millones clasificados por Alexa, W3Techs toma los primeros 10 millones y determina los idiomas identificados en los metadatos o por algoritmo de reconocimiento de la lengua. Lo interesante es notar que W3Techs actualiza diariamente sus datos, lo que permite apreciar la evolución desde la fecha de inicio del servicio, junio de 2013, hasta hoy día³.

Otro elemento de interés en el trabajo de W3Techs es su capacidad de cruzar diferentes datos:

- http://w3techs.com/technologies/cross/top_level_domain/content_language permite cruzar los nombres de dominio y las lenguas de la página (por ejemplo, se sabe que el 30% de los sitios en francés pertenecerían al dominio de nivel superior .fr⁴, datos que no existen desafortunadamente para ningún país hispanohablante);

- http://w3techs.com/technologies/cross/content_language/top_level_domain realiza el cruce recíproco (así el 93% de los sitios de dominio .fr estarían en francés);
- http://w3techs.com/technologies/cross/content_language/ranking cruza la clasificación jerárquica con el parámetro de idioma (así el 60% de los sitios clasificados en los 1.000 primeros rangos estarían redactados en inglés).

Hoy día, la estimación de W3Techs sobre el lugar del inglés (55,5% de todas las páginas Web) es mucho más alta que la indicada por el estudio de Funredes / Unión Latina en 2007 (44%) y mucho más alta que lo que sería nuestra proyección actual (en torno al 34%). Y es que hay varias razones para obtener, con ese método, cifras de la lengua inglesa muy por encima de la realidad:

1. Funredes / Unión Latina (Paolillo, 2015) o el *Language Observatory Project* (Suzuki y otros, 2002) centraban su investigación en el contenido de *las páginas web*, lo que permitía medir, dentro de cada sitio web, el idioma de cada página y no solo de la principal (*home page*); en cambio W3Techs centra su investigación *en los sitios web*, lo que provoca la toma en cuenta para el sitio entero de la página principal, muchas veces conteniendo inglés (lengua de entrada principal de muchos sitios multilingües), incluso si el resto está escrito en otros idiomas.
2. Dado que Alexa se instala voluntariamente por el usuario, es un buen instrumento para medir los sitios que visitan los usuarios que lo han instalado, pero es un instrumento muy utilizado en ciertas regiones del planeta, y mucho menos en otras, lo que introduce muchos sesgos en materia de poblaciones estudiadas.
3. Alexa, por su naturaleza, mide el uso y no la existencia; de modo que las páginas no visitadas o poco visitadas por sus usuarios no llegan a ser identificadas y, por lo tanto, contabilizadas por W3Techs.
4. Pero la principal razón es que W3Techs sólo considera los 10 millones de sitios web más visitados según Alexa, es decir, un poco más del 1 % de los sitios existentes o, para ser más justos, un poco menos del 6% de los sitios considerados válidos. Por lo tanto, los sitios visitados incluirán necesariamente los medios de comunicación y los sitios comerciales más reputados de los diferentes países, mayoritariamente los países occidenta-

les, con los Estados Unidos en primer lugar. Es así como aquellos sitios que presentan un interés mundial menor (sitios científicos, culturales, administrativos nacionales, sitios localizados en lenguas menos utilizadas, etc.), no figurarán en dicha clasificación. Entre los posibles sesgos lingüísticos de este estudio, cabe señalar que la República Checa tendría más páginas que Corea, o bien que habría menos páginas en chino, en alemán o en ruso, con una población en línea casi 10 veces más grande.

A pesar de sus limitaciones, W3Techs representa la fuente más atractiva entre los indicadores disponibles en la actualidad y se deben aceptar con satisfacción los progresos que representa para el campo de estudio.

Evolución de los motores de búsqueda desde 2008

A partir de 2008 se ha reducido el número de motores de búsqueda y aquellos que han quedado (Google, Yahoo, Bing / Live Search, Ask, AOL, Lycos, Excite, Exalead, Teoma⁵) han evolucionado de la misma manera:

- reducción significativa del porcentaje de la parte indexada de la Web (más del 80% a menos del 5% del espacio total);
- pérdida total de credibilidad de las cifras publicadas por el número de apariciones de una palabra clave determinada;
- implementación de la "*búsqueda inteligente*" por palabra-clave dando lugar a la pérdida de asociación entre palabras-clave y resultados (ya sea causada por la traducción automática, o por algoritmos de sinonimia o de corrección ortográfica).

A partir de 2008, y de manera gradualmente amplificada, el tamaño de la Web se ha convertido en incontrolable y puede considerarse en términos prácticos, como tendente hacia el infinito. Es decir que los motores de búsqueda no pueden, por razones de costo, llevar a cabo un rastreo sistemático integral de toda la web, contentándose con un aproximado 5% de páginas indexadas en total⁶. Si hay un área en donde la falta de transparencia es la regla, es en el tamaño de los índices. En función del motor de búsqueda, se utilizan varios trucos para ocultar esta limitación de exploración de todas las páginas de un sitio web, la cual no se aplica de la misma manera a todos los idiomas en clara desventaja de las lenguas minoritarias, menos susceptibles de atraer publicidad o audiencia.

Junto con el auge de la Web 2.0 la naturaleza de la Internet ha cambiado y las páginas estáticas (HTML de base) han dejado un mayor espacio para las páginas dinámicas. En el mismo período, la topología de la Internet en materia de lenguas ha cambiado radicalmente (Prado y otros, 2010), con una reducción en términos de crecimiento de las lenguas inicialmente bien representadas (lenguas occidentales y japonés, en particular) y crecimientos exponenciales para los idiomas de Asia y, más recientemente, de la lengua árabe. Al mismo tiempo, la naturaleza del contenido ha evolucionado mediante la reducción de la proporción de información textual y un aumento masivo de documentos multimedia, significando, para el video en particular, un 79% por ciento de todo el tráfico estimado en 2018, cuando representaba un 66% en 2013. Ver *Cisco Visual Networking Index* para las predicciones 2014-2019 (CISCO, 2014).

En tales condiciones, el porcentaje de páginas textuales sigue siendo un indicador de cierta importancia pero, frente a una realidad más compleja, se deben crear otros indicadores que reflejen mejor dicha complejidad y aceptar que se deben tratar separadamente elementos parciales de un mosaico, en lugar de tratar con un número limitado de indicadores integrales.

En 2010, la Unión Latina, en colaboración con Funredes, llevó a cabo un primer intento de captar la percepción de la realidad compleja de las lenguas en diferentes espacios de la Web, y aunque dicha experiencia no dio pie a una publicación específica, permitió contribuir indirectamente a algunas publicaciones (Pimienta, 2011, UIT, 2010).

El trabajo del año 2010 exploró la presencia de un gran número de lenguas en las aplicaciones más conocidas (redes sociales, blogosfera, peer-to-peer, motores de búsqueda, VOIP⁷, Wikipedia, Youtube, etc.). A menudo, en ausencia de otras alternativas, los datos se construyeron a partir del origen geográfico del tráfico de dichas aplicaciones, cruzando los datos por país con fuentes demolingüísticas. Estos principios que han contribuido a perfilar un nuevo enfoque metodológico se ampliaron y extendieron en este nuevo estudio.

3. METODOLOGÍA

La metodología propuesta se basa en dicho enfoque del 2010 y se aplica a un determinado idioma, el español en este caso, comparándolo con los principales idiomas utilizados en la Internet, extendiéndola al mayor número posible de espacios y de aplicaciones. El enfoque inicial que estaba basado a menudo en el tráfico por país se complementa con la investigación intensa y sistemática sobre el contenido lingüístico de esas aplicaciones.

La metodología se basa entonces en un trabajo amplio y abierto de colecta de datos sobre el uso de las lenguas en diferentes aplicaciones y espacios de Internet, a partir de fuentes diversas y múltiples, seguido por una recopilación y organización de esos datos, luego, su evaluación y validación, finalmente el cruce de los datos organizados con diferentes estudios de referencia considerados como fiables, y en última instancia, la puesta en perspectiva de los resultados obtenidos con el fin de comprobar las tendencias y los indicadores compuestos que dan cuenta sobre los nuevos acontecimientos.

Los elementos metodológicos clave de este enfoque son:

- espacios y aplicaciones seleccionados;
- búsqueda, selección y análisis de fuentes de información ofreciendo datos sobre el lugar del español en relación con esos espacios y aplicaciones;
- cruce con datos demo-lingüísticos que permiten poner en perspectiva los datos recogidos;
- síntesis a partir del conjunto de datos dispersos obtenidos permitiendo informar, de manera significativa, sobre el lugar del español en la Internet en comparación con otras lenguas.

Se debería advertir que:

1. Los datos expuestos son extraídos de fuentes analizadas entre finales del 2013 e inicios del 2014 y no reflejan la situación de hoy ni el espacio abrumador que ha tomado la Internet móvil y sus aplicaciones propias en el 2015.
2. Como sea, la confianza que se le puede atribuir a las numerosas fuentes procesadas no es homogénea según las aplicaciones o espacios, y en muchos casos no es muy alta. Se deben entonces tomar los resultados específicos con cierta precaución, aún más cuando una de las categorías comporta muy pocas fuentes.

Lo que sí consideramos es que el conjunto de todas esas fuentes ofrece globalmente una estimación relativamente confiable sobre lo que es la situación del español en la Internet.

Se han identificado un centenar de aplicaciones y espacios de Internet como los más apropiados para informar sobre el espacio de las lenguas en la Internet. Treinta de ellos no ofrecían datos fiables o utilizables de manera uniforme, por lo que han sido excluidos provisoriamente y se integrarán tan pronto como lo permitan. Igualmente, ciertos espacios utilizados estudiados no presentan datos utilizables para el objetivo de este artículo por lo que se han descartado, como ha sucedido con los navegadores y los motores de búsqueda.

La tabla I muestra la lista inicial de los espacios y aplicaciones elegidos (incluyendo los que se desecharon), organizada según el tipo de aplicación o el espacio estudiados.

4. LAS FUENTES ANALIZADAS

Se necesitó un importante esfuerzo para recopilar fuentes sobre la presencia de las lenguas en la Internet en general, y en español, en particular. Este capítulo está dedicado a las fuentes que alimentaron los resultados presentados.

La falta de producción de datos en materia de cibermetría de lenguas ha creado un vacío en donde la gran mayoría de las fuentes provienen de empresas de negocios o de mercadeo en línea, que filtran información parcial de forma gratuita (y por lo general sin revelar la metodología) como medio para promover sus servicios de pago. Fuera de las estadísticas tradicionales y confiables que proporcionan cierta información útil sobre las lenguas (Organización de las Naciones Unidas, UNESCO, Unión Internacional de Telecomunicaciones, Organización para la Cooperación y Desarrollo Económicos, Unión Europea, etc.), no sin

Tabla I. Espacios y aplicaciones de Internet seleccionados para analizar la presencia de las lenguas en Internet. En cursiva, las fuentes que han sido descartadas en las ponderaciones finales para la lengua española o que no han sido utilizadas en el presente artículo

Infraestructura	Libros y artículos en línea	Teléfonos y tabletas	Mensajería y telefonía IP
Usuarios de Internet por lengua	<i>Biblioteca virtual</i>	Smartphones	Skype
<i>Computadoras por país</i>	<i>GoogleBooks</i>	<i>Tabletas G3</i>	<i>QQ</i>
Servidores por habitantes	Amazon	<i>Data Sims</i>	AIM
<i>Sitios web por internautas</i>	<i>GoogleScholar</i>	3G	ICQ
<i>Penetración de Internet</i>			Yahoo!
<i>Penetración de banda ancha</i>			
<i>Tráfico de traducción automática</i>			
<i>Herramientas lingüísticas</i>			
<i>Telefonía móvil por habitante</i>			
Descarga de archivos y P2P	Redes sociales y videos	Blogs	Conteo de páginas Web
<i>Megaupload</i>	Wikipedia	<i>Technorati</i>	W3 Techs
Rapidshare	Facebook	Blogs	<i>Webboar</i>
<i>Filefactory</i>	Twitter	Wordpress	Internetarchive
<i>Depositfiles</i>	Linkedin	<i>Google Blogs</i>	<i>Google</i>
<i>Hotfile</i>	Viadeo	<i>Blogger</i>	<i>Baidu</i>
<i>Uploading</i>	<i>Xing</i>	<i>Blogspot</i>	<i>Wolfram Alpha</i>
<i>Uploaded</i>	<i>Yahoo</i>	<i>Sina Weibo</i>	MSN
<i>Fileserve</i>	Google+	<i>Technorati</i>	<i>Bing</i>
<i>Mediafire</i>	Windows Live Profile	Tumblr	Foofind
Gigasize	<i>Myspace</i>		<i>Rtbot</i>
Bitshare	Livejournal		Ccsearch
<i>4shared</i>	<i>Secondlife</i>		<i>Altavista</i>
	Ning		<i>Yandex</i>
	<i>Tuenti</i>		<i>Wikia Search</i>
	Hi5		Open Directory Project
	Orkut		
	Badoo		
	Instagram		
	<i>Sonico</i>		
	<i>Qzone</i>		
	Youtube		
	<i>Googleplay</i>		

Correo electrónico	Motores de búsqueda	Navegadores	Sistema operativo y aplicación ofimática
Gmail	<i>Google</i>	<i>Chrome</i>	<i>Windows</i>
Hotmail	<i>Bing</i>	<i>Firefox</i>	<i>Linux</i>
<i>Yahoo</i>	<i>Yahoo!</i>	<i>IE</i>	<i>Mac</i>
<i>Yandex Mail</i>	<i>Yandex</i>	<i>Safari</i>	<i>Ios</i>
<i>Icloud</i>	<i>Baidu</i>	<i>Opera</i>	<i>Android</i>
<i>Outlook</i>	<i>Otros</i>	<i>Otros</i>	OpenOffice
			Microsoft Office
			<i>Otros</i>

En cursiva, las fuentes que han sido descartadas en las ponderaciones finales para la lengua española o que no han sido utilizadas en el presente artículo

dificultades en el momento de discriminar el aspecto lingüístico en la Internet, algunos consultores o expertos difunden la recopilación de información sobre un espacio o aplicación determinados, con el fin de promover sus conocimientos. Es el análisis del conjunto de esos sitios profesionales, aunque a menudo imperfectos, lo que permite, no sin cierta dificultad, recopilar un número interesante de datos relativamente confiables segmentados por aplicación (Wikipedia, Twitter, Youtube, Facebook, etc.) o por espacio (motores de búsqueda, correo electrónico, VoIP, etc.).

Otra limitación a tener en cuenta en este tipo de análisis es que el grado de globalización de los espacios y aplicaciones es cada día más variable, puesto que hay países o regiones que adoptan aplicaciones locales específicas en detrimento de las principales aplicaciones con vocación mundial (Baidu en lugar de Google en China y V Kontakte y Odnoklassniki en lugar de Facebook en Rusia, por ejemplo). Es importante tener en cuenta este factor cuando se establecen los datos cuantitativos sobre el uso de las lenguas (o sobre el uso por países) para cada aplicación o espacio determinado.

Por ejemplo, concluir que en las redes sociales no profesionales, el francés sería la cuarta o la primera lengua a partir de sus resultados respectivos en Facebook o Viadeo, debe interpretarse de manera relativa puesto que ninguna de estas aplicaciones tiene una penetración geo-equilibrada o *linguo-equilibrada*. Para obtener un indicador creíble sobre el lugar que cada lengua ocupa en una jerarquía de uso, cada aplicación debe ser ponderada en función de su distribución en el mundo (peso total y presencia relativa de un país).

Una visión rápida permite apreciar que las aplicaciones más globalizadas son Wikipedia, Twitter y YouTube⁸; otras son relativamente globalizadas, como Facebook y Google (ésta, con competencias de motores locales en China, Rusia, Kazajstán, Corea, etc.), o que muestran diferentes hábitos (en Japón, Yahoo! es mucho más utilizado que Google, por ejemplo).

Para la mayoría de las aplicaciones y espacios estudiados se debe tener cuidado y matizar las conclusiones lingüísticas de los resultados en función de posibles sesgos en los usos.

Desde 2010, se han analizado y almacenado varios cientos de fuentes potenciales, y se ha mantenido una observación permanente mediante búsqueda de palabras clave o consultando los enlaces externos de las fuentes ya seleccionadas. Muchas fuentes han sido excluidas por alguna o algunas de las causas siguientes:

- el campo de estudio es demasiado pequeño o parcial
- los datos parecen ser demasiado parciales
- las estadísticas no han sido actualizadas, o bien no había diferencias significativas entre dos fechas de consulta
- la metodología utilizada no permite comparar los datos
- la metodología utilizada es poco pertinente, no adecuada o poco creíble
- quedan dudas sobre la fiabilidad de la fuente.

Finalmente se han seleccionado, clasificado, evaluado y valorado cerca de 200 fuentes diferentes (URL, artículos, libros u otros) que permitirían identificar distintos indicadores sobre la presencia de las lenguas en diferentes áreas.

Clasificación de las fuentes

Cada elemento de este muestreo de fuentes fue notado y clasificado de 5 (promedio) a 10 (excelente) y aquéllos que obtenían menos puntaje fueron rechazados automáticamente. Las notas han sido asignadas en base a varios criterios: relevancia, confianza, alcance, transparencia de método, etc.

Para cada fuente se registraron los siguientes parámetros:

- El último año de publicación
- El ámbito (mundial, América Latina, España ...)
- Si la fuente se actualiza con frecuencia o no
- El tipo de fuente (por ejemplo, meta-información)
- El área de aplicación de la fuente (por ejemplo, Facebook)
- Si es específica para el idioma estudiado o no.

Las diferentes fuentes consultadas serán divulgadas próximamente en un metasitio con los parámetros correspondientes a fin de mantener un observatorio permanente. Mientras tanto, el grado de rápida obsolescencia de las fuentes y la dinámica de creación y eliminación de páginas de Internet es tal que no es apropiado citarlas, puesto que esta encuesta se terminó de realizar a finales del año 2013. Es el caso, por ejemplo, del sitio Socialbakers (Socialbakers, 2015), utilizado para medir la utilización de algunas redes sociales (LinkedIn, Facebook, Twitter, Google+), que ha cambiado la disposición de sus páginas desde entonces.

Datos demográficos y demo-lingüísticos

Como bien se sabe, los países monolingües son excepción, siendo el multilingüismo la regla, incluso si la adopción de una lengua oficial en la mayoría de los países pueda crear confusión. Ya sea en España, México, Estados Unidos, China o Camerún, o incluso en micro-estados como Mónaco, Malta o Singapur, el multilingüismo está siempre presente.

La mayoría de los estudios que han sido consultados proporcionan datos por país o región y tienden a extrapolar los resultados a la lengua oficial de cada país, sin tomar en cuenta las otras lenguas habladas⁹, lo que es una simplificación que conlleva a errores importantes. Repartir los datos de número de hablantes proporcionalmente a su presencia en el país sería otra simplificación, por supuesto, porque la brecha digital no se distribuye uniformemente en la población y los inmigrantes a menudo tienen menos acceso a la Internet. Sin embargo es un error más aceptable, pues así se asegura una mejor toma en cuenta de los idiomas de un país, en vez de ocultar totalmente esta diversidad.

En la mayoría de los países del mundo se necesita contar los hablantes de diferentes lenguas si se quiere transformar los datos por país (los cuales son las fuentes primarias en la mayoría de los casos) en datos por idioma (necesarios para este estudio).

Para poder comparar una lengua (en este caso el español) con las otras, es necesario establecer estadísticas fiables, no necesariamente para todos

los idiomas del mundo, pero por lo menos aquellos que son hablados por una amplia mayoría. Existen estadísticas confiables para algunas lenguas habladas en territorios bien definidos y que conocen un importante desarrollo (por ejemplo, en Europa), siempre que tengan un estatus oficial o instituciones de tutela¹⁰ y que la diáspora que aún lo habla esté bien estudiada (es el caso del español en los Estados Unidos, en Francia, en Brasil, por ejemplo). Pero esta tarea se complica para los idiomas que no cumplen ninguna de estas condiciones.

Dificultades con lenguas sin organismo de tutela

Podemos, por ejemplo, señalar el caso de lenguas como el occitano (hablado en España, Francia e Italia), el piemontés (hablado en Italia) o el franco-provenzal (hablado en Francia, Italia y Suiza), para los cuales, a pesar de que se hablan en zonas desarrolladas y en territorios bastante bien definidos, la falta de institución de tutela para el conjunto de la lengua¹¹ afecta a la calidad de las cifras sobre el número de hablantes.

Dificultades con lenguas habladas en territorios extensos

Para algunos idiomas hablados en extensos territorios y diversas condiciones socioeconómicas, hay estadísticas relativamente fiables; este es el caso del francés o español, por ejemplo, no así para otros idiomas con características similares, tales como el inglés, el chino, el portugués¹² o el árabe para los cuales hay diferencias muy importantes entre las fuentes, en particular en el conteo de los que lo hablan como segundo idioma.

Fuentes demo-lingüísticas en conflicto

A las complicaciones ya mencionadas se suman las divergencias en las metodologías de conteo de los hablantes en los numerosos estudios demo-lingüísticos. Poder comparar todos los estudios disponibles sobre el número de hablantes de todas las lenguas en el mundo para crear una tabla comparativa completa es una tarea enorme, por encima de los recursos de este estudio. Por lo tanto, es impensable poner en paralelo los resultados de diferentes estudios sin un análisis detallado, porque pondría al mismo nivel cifras obtenidas por definiciones no comparables o métodos no homogéneos.

Tipología de lenguas

Un problema adicional que hay que resolver es la tipología de las lenguas. ¿Debe ser tomado el idioma alemán como un todo incluyendo los distintos

dialectos, a veces muy distantes unos de otros? ¿Debemos considerar el árabe literario -o sólo el clásico- o todos los árabes dialectales (árabe argelino, árabe levantino, árabe egipcio, etc.)?

Algunos especialistas aprobarán unas, otros aprobarán otras, pero no están calificados los autores de este estudio para tomar decisiones al respecto.

Segunda lengua (L2)

Pero sin duda, el dilema más agudo es la forma de tener en cuenta los hablantes "L2" (quienes, sin ser su lengua materna, la utilizan seguido o la dominan en un grado satisfactorio) para las lenguas vehiculares. En este estudio, anotamos L1 para los hablantes de lengua materna y L2 para los hablantes que utilizan con fluidez ese idioma, no materno para ellos.

Este aspecto es particularmente importante para una serie de lenguas que permiten un mayor acceso a la información o su utilización posibilita una mayor difusión de ideas, puesto que aquellos que la dominan suficientemente, sin que sea su propia lengua, la utilizarán en prioridad en ciertos contextos específicos, e Internet no escapa a dicha situación.

Hay muchas lenguas vehiculares en el mundo, pero son aquéllas que han tenido un pasado colonial o imperial importante las que han adquirido un lugar relativo de *prestigio* (entendido como la influencia o autoridad que le confiere la comunidad a dicha lengua) y, por lo tanto, interesa medir la población que pueda utilizarla en la Internet como lengua de comunicación. Es por dicha razón que el inglés se posiciona en primer lugar entre las lenguas mundiales (a pesar de tener muchos menos hablantes vernáculos que el mandarín o el español) y que el francés, el portugués, el ruso, el alemán, el mandarín o el malayo indonesio son porcentualmente mucho más utilizados de lo que el número de hablantes vernáculos dejaría suponer.

Ahora bien, el número de hablantes vernáculos es consecuencia del grado de penetración de una lengua en el territorio que fuera dominado. Es sabido que el español tuvo un grado de penetración mucho mayor en los territorios colonizados que el inglés, el francés o el portugués, por ejemplo. Es así como habría unos 400 millones de hablantes de español lengua materna en los países en que la misma es oficial (en los que viven unos 440 millones de habitantes), mientras que el francés, por ejemplo, es lengua materna de unos 70 millones de individuos, pero es oficial en países en los cuales viven unos 800 millones de habitantes, lo que le permite ser lengua primera y segunda (entendida

esta última como lengua de uso sostenido) de unos 230 millones de individuos.

Es así que para lenguas como el inglés, el portugués, el ruso o el alemán (aun muy utilizado como lengua segunda en el Este de Europa), es importante el número de hablantes vehiculares a la hora de medir su uso en la Internet. Es mucho menos pertinente su cálculo para la lengua española, puesto que el número de hablantes de lengua segunda (L2) es poco significativo (aunque numeroso en cifras) con respecto al número de hablantes de lengua materna (segunda lengua materna en el mundo luego del mandarín).

Por esta razón, al realizar mediciones que comparen los hablantes de lengua materna exclusivamente, los resultados muestran una neta predominancia del español con respecto a otras lenguas (como el francés, el portugués, el alemán, el ruso, etc.), pero no así en ciertos segmentos especializados en los cuales otras lenguas pueden ser más utilizadas por su carácter vehicular en grandes regiones (y el reservorio de hablantes de lengua segunda que eso representa) y la lengua española puede retroceder en uso global. Es importante notar que este aspecto es válido para segmentos especializados pero mucho menos para segmentos en los cuales el discurso es informal y familiar, en los cuales el uso de la lengua materna se afianzará.

Opciones demo-lingüísticas

Para tomar mejor cuenta de todos estos elementos se ha seleccionado una serie de opciones para el mejor equilibrio entre la homogeneidad y fiabilidad de los datos demo-lingüísticos.

1) Ethnologue para contabilizar los hablantes de lengua materna (L1)

Al contabilizar los hablantes de la lengua materna, la elección se hizo de forma natural en Ethnologue (Ethnologue, 2013), la única fuente que proporciona cifras actualizadas continuamente sobre todas las lenguas del mundo. Esta fuente es a menudo inexacta en sus cifras, frecuentemente incompleta y sus actualizaciones no son homogéneas, sin embargo es la única que puede proporcionar datos dinámicos de todos los idiomas del mundo mediante la aplicación de una metodología relativamente consistente¹³. La resolución expuesta y explicada¹⁴ por los autores del "Barómetro Calvet" (Calvet, 2012) inspiró esa decisión.

Wikipedia (Wikipedia, lista de países por población, 2014) era una opción, pero la famosa enciclopedia no ofrece una metodología estable

para el recuento de los hablantes (y por una buena razón, ya que la naturaleza de la Wikipedia no es proporcionar información centralizada, cada artículo es independiente de los otros). Mientras que Wikipedia está tomando Ethnologue como su fuente principal para la mayoría de los idiomas inventariados, para otros idiomas toma como fuentes diversos estudios, ofreciendo diferentes metodologías y, por tanto, no comparables entre sí. Por otra parte, para algunos idiomas, Wikipedia es cauteloso y da rangos y figuras no precisas¹⁵. El caso de Albania, por citar una lengua que podría tener datos más precisos debido a la antigüedad de los estudios que se han dedicado, es esclarecedor. Ethnologue menciona 15.000 hablantes en Turquía, mientras que el artículo de Wikipedia, en diciembre de 2013, mencionaba cerca de tres millones de hablantes en este país, al mismo tiempo que incluía Ethnologue entre las fuentes.

Los profesionales de la lengua son, a menudo, muy críticos con los datos de Ethnologue¹⁶ y haberlo elegido para este estudio no acarrea, obviamente, sólo ventajas; sin embargo, los inconvenientes afectan sólo marginalmente este estudio porque se aplican sobre todo en idiomas identificados en estadísticas para las lenguas que no son objeto de este estudio.

2) Wikipedia para los datos demográficos

Después de estudiar varias fuentes para obtener buenas cifras de datos demográficos de todos los países del mundo, se utilizó Wikipedia en francés, que parecía federar (a fines de 2013) las mejores y más actualizadas fuentes¹⁷.

3) Situaciones específicas

Se tomaron ciertas decisiones arbitrarias en determinados casos específicos, por falta de precisiones por parte de las fuentes consultadas, puesto que algunas de ellas tuvieron en cuenta la macrolengua, mientras que otras sólo una de las lenguas de la familia. A modo de ejemplo, el alemán, el árabe, el chino, etc. En este caso la macrolengua se ha tomado como única referencia con el fin de permitir la comparación. Por lo tanto, cuando hablamos de chino, será todos los idiomas hablados en China, así como para el árabe, alemán, malayo, etc.

Aunque estas consideraciones tienen un impacto pequeño en los resultados de este estudio, centrado en el español, se reportan desde un punto de vista crítico porque estas decisiones podrían parecer contradictorias con un proceso normal de tratamiento de variantes lingüísticas.

5. MÉTODO DE EVALUACIÓN GLOBAL

En las páginas precedentes se han presentado muchos registros diferentes y, en general, independientes, de la presencia del español en la Internet en relación con una selección de espacios o aplicaciones. Lo que sigue es un intento de dar un resultado significativo y completo sobre el lugar que ocuparía el español en la Internet, a partir de esta amplia colección de datos. En otras palabras, la intención de ese capítulo es la de ofrecer una evaluación global de la presencia de la lengua española en la Internet, tratando de resumir el conjunto de datos en un resultado único, a sabiendas de que cada dato, tomado individualmente, sólo expresa el rango que la lengua española tiene en el espacio o aplicación estudiado.

Una aproximación simplista de determinar un coeficiente común consistiría en calcular el **promedio del rango** de la presencia del español a partir de la suma de datos y tomarlo como la expresión resumida y global del rango del español en la Internet.

Una aproximación más satisfactoria consistiría en atribuir a cada espacio o aplicación un peso relativo, de acuerdo con su importancia como indicador de la presencia de la lengua en la Internet, y calcular el **promedio ponderado de los rangos**, lo que daría más sentido a una estimación global.

Una última posibilidad, más satisfactoria aún, sería la de establecer una serie de parámetros de calificación para cada resultado de la serie de datos, en base a elementos de credibilidad, y calcular así el **promedio ponderado multi-criterio** a partir de una ecuación que tome debidamente en cuenta este conjunto de parámetros.

Se decidió entonces mostrar las tres aproximaciones, para poder compararlas entre sí, dando paso a una clasificación global que pudiese integrar todos los resultados y reflejar con mayor precisión el lugar estimado del español en la Internet.

Las clasificaciones obtenidas se presentan en la tabla de presentación de los resultados (tabla II), ordenadas por aplicación o espacio. La clasificación del español en la Internet varía, según la aplicación o espacio estudiado, **entre la posición 1 y la 7** con un **promedio simple de 3,12**, lo que, en definitiva, no la aleja mucho de su posición de tercera lengua en términos de número de internautas. Esto está graficado en la columna L1 de la tabla II, la cual expresa el rango medido del español en relación con cada espacio o aplicación. Así, por ejemplo, 1 indica que el español es la primera lengua para la aplicación Gigasize o 7 indica que el español es la séptima lengua, en el caso de Amazon.

Una ponderación simple permitiría asignar un peso para cada elemento (indicada "P" en la tabla de presentación de los resultados), para reflejar la importancia relativa de tal espacio o aplicación como indicador de la presencia de la lengua en la Internet. Así, por ejemplo, se ha asignado una nota de 10 a los estudios específicos que analicen directamente la lengua de los usuarios de Internet o a aquéllos que mencionen un "porcentaje de páginas en español" y un peso de 3, a aplicaciones como Hi5 o ccSearch, que no cuentan con especificación lingüística específica.

Con los valores propuestos, se puede establecer un **promedio ponderado**, suma de los productos (peso x rango) dividido por la suma de los pesos. En este caso, el **promedio ponderado sería de 3,27**.

Una ponderación relativamente más complicada (que llamamos **multi-criterio**) y que aporta una mayor precisión sobre la importancia de los parámetros, sería la de calcular "I" (indicador del lugar de la lengua española en la Internet) a partir de la ecuación: $I = A \times B \times C \times D / 100$, graficada en la Tabla II, en donde:

A = Grado de globalización del espacio o la aplicación considerado (0 a 10). Parámetro que refleja la relevancia del espacio o aplicación en cualquier lugar del globo. Así, Viadeo, red social especialmente usada en Francia tiene una nota de 2 en este estudio sobre la lengua española mientras que Facebook que es utilizado a nivel casi mundial (con las excepciones notables de China y Rusia, entre otros), llega a una nota de 8.

B = Grado de fiabilidad de los valores de este parámetro (0 a 10).

C = Nivel de confianza de los datos obtenidos para el español (0 a 10).

D = Relevancia del parámetro para la lengua española (de 0 a 10).

L1 = Clasificación del español para cada espacio o aplicación estudiada.

P = Ponderación simple del espacio o aplicación.

En este caso, el promedio resultando de la ecuación sería de 3,52. Por tanto, es razonable concluir que, en base a los parámetros establecidos y los resultados reunidos, el lugar relativo del español respecto a las otras lenguas en la Internet, según todos los criterios incluidos, sería **entre la tercera y la cuarta lengua**.

Si se comparan estos resultados con el lugar que ocupa la lengua española respecto a otras lenguas en el concierto mundial, se aprecia una situación aparentemente holgada gracias tanto a la importancia demográfica como a la relativa buena penetración de Internet en los países en los cuales se habla.

Sin embargo, incluso si la lengua española es la segunda en términos demográficos, y tercera tanto en número de países en los cuales es oficial como en cuanto lengua de enseñanza, se encuentra aún rezagada en términos relativos en el mundo virtual y queda trabajo si se desea que haya una coherencia entre uso y contenido.

Incluso si las instituciones encargadas de su tutela afirman sostenidamente un presunto segundo rol en el concierto de lenguas internacionales, es difícil poder afirmarlo en materia de relaciones internacionales o económicas, de comunicación científica, de normativa internacional o de patentes industriales, así como en materia de número de traducciones tanto literarias como especializadas, de industrias culturales o de turismo. Las lenguas francesa, alemana, china, rusa o japonesa ocupan muchas veces una situación superior a la castellana en esos sectores. Lo mismo pasa con la posición del español en la Internet, potencialmente importante como vector de comunicación pero que requiere más esfuerzos en términos de producción de contenidos especialmente en materia de presencia de libros y de páginas web como lo muestra la tabla III.

El control regular de todos estos resultados permitiría una observación del lugar del español por Internet, muy útil para determinar las políticas a seguir por el apoyo y la mejora de su presencia.

6. CONCLUSIÓN

Los enfoques metodológicos, propuestos en un campo caracterizado por una crisis prolongada, permiten superar parcialmente las lagunas existentes en la información sobre la presencia de las lenguas en la Internet y dan una contribución original al tema. Hay una probabilidad razonable de que puedan adaptarse, sin muchos cambios, para otros idiomas fuera del francés y el español.

Los resultados del español en la Internet son prometedores dada la gran proporción de internautas hispanohablantes,¹⁸ pero la presencia de contenidos en lengua española parecería no estar acorde con el número de utilizadores. Es probable que sean necesarias ciertas acciones voluntaristas (digitalización de libros, incentivos a la creación de artículos científicos en español, incentivos a crear artículos Wikipedia, alfabetización digital, e-gobierno o e-administración, etc.) tal como lo han promovido ciertos grupos lingüísticos (francés, lenguas nórdicas, lenguas minoritarias, etc.) permitiendo una presencia lingüística proporcional o superior a la de usuarios.

Tabla II. Posición de la lengua española en Internet. Evaluación general por espacios y aplicaciones

Espacio o aplicación	A	B	C	D	I	L1	P	L1x I	L1 x P	Tipo
3G	10	6	9	6	32	4	3	130	12	Infraestructura
Amazon	7	9	9	8	45	7	6	318	42	Libros
AOL/AIM	5	7	7	6	15	3	3	44	9	Aplicaciones
Badoo	6	5	7	5	11	2	3	21	6	Redes sociales
Banda ancha	10	9	9	7	57	4	4	227	16	Infraestructura
Bitshare ++	7	7	7	6	21	3	4	62	12	P2P
Blogs.com	6	7	7	5	15	3	5	44	15	Blogs
Ccsearch	6	7	7	6	18	3	3	53	9	Aplicaciones
Facebook	8	7	7	6	24	2	7	47	14	Redes sociales
Foofind	7	7	7	6	21	3	3	62	9	Aplicaciones
Gigasize	7	7	7	6	21	1	4	21	4	P2P
Gmail	7	5	8	6	17	3	6	50	18	Aplicaciones
Google +	8	7	7	6	24	3	7	71	21	Redes sociales
Hi5	7	6	7	4	12	1	3	12	3	Redes sociales
Hotmail	5	5	6	6	9	2	4	18	8	Aplicaciones
Icq	5	7	7	5	12	5	3	61	15	Aplicaciones
Instagram	7	5	7	6	15	2	5	29	10	Redes sociales
Internet archive	9	7	7	9	40	3	7	119	21	Contenidos
Internetworldstats	10	6	10	10	60	3	10	180	30	Usuarios
Ixquick	6	7	7	6	18	2	3	35	6	Aplicaciones
Linkedin	7	7	7	7	24	3	6	72	18	Redes sociales
Live journal	8	7	7	7	27	3	5	82	15	Blogs
MSN	7	7	7	6	21	3	5	62	15	Aplicaciones
Ning	7	7	7	8	27	3	5	82	15	Redes sociales
Open directory	9	10	7	9	57	5	7	284	35	Contenidos
Open office	9	9	9	8	58	5	5	292	25	Aplicaciones
Orkut	2	5	7	3	2	3	6	6	18	Redes sociales
Rapidshare	7	7	7	6	21	2	4	41	8	P2P
Servidores / hab.	10	9	9	7	57	2	5	113	10	Infraestructura
Skype	8	7	7	8	31	4	7	125	28	Aplicaciones
Smartphones	10	6	8	9	43	4	4	173	16	Infraestructura
Telefonía móvil	10	9	9	7	57	2	3	113	6	Infraestructura
Tumblr	6	6	7	6	15	2	4	30	8	Redes sociales
Twitter	9	8	7	9	45	2	8	91	16	Redes sociales
Viadeo	2	5	7	10	7	5	6	35	30	Redes sociales
W3techs	10	7	10	10	70	5	10	350	50	Páginas
Wikipedia	9	10	10	10	90	6	8	540	48	Contenidos
Windows live	7	7	7	6	21	3	4	62	12	Redes sociales
Wordpress	8	7	7	7	27	2	5	55	10	Blogs
Yahoo!	5	5	6	6	9	3	4	36	16	Aplicaciones
Youtube	8	7	7	8	31	2	7	63	14	Video
Promedios						3,12		3,27	3,52	

Ver definición de las columnas en el texto.

Tabla III. Posición de la lengua española en Internet. Valores por tipo de segmento

TIPO DE SEGMENTO	Posición	Número de aplicaciones o espacios estudiados
Aplicaciones	3,7	11
Blog	2,6	3
Contenidos y páginas	5	4
Infraestructura	3,1	5
Libros	7	1
P2P	2	3
Redes sociales	2,8	12
Usuarios (Número de)	3	1
Vídeo	2	1

8. NOTAS

1. Lenguas romances (catalán, español, francés, italiano, portugués, rumano) así como el inglés y el alemán.
2. Ver la rúbrica «Usage of content languages for websites».
3. Ver la secuencia del español en <http://w3techs.com/technologies/details/cl-es-/all/all> o también la de las lenguas con más de 0,1% de presencia en: http://w3techs.com/technologies/history_overview/content_language
4. En 2007, el trabajo de Funredes / Unión Latina calcula un valor de poco más de 26% para este indicador, y un valor de poco más de 57% para el porcentaje de francés en las páginas situadas en Francia (incluyendo las del .fr tanto como las de los dominios genéricos como .com o .org).
5. Según Netmarketshare Google sería, en abril 2015, el motor de búsquedas más utilizado con 62% pero con tendencia a la baja desde 2010: <http://market-share.hitslink.com/search-engine-market-share.aspx?qprid=4&qpcustomd=0&qpcustom=>
6. En 2008, varias fuentes mencionaban unas 127 mil millones de páginas indexadas (en particular el motor de búsqueda CUIL, ahora desaparecido, que afirmaba explorar la Web entera: <http://web.archive.org/web/20100916001435/http://www.cuil.com/>) y se puede encontrar fácilmente varias fuentes que mencionan la cifra de 40 mil millones para las páginas indexadas por Google (<http://www.openinnova.es/como-hace-google-para-indexar-las-paginas-web/>), cuando esta cifra es de 2008.
7. Voz por Internet
8. Aunque Youtube tiene un competidor en segunda posición, Dailymotion, con fuerte presencia francesa, puesto que pertenece a capitales franceses.
9. Es el caso de InternetWorldStats, por ejemplo, y uno de los escasos contraejemplos es una vez más Wikipedia que provee datos lingüísticos de calidad en http://meta.wikimedia.org/wiki/List_of_Wikipedias/sortable

7. AGRADECIMIENTOS / ACKNOWLEDGEMENTS

Esta investigación se realizó en base a un trabajo específico para la lengua francesa, financiado originalmente por la *Organisation internationale de la Francophonie* (OIF) y consultable en línea. Álvaro Blanco colaboró en la etapa inicial de recopilación de fuentes.

This research is a result of a specific work for the French language, originally funded by the *Organisation internationale de la Francophonie* (OIF) and available online. Álvaro Blanco contributed in the initial stage of gathering sources.

10. Entendemos por organismos de tutela, toda aquella entidad pública, parapública o asociativa que tenga por misión la de promover la difusión, promoción y/o desarrollo de la lengua en conjunto, como lo pueden ser diferentes direcciones de políticas lingüísticas, academias de la lengua, asociaciones locales o regionales de promoción cultural, etc.
11. Mencionamos la falta de organismo de tutela para el conjunto de la lengua, dado que muchas veces hay entidades que cumplen con dicha función, pero a nivel local y no global. Es el caso del aranés en España, variante del occitano, que cuenta con una protección de parte de las autoridades de Cataluña, pero no existen entidades de tutela para el conjunto de la lengua occitana.
12. El instituto Camões ofrece desde hace poco cifras con un margen de error aceptable.
13. Ethnologue publicó una nueva versión en línea con muchos cambios en los datos estadísticos después que terminamos ese estudio. Debe quedar claro que los datos demolingüísticos de Ethnologue tomadas en este estudio son aquellos publicados en mayo 2013.
14. <http://wikilf.culture.fr/barometre2012/tmpl.php?data=doc/methodologie/index>
15. Es el caso el inglés por el cual indica 309 < L1 < 380 millones.
16. Varios problemas también han sido identificados y registrados durante el curso de este estudio.
17. http://fr.wikipedia.org/wiki/Liste_des_pays_par_population
18. Según InternetWorldStats, el porcentaje de hispanohablantes conectados sería de 50,6% en 2013 a comparar con las cifras de 39% por la población mundial, de 58,4% para el inglés, 46,7% para el portugués y de 20,9% para el francés.

9. REFERENCIAS

Nota: Salvo indicación específica, las fuentes web citadas en este artículo se entienden consultadas entre finales del año 2013 e inicios del año 2015.

- Alexa - <http://alexa.com> [diciembre 2014].
- Calvet L.-J. (2012). Poids des langues (Baromètre Calvet) - <http://wikilf.culture.fr/barometre2012/> [marzo 2014].
- CISCO, *Cisco Visual Networking Index* - <http://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/> [diciembre 2014].
- Dilinet - <http://dilinet.org> [diciembre 2014].
- Estudio lenguas y ciberespacio en *Unión Latina* http://dtil.unilat.org/LI/2007/index_es.htm [julio 2012].
- Ethnologue: Languages of the World - <http://www.sil.org/ethnologue> [mayo 2013].
- Funredes, Fundación Redes y Desarrollo - <http://funredes.org> [diciembre 2014].
- Grefenstette y Nioche J. (2001). Estimation of English and non English Language use on the WWW. Technical report from Xerox Corporation Center Europe. - <http://arxiv.org/ftp/cs/papers/0006/0006032.pdf> [marzo 2014].
- InternetWorldStats - <http://www.internetworldstats.com/stats7.htm> [diciembre 2013].
- LOP, Language Observatory Project - <http://gii2.nagaokaut.ac.jp/gii/blog/lopdiary.php> [julio 2012].
- Maaya, Red Mundial por la diversidad lingüística - <http://maaya.org> [mayo 2015].
- Netcraft - <http://news.netcraft.com/archives/category/web-server-survey/> [marzo 2015].
- Net.Lang : Réussir le cyberspace multilingue / Towards a multilingual cyberspace* (2012), Paris, C&F Éditions. Consultable igualmente en <http://net-lang.net/> [diciembre 2014].
- Observatorio de lenguas y culturas en la Internet* en Funredes - <http://funredes.org/lc> [julio 2012].
- OIF (Organisation internationale de la francophonie). La langue française dans le monde (2014) - <http://www.francophonie.org/Langue-Francaise-2014/> [junio 2016].
- Paolillo J.; Pimienta D.; Prado D.; Mikami Y.; Zaki abu Bakar A.; Sonlertlamvanich V.; Vikas O.; Pavol Z.; Zaidi Abdul Rozan M.; Nagy János G.; Takahashi T.; Fantognan X. (2005). *Mesurer la diversité linguistique dans l'Internet*, Paris, UNESCO – Consultable igualmente en <http://unesdoc.unesco.org/images/0014/001421/142186f.pdf> [diciembre 2014].
- Pimienta D. (2011). Chapter 12, Language and Content, pp183-197, in *Accelerating Development Using the Web*. Gorge Sadowsky (ed.), Word Wide Web Foundation. http://g3ict.org/download/p/fileId_928/productId_241.
- Pimienta D.; Prado D.; Blanco A. (2009). *Douze ans de mesure de la diversité linguistique dans l'Internet: bilan et perspectives*, Paris, UNESCO – Consultable igualmente en <http://unesdoc.unesco.org/images/0018/001870/187016f.pdf> [diciembre 2014].
- Prado D.; Pimienta D.; Lemoulinier A. (2010). Diversité linguistique et cyberspace : état de l'art, enjeux et opportunités en *Cosmopolis*, - http://agora.qc.ca/cosmopolis.nsf/Articles/no2010_1_Diversite_linguistique_et_cyberspace__etat_de_l?OpenDocument [diciembre 2014].
- Socialbakers - <http://www.socialbakers.com/statistics/> [diciembre 2015].
- Suzuki I.; Mikami Y.; Ohsato A.; Chubachi Y. (2002). A Language and Character Set Determination Method based on N-gram Statistics, ACM Trans. on Asian Language Information Processing, Vol 1 N3, pp. 270-279. <http://dx.doi.org/10.1145/772755.772759>
- UIT, (2010). Monitoring the WSIS Targets. A mid-term review, Target 9 (content) en *World Telecommunication/ICT Development Report 2010 (pp175-192)*. Ginebra, UIT – Consultable igualmente en <http://www.itu.int/pub/D-IND-WTDR-2010> [diciembre 2014].
- Unión Latina - <http://unilat.org/> (consultado en mayo 2015).
- Union latine (2010): *Présence, poids et valeur des langues romanes dans la société de la connaissance, actes de la journée d'étude du 30 avril 2010* (sous la direction de Daniel Prado), Paris. Unión Latina
- W3Techs - http://w3techs.com/technologies/overview/content_language/all [diciembre 2014].
- Wikipedia, lenguas del mundo - https://es.wikipedia.org/wiki/Familia_de_lenguas [diciembre 2014].

ESTUDIOS / RESEARCH STUDIES

The impact of the economic crisis on Spanish university libraries

José Simón-Martín*, Alicia Arias-Coello**, Clara Simón-Blas***

*Instituto Universitario de Evaluación Sanitaria. Universidad Complutense de Madrid.

**Facultad de Ciencias de la Documentación. Universidad Complutense de Madrid.

***Departamento de Estadística e Investigación Operativa. Universidad Rey Juan Carlos. Madrid.
Correo-e: arcoello@gmail.com

Received: 05-11-2015; 2nd version: 14-12-2015; Accepted: 13-01-2016

Cómo citar este artículo/Citation: Simón-Martín, J.; Arias-Coello, A.; Simón-Blas, C. (2016). The impact of the economic crisis on Spanish university libraries. *Revista Española de Documentación Científica*, 39(3): e142. doi: <http://dx.doi.org/10.3989/redc.2016.3.1346>

Abstract: The scope of this study is to examine the impact of the financial crisis on the libraries of Spanish public universities. For this purpose, data from staff, finances, and services provided by the 47 public university libraries during 2008-2014 have been analyzed. Overall, the analysis of the data leads us to assert the following observations: these libraries have experienced a reduction in staff and expenditures for the purchase of information resources, whereas the volume of library lending has matched the evolution of student enrollment. Both interlibrary lending and the number of days that libraries were open during this period have decreased. Expenditures on electronic resources made by libraries and library consortia have increased in the period under study to the detriment of expenditures on monographs and printed journals. In this new economic environment, we consider that library managers should design actions for establishing strategies to ensure efficient and high-quality services to users.

Keywords: University libraries; economic recession; Spanish universities; financing libraries; library expenditure.

Impacto de la crisis económica en las bibliotecas universitarias españolas

Resumen: El presente trabajo tiene como objetivo examinar el impacto de la crisis financiera en las bibliotecas de las universidades públicas españolas. Para alcanzar este propósito se han analizado los datos de personal, los datos económicos, y los referidos a los servicios prestados por las 47 bibliotecas universitarias públicas durante el periodo 2008 a 2014. El análisis de los datos nos lleva a aseverar, a nivel global, las siguientes consideraciones: estas bibliotecas han experimentado una reducción en los efectivos de personal y en los gastos de adquisiciones de recursos de información, mientras que el volumen del préstamo domiciliario ha seguido una evolución similar al del número de estudiantes matriculados. Tanto el préstamo interbibliotecario como el número de días que permanecieron abiertas las bibliotecas durante estos años también han disminuido. El gasto en recursos electrónicos, realizado por las bibliotecas y los consorcios bibliotecarios, se ha incrementado en el periodo estudiado a expensas del gasto en monografías y en revistas en soporte papel. En este nuevo entorno económico, consideramos que los gestores bibliotecarios deben diseñar acciones orientadas al establecimiento de estrategias que aseguren una oferta de servicios eficientes y de calidad para los usuarios.

Palabras clave: Bibliotecas universitarias; crisis económica; universidad española; financiación bibliotecas; gasto bibliotecario.

Copyright: © 2016 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Attribution (CC BY) España 3.0.

1. INTRODUCTION

As a result of the world financial crisis, since the second quarter of 2008, Spain has experienced what economists have called the Big Recession (Bandrés and Gadea, 2013). This period has persisted in time, affecting the Autonomous Communities (ACs), which had their expenditures reduced by more than 20% on average between 2010 and 2012 (Lago Peñas and Fernández Leiceaga, 2013). This crisis situation has also been reflected in Spanish public universities, which experienced a 16% reduction in funding during the 2008-2014 period, according to data from the European University Association Observatory (2014). However, this funding reduction has not been homogeneous and has depended on the education and financial policies of each regional government.

In response to this context of general recession, both central and regional governments have initiated a series of actions with negative impacts on the operation of public universities, among which the following can be highlighted:

- In May 2010, five thousand million euros were cut from social public expenditures. Amid the items cut, there was a 5% reduction on average in staff salaries.
- The General State Budgets approved for the year 2012 established a 13.4 thousands million euros cut in public expenditures. Subsequently, on April 9, there was an additional 10 thousands million reduction, split between Education (3 thousand million) and Health (7 thousand million).
- The Organic Law 2/2012 of Budgetary Stability and Financial Sustainability was published on April 27, forcing universities to conform to the principles of budgetary stability and financial sustainability.
- The Royal Decree-Law 14/2012 of April 20 contained an urgent measure for public expenditure rationalization in the area of education. This decree included, among other things, the following measures:
 - Universities were urged to approve a maximum annual expenditure limit, with no possibility to exceed such limit.
 - Budgets and their executions should make explicit reference to their compliance with financial equilibrium and sustainability.
 - In the case of budget execution with a negative cash carry over, the Social

Council will have the power to compel the university to reduce it to the same extent as the deficit produced. This action should occur in the first session addressing expenditure reduction in the new budget. In this sense, it is important to note that transfers to universities charged to the budgets of an Autonomous Community require the approval and commencement of expenditure reduction.

- Increasing public undergraduate and master program fees.
- Staff costs should be authorized by the Autonomous Community. Appointments of interim staff and temporary staff to work in universities should comply with basic state regulations on the matter.

These measures were approved, and more measures related to the reduction of replacement rates for staff were added. Before 2010, it was 100 per 100, but from that year on, it was reduced by 10 per 100 per year through the year 2015, when it was set at 50%.

In Spain, the first direct mention of the economic crises in the university libraries context is found in the 2009 Memory of the Consortium of Catalonia University Libraries (Consortio de Bibliotecas Universitarias de Cataluña - CBUC). In this document, economic sustainability for the Catalan Digital Library is introduced as one of the challenges for the future, including the increase in product prices and the crisis situation. Its core concept is that the economic crisis has a negative impact on library funding. Gutiérrez Santana et al. (2010) agree with this premise, indicating that 33% of those in charge of university libraries in Andalusia already had their budgets reduced by the year 2010. This crisis environment has also impacted public libraries, as observed in the studies by Arroyo-Vazquez et al. (2011) and Villoldo et al. (2011), who pointed to the relative stagnation of the economic provisions assigned to Spanish libraries in 2010, which consequently had obvious reductions in funding. Indeed, only the libraries of higher education institutions preserved their funding in 2010, with a slight 2.2% increase with respect to 2008. Merlo Vega (2012) and Pérez-Salmerón (2014) describe the actions undertaken by Spanish libraries to preserve librarian services during the recession period. Similarly, Fernández and Rubio (2013), studying the relationship between the library budget and research productivity at the Polytechnic University of Valencia, indicate that the average budget assigned by university libraries to acquire information resources began to decline in

2010. Hernández-Sánchez and Arroyo-Vázquez (2014) hold the same position in a study that included all Spanish public libraries, highlighting the fact that the economic crisis in Spain has produced a reduction in acquisition expenditures, a decline in service sites, and staff shortages.

The consequences of this economic recession in Spanish university libraries have been clearly stated in the V BUCLE conference held on September 20-21, 2012. Among its conclusions, the following can be highlighted:

- The crisis is jeopardizing access to scientific information, with an impact on research results.
- As librarians, we have to cope with collection resizing, which should be oriented toward the actual use of the collection's contents. Only the preservation of collections suited to our researchers' needs will allow us to have economically sustainable library services. Continuing with an inflationary model based on exponential and unlimited content growth is no longer possible.

In the international context, the first documented reference regarding the impact of the economic crisis on university libraries was the publication of the findings of a survey of 886 librarians from different countries attending the Charleston Library Conference held in 2008. There, participants were required to reveal their opinions on the effect of the recession on their own libraries. The results indicated that 33.4% of university librarians stated their beliefs that budget cuts will be extended, at least to the year 2012 (Nicholas et al., 2010). In the US, Guarria (2011), in a study conducted on 1.442 university libraries, found that by 2009, 33.2% had already their budget reduced, with the number increasing to 43.7% in 2010. The same year, Guarria and Wang (2011) presented numerous examples of the effects of cuts on university library funding in the US.

Regarding Europe, Greece has been the country where the effects of this crisis have been the worst, with a negative impact on the operation of university libraries. Studies conducted by Kostagiolas and Margiola (2011) and Kostagiolas et al. (2013) reveal that the increase in the funding of university libraries implemented up to 2009 due to grants from the European Union was dramatically reduced during the 2010-2012 period by 50% for operational expenditures and 20% for staff expenditures. Giannakopoulos et al. (2014) indicate that in 2013, 85.2% of Greek university libraries had their funding reduced, producing an outcome that translated into the reduction

of information resources and maintenance expenditures and investments. Additionally, 79.1% of librarians considered the most negative effect of the crisis to be in the area of staff, with salary reductions, layoffs, and unfilled vacancies.

Ireland is another European country greatly affected by the economic crisis in the university library sector, with a 9% forced reduction in university staff during the 2008-2011 period and a 15% decline in the purchases of information resources (Cox, 2010a, 2010b).

Since 2009, British university libraries have experienced funding reductions in digital format journal purchases due to the depreciation of the pound against the euro and dollar, with effects in collection development (JISC, 2009; Kid, 2010; Harper and Corral, 2011). Similarly, Harper and Corral (2011) note a significant reduction in librarian staff funding.

In the Netherlands, funding cuts in public libraries started in 2009 (Nijboer, 2010), with a 6%-10% reduction in the following years, causing a decline in library staff in the libraries of the country's main cities.

Librarian consortia have also suffered impacts from the economic recession, as publicly acknowledged by the International Coalition of Library Consortia (ICOLC) in its declaration "The Global Economic Crisis and its Impact on Consortia Licenses" in January 2009 and again in June 2010. This declaration indicates that budgetary cuts to libraries and consortia would not be sporadic but significant, widespread, and lasting. Several proposals intended for consortia and publishers to preserve information context as much as possible are suggested in this declaration.

In Spain, the five university library consortia, two library associations, and the Spanish National Research Council Archive and Library Network initiated a consortium board in 2014, to coordinate a common strategy to negotiate with large journal publishers. In a document released by the board in 2015, this effort was highly commended, and the document specified in detail some of the strategies intended by the parties to negotiate with the large publishers.

In short, this period of recession has produced diverse effects: on one hand, a salary cut and, on the other hand, a dramatic reduction in funding both for the purchase of information resources and to meet operational expenses. In addition to the previous effects, there is the position of the large groups of journal and electronic resource publishers, which, despite the economic context, have continued to raise their prices.

In response to this contextual change, libraries have developed several strategies to maintain the same level and quality of service provision. The proposals that libraries have formulated can be classified into three major categories: first, those suggesting that the key is in librarians' professional development as a crucial factor for any change in strategy (Guarria and Wang, 2011; Varela Pardo and Baiget, 2012; Giannakopoulos et al., 2014); second, those holding that the key is to make their social role more visible and to establish links with society (Lloret Romero, 2011); and, third, those proposing the establishment of improvements at the level of resource management to face the cuts to funding for the purchasing of information resources in both the medium term and the long term. Kostagiolas et al. (2013) perform an excellent literature review on the different proposals from different researchers. Among them, and because of their potential application to the Spanish context, the following are worth highlighting:

- Optimizing the potential purchasing capacity through consortium conformation and realistic negotiation with providers.
- Library collaboration for the sharing of highly specialized information resources.
- Better exploitation of the information provided by analysis of the use of electronic publications databases and information packages with a high reoccurrence rate of periodic publications to improve decisions on subscription renewal. To this, it should be added that it would be key to take into account the opinion of interest groups regarding such decisions.
- Enhancing the use of open access resources.
- Reviewing and improving internal processes and, in collaboration with interest groups, redesigning new, more efficient methods for providing services.
- Establishing improvement programs for the efficient use of resources, such as computers, paper, energy, outsourcing, etc.

In the Spanish environment, the effort undertaken by several institutions to establish the relevance of libraries for society should be noted (Pérez-Salmerón, 2014).

In this paper, we first intend to identify and quantify the effect of the economic recession on Spanish public university libraries regarding the following parameters:

- Staffing.
- Expenditures on information resources.

- The volume of bibliographic material purchased.
- Library services, estimated through the volume of loan services and interlibrary loan requests and the number of days that libraries are open annually.

These data were compared with the effects of the economic crisis on funding, staffing, and the number of students enrolled in Spanish public universities.

Lastly, proposals made by Spanish university libraries to mitigate the effects of the economic recession were analyzed.

2. METHODOLOGY

This study takes the 47 Spanish public universities as reference. We should clarify that only on-site universities were considered. The library data used in this study have been collected from yearly statistical reports published by the University Libraries Network (Red de Bibliotecas Universitarias - REBIUN) corresponding to the years 2008 to 2014. With the aim of improving the data contained in this statistical network, we have first considered their yearly evolution to verify that there were no inconsistent figures in the series. Second, we have compared the data from REBIUN to the data collected from the different library reports, although we must note that this last approximation was conducted in only 25 cases of the total analyzed, given that they were the libraries with yearly activity reports. Lastly, data on expenditures have been verified in some occasions against the data collected in the universities' budget executions.

For librarian consortia, we considered it convenient to use the expenses incurred during the 2008-2013 period corresponding to the purchase of electronic resources. We obtained the data from the following sources:

- The Consortium of Galician University Libraries (Bugalicia). Data provided by Bugalicia.
- The Catalan University Libraries Consortium (Consorcio de Bibliotecas Universitarias de Catalunya - CCUC). Data obtained from its yearly activity reports and verified against the budgets approved annually by the Government of Catalonia
- The Andalusia University Libraries Consortium (Consorcio de Bibliotecas Universitarias de Andalucía - CBUA). Data obtained from its resource evaluation yearly reports.

- The Madrid University Libraries Consortium (Consortio de Bibliotecas Universitarias de Madrid - MADROÑO). Data officially provided by one of the consortium members.
- Regarding the Castilla León University Libraries Consortium (Consortio de Bibliotecas Universitarias de Castilla León - BUCLE). Consortium management has refused to provide data on its expenditures on information resources.

Regarding the universities, the data presented in this paper come from the following sources:

- Data on university approved budgets were collected from the universities' webpages and Autonomous Law Gazette.
- Data on the university staff were obtained from the series *Statistics on staff serving at universities*, published by the Ministry of Education, Culture, and Sports.
- Data on the number of students were obtained from the series *Statistics on students*, published by the Ministry of Education, Culture, and Sports.
- The data on university financial accounts come from the series *Statistics on Spanish public university funding and expenditures*, which registers the evolution from 2009 to 2012, prepared by the Ministry of Education, Culture, and Sports. For 2008, the publication *Spanish Universities in numbers (2010)*, by the Spanish University Council of Chancellors (Consejo de Rectores de las Universidades Españolas - CRUE, 2010), was used.

The measures adopted by libraries to face this period of crisis were obtained from two sources: their yearly library reports and the strategic planning documents found on their webpages.

3. RESULTS

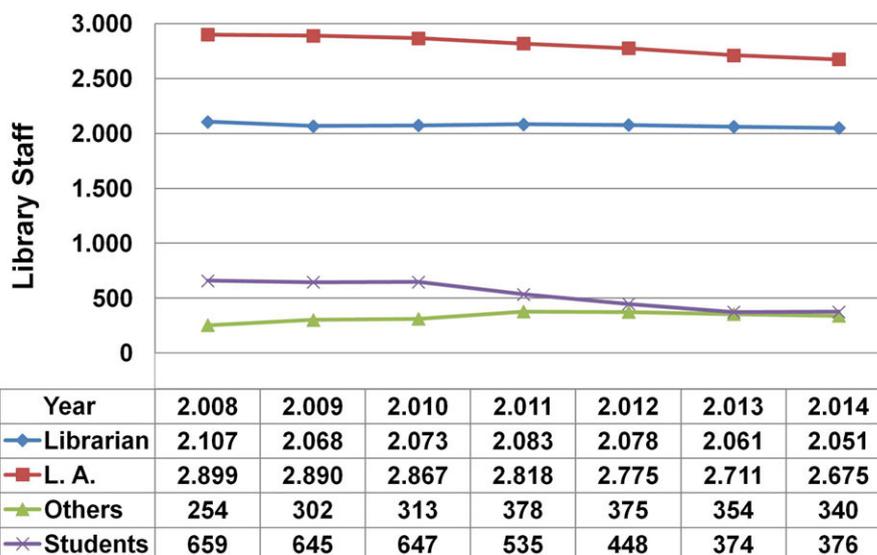
3.1. The evolution of staffing in Spanish university libraries

Figure 1 shows that the evolution of staff varies according to professional categories. The number of librarians and assistants has been progressively declining in the period between 2008 and 2014, ultimately resulting in a reduction of 2.7% for librarians and 7.2% for assistants. Regarding the category of "other staff", which comprises the clerical and specialized staff, should be highlighted a 33.9% increase in the period under study. Finally, the number of scholarships (standardized to full-time) registers a 42.9% reduction.

In absolute numbers, from 2008 to 2014, there was a net reduction of 477 personnel, with 56 being librarians, 224 assistants, and 283 students. Simultaneously, there was a reduction of 86 people in the category "other staff". Altogether, the permanent staff was reduced to 194 people, corresponding to 3.89% of the total number of personnel in 2008.

In the same period, this percentage is lower than the reduction in the teaching and research staff (4.4%) and similar to the reduction in managerial and service positions (3.97) at universities, as shown in the data presented in Table I.

Figure 1. Evolution of staffing in Spanish public university libraries



L. A.: Library Assistant.

Table I. Evolution of staffing in on-site Spanish public universities

Year	Teaching and research staff			Services and managerial staff		
	Total	Permanent*	Contracted**	Total	Administrative	Non-Administrative
2008/09	97.303	50.208	47.095	50.349	29.353	20.996
2009/10	99.203	50.098	49.105	52.083	30.625	21.458
2010/11	98.930	48.673	50.257	51.404	30.842	20.562
2011/12	97.662	48.267	49.395	50.663	30.835	20.428
2012/13	94.508	47.667	46.879	49.508	30.481	19.027
2013/14	93.000	46.348	46.040	48.351	30.060	17.990
%2013/08	-4.44	-7.7	-2.2	-3.97	2.4	-14.3

Sources: Statistics on university staffing (Ministry of Education, Culture, and Sports). Data from december in the first year of the period. (Permanent* include: Catedráticos and Titulares). (Contracted** include: Associated, Doctor Contracted and others)

3.2. The evolution of funding for the purchase of bibliographic material

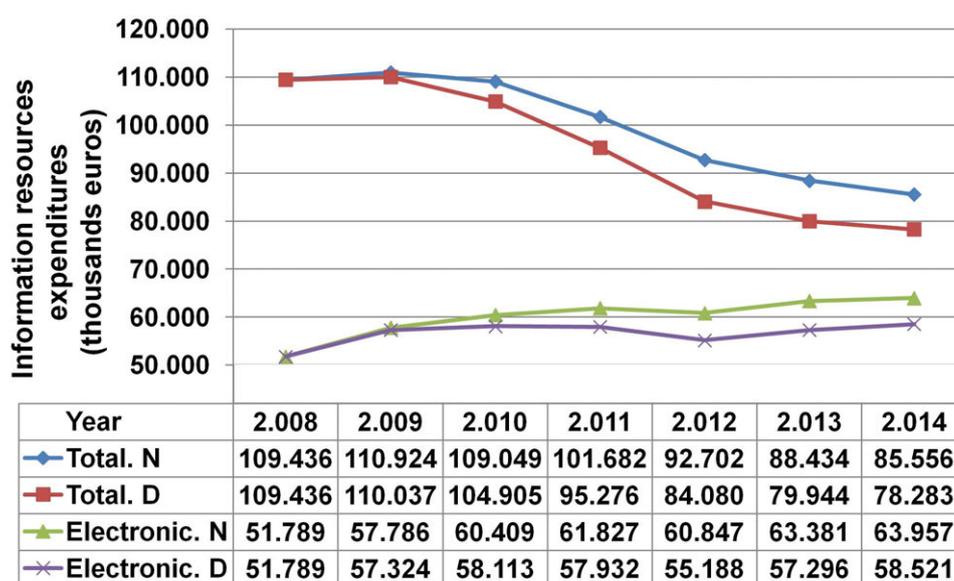
Figure 2 shows the evolution of library expenditures on bibliographic material purchases. Evidently, the total expenditures of university libraries increased in the year 2009 to decline afterward, both in constant euros (taking 2008 as reference) and in current euros, becoming more manifest starting in 2012.

This reduction in total expenditures of 21.82% in current euros and 28.47% in constant euros is coupled with a restructuring; thus, expenditures on

electronic format purchases grew in this period (see Figure 2), both at the nominal level (23.49%) and in constant euros (13.0%), to the extent that the percentage of electronic materials against the total went from 47.3% in the year 2008 to 74.7% in 2014.

In particular, this reduction in bibliographic product expenditures was uneven, given that it varied from a 10% increase to a 52% reduction at different libraries. This percentage difference depends on the Autonomous Community (AC) to which the university belongs. Figure 3 shows that university libraries situated in La Rioja, Asturias,

Figure 2. Library expenditures on information resources



Total. N: total expenditures in current euros. Total. D: total expenditures in constant euros. Electronic. N: expenditures on electronic resources in current euros. Electronic. D: same but in constant euros.

Madrid, Extremadura, Castilla-La Mancha, and Galicia ACs altogether experienced a reduction of more than 25%. By contrast, the university libraries in the País Vasco, Valencia, and Castilla y León ACs had a decrease of less than 15%.

Comparing the data from information resources expenditures against the data presented by public university budgets, we can observe the following facts (see Figure 4):

- First, the initial university budget increased up to the year 2010 and then started decreasing in 2011. This behavior is similar to the library expenditures on information resources. However, when we examine revenue recognition, that is, the real university income, we observe their downward trend starting in 2010.
- Second, the reduction in universities' initial budget global numbers from 2008 to 2014 in current euros was 9.4%. When the variation in the Consumer Price Index (CPI) is taken into account, the reduction reached 17.4%, a lower value than the reduction rate for bibliographic material expenditures.

- Third, as shown in Figure 4, the data from the approved 2015 budget show a slight change in the trend, with a growth of 1.7% compared to the previous year.
- Fourth, it should be emphasized that revenue recognition, or real university income, began to decline in 2010. These actual income values were very close to the income budgets approved by universities starting in 2012, when universities were forced to generate balanced budgets.

The reduction in university funding in this period was largely due to the reduction of AC subsidies for current expenditures. According to the data on the last budget execution published by the Ministry of Education, Culture, and Sports, in the 2008-2013 period, there was a decrease in the volume of subsidies of nearly 1.02 thousands million euros (15.8%), which was only partially compensated for with 300 million euros coming from the increase in registration fees, which is very similar to the numbers in the study conducted by Hernández Armenteros and Pérez García (2015) for the 2010-2013 period.

Figure 3. Percentage of reduction on average expenditures in current euros on bibliographic material during the 2008-2014 period by university libraries according to their regional location

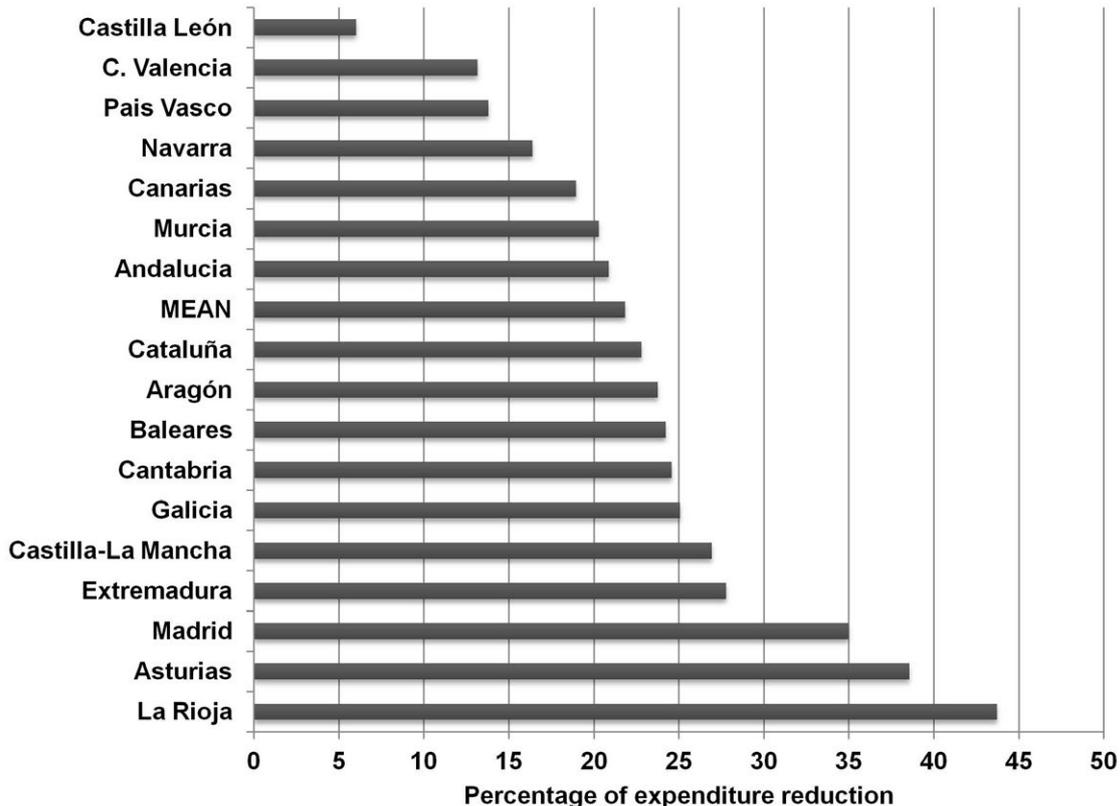
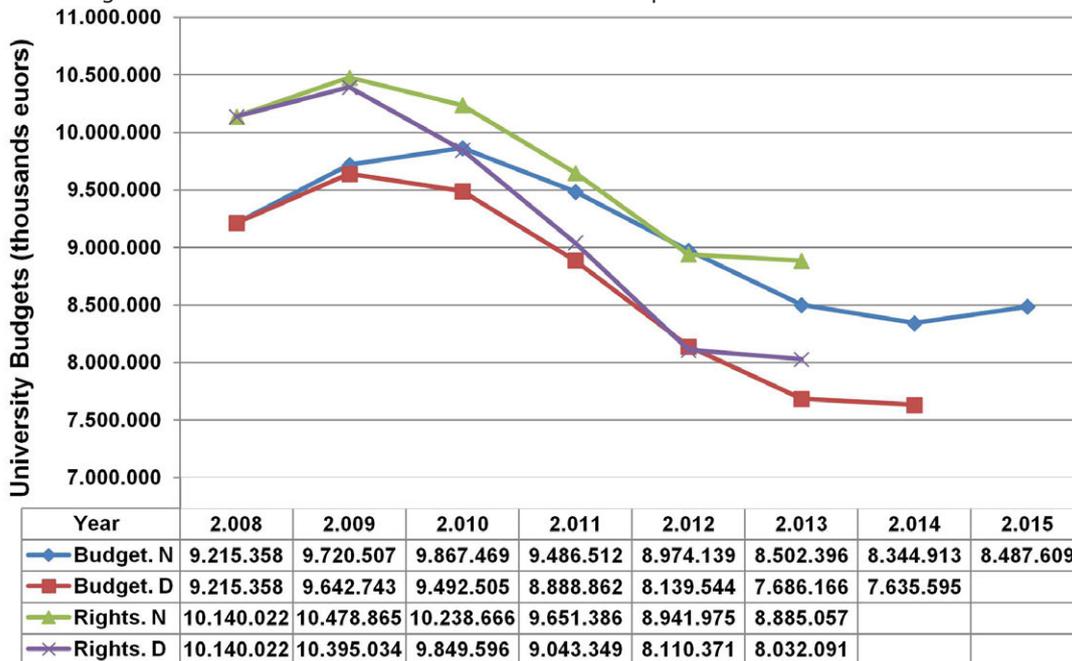


Figure 4. Budget evolution of initial and executed revenue of public on-site universities

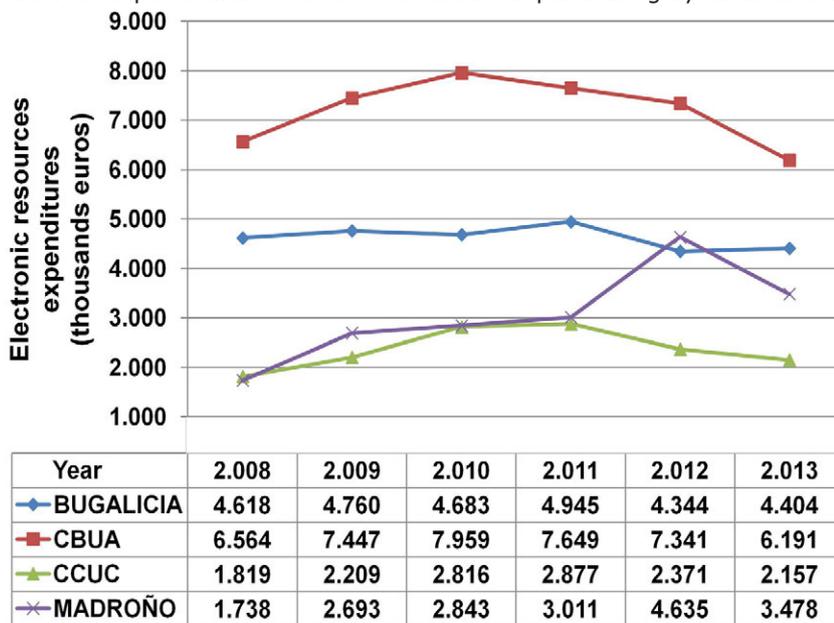


3.3. The evolution of consortia funding

In Spain, there are five university library consortia and two library associations that were created to consolidate the purchase of information resources. Figure 5 shows that expenditures by the library consortia on electronic resources started to decline in 2011. However, this decline was not the

same for each of the consortia; the consortium of Catalan universities had the largest reduction. Nevertheless, it should be noted that the Madroño Consortium shows a different trend during that period, experiencing an increase in expenditures in the year 2012. That year, the Madrid AC reduced the subsidies for the abovementioned consortium by half.

Figure 5. Evolution of expenditures on electronic resources purchasing by libraries consortia



BUGALICIA: The Consortium of Galician University Libraries. CBUA: The Andalusia University Libraries Consortium. CCUC: The Catalan University Libraries Consortium. MADROÑO: The Madrid University Libraries Consortium.

3.4. The evolution of bibliographic material purchase

3.4.1. Printed information resources

Figure 6 shows the number of monographs and printed journals purchased in the 2008-2014 period. The most relevant information is the continuous reduction in the number of issues purchased by libraries starting in 2008. As a result, there was a reduction of 47.0% for monographs and 52.8% for printed journals. This result matches the reduction in total information resources expenditures produced in that same period.

3.4.2. Electronic information resources

Figure 7 shows the evolution of the number of electronic monographs and journals purchased during the 2008-2014 period. The 131.3% increase in electronic book purchases in 2014 compared to 2008 is noteworthy, in addition to the 44.9% increase in electronic journals.

3.5. The evolution of the volume of services provided by libraries

As variables representing the service provided by libraries to their users, we have selected three activities: circulation, interlibrary loans, and the average number of days that libraries are open.

The first and third variables are related to student demand, whereas the second variable is related to faculty demand.

3.5.1. Circulation

Figure 8 presents the evolution of the volume of book loans in the 2008-2014 period. The number of loans increased 10.6% up to 2011, when a decline through 2014 started. The final result was a 0.7% reduction in the volume of loans reached in 2014 compared to 2008.

Because this type of loan is used mainly by students, it was considered convenient to relate these results to the evolution of the number of students enrolled in undergraduate courses during the same period. As shown in Figure 9, the evolution of the number of loans follows a similar course with respect to the number of students enrolled, with a maximum number of students in 2011, after which there is a decline of up to 4.8% compared to the year 2008.

3.5.2. Interlibrary loans

For this variable, we collected the total number of loan requests from libraries belonging to REBIUN, in addition to loans requested from libraries not belonging to REBIUN, and loans from foreign institutions. Figure 10 shows the evolution

Figure 6. Evolution of the number of printed monographs and journals acquired by university libraries

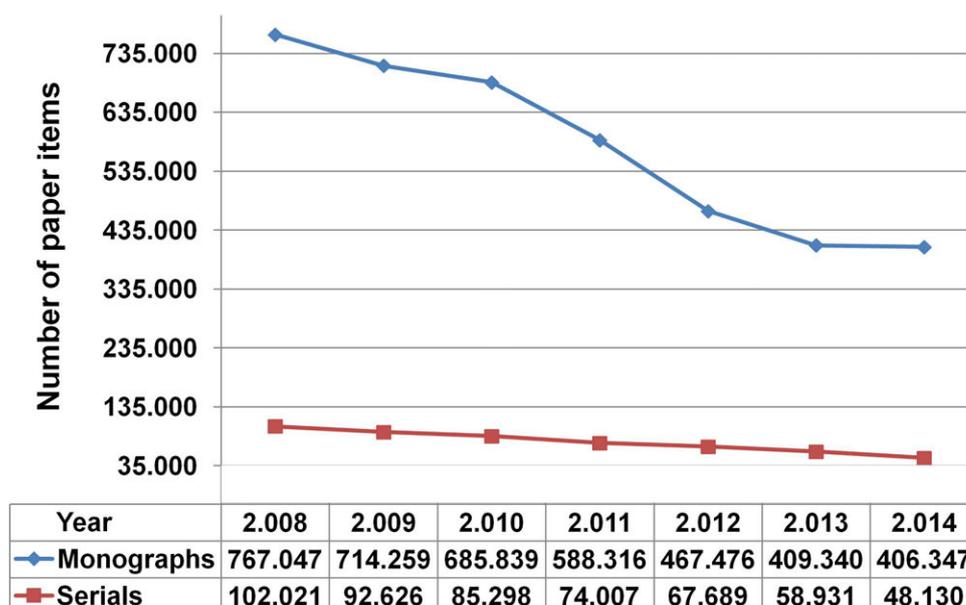


Figure 7. Evolution of electronic format resources purchasing

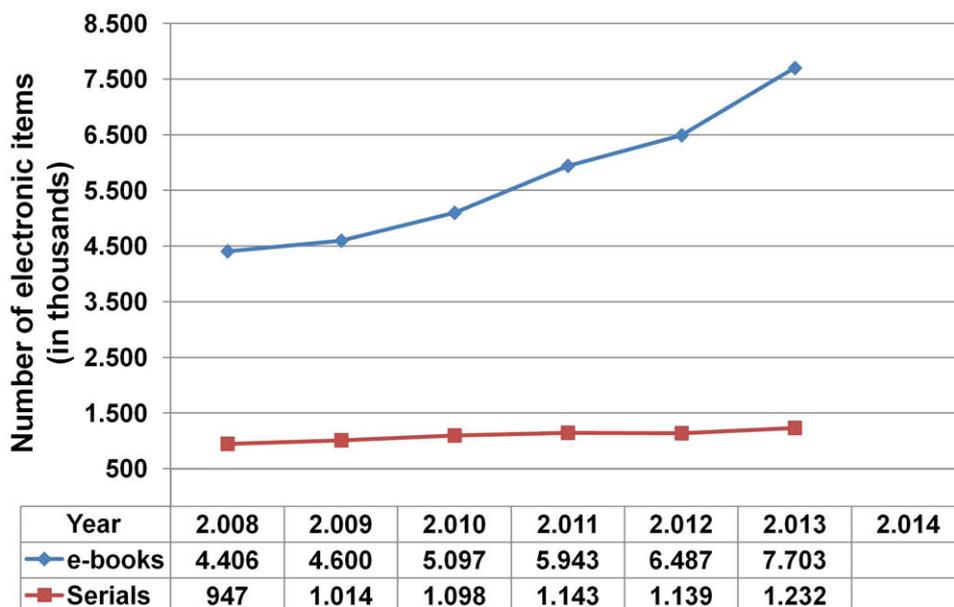
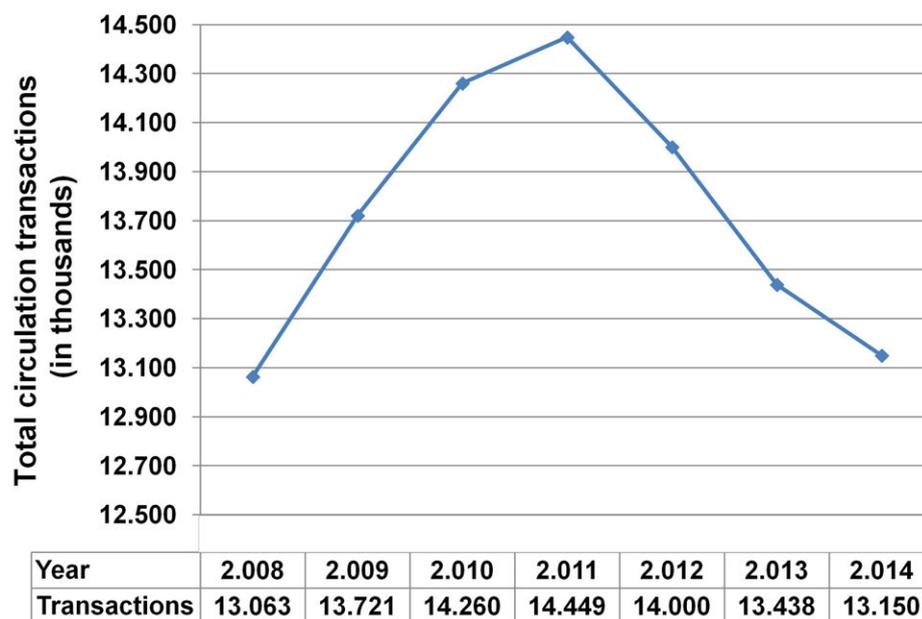


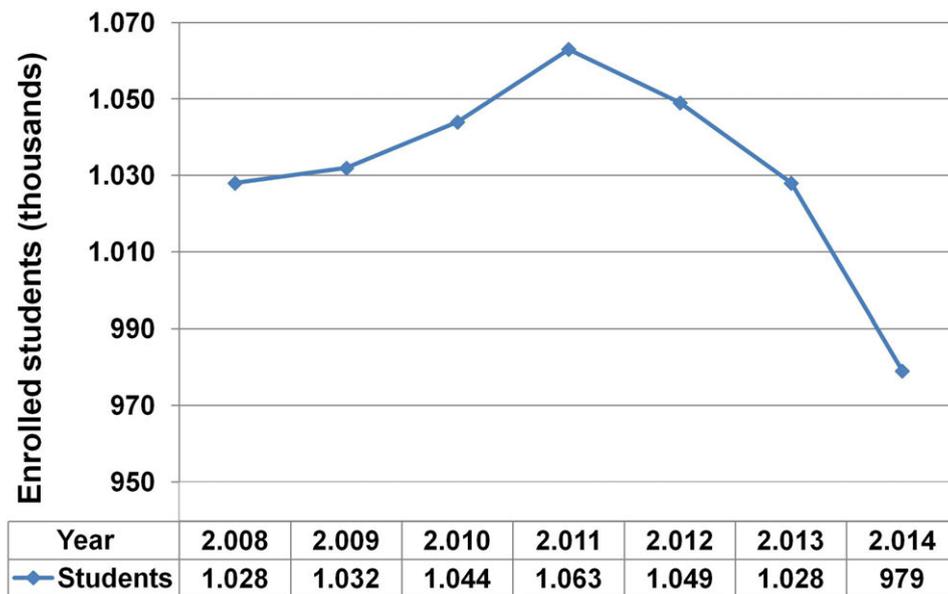
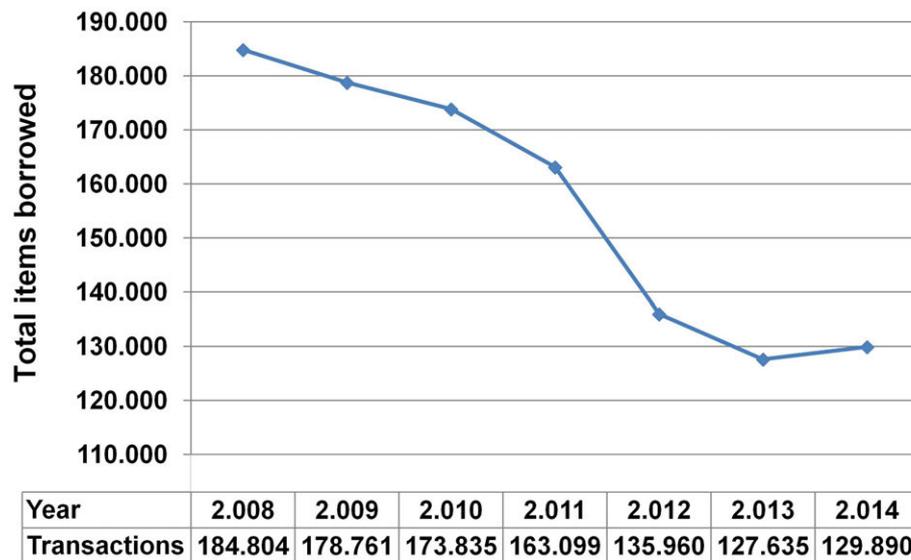
Figure 8. Evolution of book loan



of the number of loan requests for the 2008-2014 period. Evidently, there was a continuous decline in the number of requests, producing a decrease of 29.7% by the year 2014. This decline may be due to the enlargement of the electronic journal collection with direct access to users.

3.5.3 Yearly operation days

Only 38 libraries could be selected to analyze the evolution of this variable. This was due to the libraries' lack of understanding regarding this variable: some libraries provided information corresponding to the type of service that operates

Figure 9. Evolution of the number of students enrolled in undergraduate courses in Spanish public on-site university centers**Figure 10.** Evolution of interlibrary loan requests

the most days during the year, whereas other libraries provided the average number of days that all their services operate. For this study, we have chosen the second option, considering that it better represents the overall effort that libraries make.

Figure 11, which reflects the evolution of this variable, shows that there was a steady reduction in the number of days that libraries were open,

which, in 2014, was 4.05% fewer days of operation than in 2008. The implication is that in 2014, the 38 university libraries remained open an average of 254.8 days per year, whereas in 2008, the average was 267.8 days. Regardless, it must be noted that only 25 of the 38 libraries reduced the number of days that they were open during this period. This 10-day average reduction could be due to, among other causes, closings for holidays such as Easter,

Christmas, and summer vacations, which have been occurring in some universities over the last few years as a way to economize; and as the library at Carlos III University states on page 13 of its 2013 activities report: "During 2013, the cutback has continued by closing libraries on the weekends and suppressing night shifts during those periods".

3.6. Expression of the economic crisis in reports and the strategic planning developed by libraries

There are 20 university libraries with strategy documents in their webpages, of which only 11 have the period comprising the years 2010-2012 as the time when their planning went into effect. Regarding their yearly activity reports, only 25 libraries write such documents, with very uneven contents and to very uneven extents. With the scope of completing our analysis, the strategy documents produced by the consortia and library associations were also analyzed.

In these documents, the following aspects can be identified:

- References to the economic crisis and its impact on the library.
- Strategies or actions developed by libraries to reduce the effects of the economic crisis.

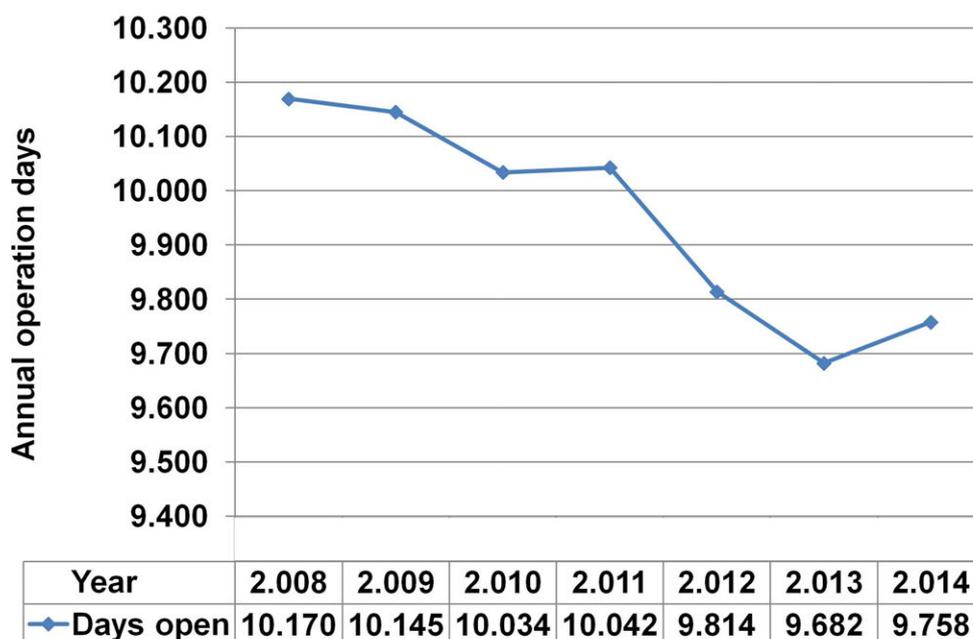
In the 2009 reports written by the Catalanian Library Consortium (Consortio de Bibliotecas Universitarias de Cataluña - BDC), for the first time

in Spain, here is a clear reference to the fact that an economic crisis will negatively affect Spanish university libraries. For the first time, the idea that the economic sustainability of BDC will be one of the challenges for the future, due to the increase in product prices and the noted crisis, emerges in this document. A year later, two other libraries, those of the University of La Rioja, in its 2010-2015 Management Plan, and Complutense University of Madrid, in its 2010-2013 Strategic Plan, precisely identify the degraded economic scenario as one of the contextual threats. It should be noted as a differential fact that 8 university libraries that, in most cases, initiated their strategic plans in 2012 do not include in detail the topic of the economic crisis, although by that time the effects of the crisis on their funding and on the hiring of staff were evident.

Among the goals that these libraries proposed to achieve and that are included in their strategic planning, the following can be highlighted:

- Attempting to obtain external resources.
- Optimizing collection expenditures by implementing actions toward, on the one hand, evaluating the use patterns of periodic publications, with the aim of adjusting their volume to their actual demands, and, on the other hand, improving purchase processes for bibliographic resources through actions such as creating centralized purchasing, certifying suppliers, or modifying the basis of negotiations with them.

Figure 11. Evolution of yearly library days of operation



- Redirecting collection development through the reduction in subscriptions to printed journals or the purchase of monographs in electronic, instead of printed, format.
- Increasing cooperation among libraries by supporting the creation of library associations and consortia to improve their negotiation capacity against suppliers of information products.
- Improving the work atmosphere, as the University of Valladolid library proposes.

4. CONCLUSIONS

Overall, the Spanish university libraries that belong to public universities began to be aware of the effects of the general economic recession in 2009, although in most, it did not become evident until 2012, the year when their funding was dramatically reduced.

During the long period of economic recession, from 2008 to 2014, libraries suffered a set of negative consequences that, in this study, we have grouped into the following categories: staff, the acquisition of information resources and collection development, and services provided to users.

Regarding staff and similar to what has happened in other countries, in this period, Spain experienced the loss of librarian and assistant positions due to the legal prohibition against filling vacancies. In addition and linked to this issue, it could be determined that libraries suffered a dramatic reduction in the number of students with scholarships as a consequence of cuts to funding, contributing to the loss of buying power of staff due to salary reductions.

The consequences in the domain of information resources purchasing are determined by reductions in budgets assigned to libraries by the universities and, perhaps, by the reduction in contributions from research funds and the centers themselves. Similarly, the consortia have reduced their budgets for electronic resources.

Expenditures on the acquisition of bibliographic materials have fallen more than 28% in constant euros (taking the year 2008 as the base). However, these reductions have not occurred in all libraries but have depended on the Autonomous Communities in which the university is located. The consequences of this decline in funding have had a major effect on the maintenance and development of collections, especially printed material. In light of this situation, managers have taken action by prioritizing the development of electronic

resources, especially electronic periodicals, which are essential for developing university research. However, the impact on students caused by the reduction in the purchase of printed monographs remains unknown. These users have not yet changed their habits toward the use of monographs in electronic format, in addition to the fact that in many disciplines, collections of manuals in electronic format are poorly developed.

With respect to the volume of librarian services rendered during this period, it could be proven that first, the volume of book loans decreased in 2014 compared to 2008. However, the declining trend that started in 2011 should be analyzed to determine its causes. Although the data presented suggest that the cause could be due to a decrease in the number of students, we think that a more detailed study should be conducted, given that also since 2011, a decrease in the purchase of both printed monographs and manuals used by students can be observed. The effect of the electronic support platforms for teaching with the availability of a large volume of information to students should also be analyzed. Second, the declining evolution presented by the volume of interlibrary loan requests during this period could be because researchers have increased their access to the electronic publications acquired by their universities, consortia, or even to the growth of open access publications. Regarding the number of days that libraries are open annually, the reduction produced in this period, although not large, could nevertheless be a trend maintained over time as a consequence of the efficiency programs that universities have implemented since 2011.

The data on the budgets approved by universities in the years 2013 to 2015 suggest that the situation of continuous reduction in university funding may have come to an end. However, macroeconomic data suggest that the funding recovery could be very slow because of, among other factors, the volume of Spanish public debt and the goals for the budget deficit established by the European Union for the coming years. This situation suggests to us a scenario of university funding with small reductions, which means that library managers should design and develop strategies, on the one hand, to mitigate the negative effects of cuts on staff and compensation, the reduction in operational expenditures, and the purchase of information resources and, on the other hand, to maintain the levels and quality of service.

Simultaneously, library managers should address current challenges such as their participation in and maintenance of virtual teaching or internet competition. They should also think about their role,

which, up to this moment, has always been that of facilitating information. At this point in technological development, that role has expanded, producing the power to act proactively with users. Because of the amount of information niches available to work, they should be capable of offering users current and precise information for the generation of high-quality knowledge. Additionally, social networks and the crisis itself have revealed that to achieve efficiency, it is indispensable to cooperate with other entities in the integration of and access to resources, etc.

6. REFERENCES

- Arroyo-Vázquez, N; Hernández-Sánchez, H; Villoldo, F. J. (2011). Bibliotecas y crisis económica (I). Las bibliotecas españolas en 2010. *Boletín de la Asociación Andaluza de Bibliotecarios*, vol. Junio-Diciembre, 181-186.
- Bandrés, E; Gadea, M. D. (2013). Crisis económica y ciclos regionales en España. *Papeles de Economía Española*, vol. 138, 2-30.
- BUCLE.V Jornadas (2012). Conclusiones. Valladolid (2012). <http://uvadoc.uva.es/bitstream/10324/1550/1/conclusiones-BUCLE2012.pdf>. [accessed January/2015].
- Consorcio de Bibliotecas de Cataluña. (2010). Memòria d'activitats del 2009. <http://www.csuc.cat/sites/default/files/docs/memo09.pdf>. [accessed December/2015].
- Cox, J. (2010a). Sharing the pain, striving for gain. *Serials*, vol 23 (1), 12-15. <http://dx.doi.org/10.1629/2312>
- Cox, J. (2010b). Academic Libraries in Challenging Times. *An Leabharlann: The Irish Library*, vol. 19(2), 7-13.
- CRUE (2010). Financiación y gasto en 2008. La Universidad Española en cifras (2010). http://www.crue.org/Documentos%20compartidos/Publicaciones/Universidad%20Espa%C3%B1ola%20en%20cifras/UEC_2010.pdf. [accessed January/2015].
- European University Association Observatory. (2014). *EUA Public Funding Observatory 2014*. http://www.eua.be/Libraries/Governance_Autonomy_Funding/PFO_analysis_2014_final.sflb.ashx. [accessed February/2015].
- Fernández, S.; Rubio, F. (2013). ¿El dinero importa? Relación entre el presupuesto de la biblioteca y la productividad investigadora de la Universitat Politècnica de València. *Revista Española de Documentación Científica*, 36(4):e023. <http://dx.doi.org/10.3989/redc.2013.4.1043>
- Ultimately, the consequences of the economic crises have revealed, among other issues, the need for library managers, along with academic authorities, to commence a process of debate on what should be the library mission in this new economic and social environment. They should work on enhancing their capacities for planning, innovation, and cooperation to establish proactive actions to become sustainable and to continue improving the quality of the services rendered to users.
- ## 5. ACKNOWLEDGEMENTS
- American Journal Experts for the English translation of this work.
- Guarria, C. I. (2011). The recession, budgets, expectations and realities. *The Bottom Line: Managing Library Finance*, vol. 24 (4), 200-217. <http://dx.doi.org/10.1108/0888045111193299>
- Guarria, C. I; Wang, Z. (2011). The economic crisis and its effect on libraries. *New Library World*, vol. 112 (5/6), 199-214. <http://dx.doi.org/10.1108/03074801111136248>
- Giannakopoulos, G; Koulouris, A; Kokkinos, D. (2014). Libraries in Crisis: A Glimpse over Greece and Cyprus. *Procedia - Social and Behavioral Sciences*, vol. 147, 411-417. <http://dx.doi.org/10.1016/j.sbspro.2014.07.121>
- Gutiérrez Santana, F; Real Duro, A; Bustamante, A. T; Guerrero Salguero, C. (2010). Estudio sobre el impacto de la crisis económica en las bibliotecas andaluzas. *Boletín de la Asociación Andaluza de Bibliotecarios*, vol.100 (Julio-Diciembre), 119-136.
- Harper, R; Corral, S. (2011). Effects of the economic downturn on academic libraries in the UK: Positions and Projections in Mid-2009. *New Review of Academic Librarianship*, vol.17 (1), 96-128. <http://dx.doi.org/10.1080/13614533.2011.554103>
- Hernández-Sánchez, H; Arroyo-Vázquez, N. (2014). Efectos de la crisis económica en las bibliotecas españolas. *El profesional de la información*, vol. 23 (2), 158-164. <http://dx.doi.org/10.3145/epi.2014.mar.08>
- Hernández Armenteros, J; Pérez García, J. A. (2015). La Universidad Española en cifras 2013-2014. CRUE. Universidades Españolas. Madrid. http://www.crue.org/Documentos%20compartidos/Publicaciones/Universidad%20Espa%C3%B1ola%20en%20cifras/UEC_13-14.pdf. [accessed July/2015]
- International Coalition of Library Consortia (ICOLC). (2009). Statement on the global economic crisis and its impact on consortium licenses. <http://icolc.net/statement/statement-global-economic->

- crisis-and-its-impact-consortial-licenses. [accessed March/2015]. Una traducción al español realizada por el Consorcio de Bibliotecas de Cataluña se puede encontrar en: http://www.recercat.cat/bitstream/handle/2072/68139/ICOLC2010_Spa.pdf?sequence=1. [accessed March/2015].
- JISC (2009). "The impact of the economic recession on university library and IT services". <http://www.webarchive.org.uk/wayback/archive/20140615233306/http://www.jisc.ac.uk/media/documents/publications/libsitimpacts.pdf>. [accessed June/2015].
- Kidd, T. (2010). The view from the UK: the economic crisis and serials acquisitions on an offshore island. *The Serials Librarian*, vol. 59 (3), 384-93. <http://dx.doi.org/10.1080/03615261003623138>
- Kostagiolas, P; Margiola, A. (2011). A library management response model against the economic crisis. *Library Review*, vol. 60 (6), 486-500. <http://dx.doi.org/10.1108/00242531111147206>
- Kostagiolas, P; Papadaki, E; Kanlis, G; Papavlasopoulos, S. (2013). Responding to crises with alliances: evidence from an academic library survey in Greece. *Advances in Librarianship*, vol. 36, 247-279. [http://dx.doi.org/10.1108/S0065-2830\(2013\)0000036012](http://dx.doi.org/10.1108/S0065-2830(2013)0000036012)
- Lago Peñas, S; Fernández Leiceaga, X. (2013). Las finanzas autonómicas: Expansión y crisis, 2002-2012. *Papeles de Economía Española*, vol. 138, 129-146.
- Lloret Romero, N. (2011). Libraries' response to the crisis: measures to mitigate its impact. *The Bottom Line*, vol. 24 (4), 236-240. <http://dx.doi.org/10.1108/08880451111193352>
- Merlo Vega, J.A. (2012). Acción ante la recesión: el compromiso de las bibliotecas con los ciudadanos. *Boletín de la Asociación Andaluza de Bibliotecarios*, vol. 104 (Julio-Diciembre), 65-91.
- Mesa de Consorcios. (2015). Valoración de las acciones de coordinación entre consorcios y grupos de compras de bibliotecas universitarias y científicas españolas en 2014. http://www.consorcioadrono.es/info/web/blogs/blog_madro/213.php. [accessed March/2015].
- Ministerio de Educación, Cultura y Deporte. Fuentes: Estadística de personal al servicio de las universidades. <http://www.mecd.gob.es/educacion-mecd/areas-educacion/universidades/estadisticas-informes/estadisticas/personal-universitario.html>. [accessed January/2015].
- Ministerio de Educación, Cultura y Deporte: Estadística de financiación y gasto de las universidades públicas españolas (2009-2012). <http://www.mecd.gob.es/educacion-mecd/areas-educacion/universidades/estadisticas-informes/estadisticas/estadistica-de-financiacion-y-gasto-universitario.html>. [accessed January/2015].
- Nicholas, D.; Rowlands, I.; Jubb, M.; Jamali, H. R. (2010). The impact of the economic downturn on libraries: With special reference to university libraries, *Journal of Academic Librarianship*, 36 (5), 376-382. <http://dx.doi.org/10.1016/j.acalib.2010.06.001>
- Nijboer, J. (2010). How to survive the recession as a cultural entrepreneur. Paper presented at the Bobcatsss Conference, Parma, 25-27 January. Available: <http://dSPACE-unipr.cilea.it/bitstream/1889/1221/1/Nijboer.pdf>
- Pérez-Salmerón, G. (2014). Librarians facing the crisis: Action. IFLA WLIC 2014, 16-22 August 2014, Lyon, France. <http://library.ifla.org/904/1/167-salmeron-en.pdf>. [accessed February/2015].
- REBIUN. Red de Bibliotecas Universitarias Españolas. Consulta de Datos Estadísticos. Available: <https://rebiun.um.es/rebiun/admin/ManageIndicatorsPage;jsessionid=1BBF1CC4438F153FC63B584DC23E9CFD>. [Accessed 2015].
- Universidad Carlos III de Madrid-Biblioteca. (2014). Biblioteca. Memoria de actividades 2013. http://e-archivo.uc3m.es/bitstream/handle/10016/19631/memoria_biblioteca_2013.pdf?sequence=1. [accessed January/2015].
- Villoldo, F.J; Hernández-Sánchez, H; Arroyo-Vázquez, N. (2011). Bibliotecas y crisis económica (II). Las bibliotecas públicas en 2010: el desarrollo de la última década, en peligro. *Boletín de la Asociación Andaluza de Bibliotecarios*, vol. Junio-Diciembre, 187-191.
- Varela-Prado, C; Baiget, T. (2012). El futuro de las bibliotecas académicas: incertidumbres, oportunidades y retos. *Investigación Bibliotecológica*, vol. 26(56), 115-135.

APPENDIX. Libraries included in this article

Autonomous Communities	Universities libraries
ANDALUCÍA	Universidad de Almería
	Universidad de Cádiz
	Universidad de Córdoba
	Universidad de Granada
	Universidad de Huelva
	Universidad de Jaén
	Universidad de Málaga
	Universidad Pablo de Olavide
	Universidad de Sevilla
ARAGÓN	Universidad de Zaragoza
ASTURIAS	Universidad de Oviedo
CANARIAS	Universidad de La Laguna
	Universidad de Las Palmas de Gran Canaria
CANTABRIA	Universidad de Cantabria
CASTILLA-LA MANCHA	Universidad de Castilla-La Mancha
CASTILLA-LEÓN	Universidad de Burgos
	Universidad de León
	Universidad de Salamanca
	Universidad de Valladolid
CATALUÑA	Universidad de Autónoma de Barcelona
	Universidad de Barcelona
	Universidad de Gerona
	Universidad de Lleida
	Universidad Politécnica de Cataluña
	Universidad Pompeu Fabra
	Universidad Rovira i Virgili
COMUNIDAD VALENCIANA	Universidad de Alicante
	Universidad Jaume I
	Universidad Miguel Hernández
	Universidad Politécnica de Valencia
	Universidad de Valencia
EXTREMADURA	Universidad de Extremadura
GALICIA	Universidad de Coruña
	Universidad de Santiago de Compostela
	Universidad de Vigo
ISLAS BALEARES	Universidad de Las Islas Baleares
LA RIOJA	Universidad de La Rioja
MADRID	Universidad de Alcalá de Henares
	Universidad de Autónoma de Madrid
	Universidad Carlos III
	Universidad de Complutense de Madrid
	Universidad Politécnica de Madrid
	Universidad Rey Juan Carlos

Autonomous Communities	Universities libraries
MURCIA	Universidad de Murcia
	Universidad Politécnica de Cartagena
NAVARRA	Universidad Pública de Navarra
PAÍS VASCO	Universidad del País Vasco

ESTUDIOS / RESEARCH STUDIES

Influencia del programa Campus de Excelencia Internacional en la posición de las universidades españolas en el ranking de Shanghái

Teodoro Luque-Martínez*, Luis Doña-Toledo* y Domingo Docampo**

*Universidad de Granada, CEI BioTic.

**Universidade de Vigo, Atlantic Research Center for ICT.

**Charles Darwin University, Northern Institute, Ellengowan Drive, Brinkin NT 0909 Australia
Correo-e: tluque@ugr.es

Recibido: 09-10-2015; 2ª versión: 16-12-2015; Aceptado: 13-01-2016.

Cómo citar este artículo/Citation: Luque-Martínez, T.; Doña, L.; Docampo, D. (2016). Influencia del programa Campus de Excelencia Internacional en la posición de las universidades españolas en el ranking de Shanghái. *Revista Española de Documentación Científica*, 39(3): e143. doi: <http://dx.doi.org/10.3989/redc.2016.3.1339>

Resumen: En 2009 nace la iniciativa Campus de Excelencia Internacional (CEI) entre cuyos objetivos estaba mejorar la posición de las universidades españolas en los rankings internacionales con un horizonte de 2015. Este trabajo analiza el impacto de los CEI en la clasificación ARWU o ranking de Shanghái. Comparando la media de los tres años del inicio (2009-2012) y del final (2013-2015), el sistema universitario español ha mejorado sensiblemente su prestación investigadora a juzgar por el incremento en puntuación media y en posiciones. Las universidades coordinadoras de proyectos CEI presentan una evolución más estable (aumento del 9%) que las no coordinadoras (aumento del 21%) resultando un sistema universitario más homogéneo. Las universidades con proyectos CEI con una sola universidad alcanzan mejores posiciones. A pesar del contexto de crisis, escaso tiempo y financiación de la iniciativa CEI, la experiencia ha resultado positiva a la vista de la evolución de las universidades españolas en dicho ranking.

Palabras clave: Campus de Excelencia Internacional; sistema universitario español; indicadores; clasificaciones internacionales; ARWU; ranking de Shanghái.

Influence of the Campus of International Excellence initiative on the position of Spanish universities in Shanghai's academic ranking

Abstract: The Campus of International Excellence (CIE) initiative was launched in Spain in 2009. Among its objectives was to improve the position of Spanish universities in international rankings by 2015. The present work analyses the impact of CIEs on the ARWU-Shanghai Academic Ranking. The comparison between the average across the first three years of the CIE initiative (2009-11) and the last three (2013-15) reveals that the Spanish university system has notably improved its research performance, judging by the increase in average scores and the hike in positions. Universities coordinating CIE projects present a more stable evolution (an increase of 9% in scientific output) than those that don't (increase of 21%), resulting in a more homogeneous university system. Institutions with CIE projects comprising a single university achieve better positions in the rankings. Despite the context of the economic crisis, a tight timeframe and limited funding for the CIE initiative, the experience is found to have been positive in view of the evolution of Spanish universities in these rankings.

Keywords: CEI Initiative; Higher Education System; Spain; indicators; international classifications; ARWU; Shanghai ranking.

Copyright: © 2016 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Attribution (CC BY) España 3.0.

1. INTRODUCCIÓN

La Estrategia Universidad 2015 (Ministerio de Educación, 2011) impulsó la convocatoria del programa de Campus de Excelencia Internacional, por la que se concedieron ayudas con la finalidad de que los campus universitarios españoles adquieran un nivel de excelencia e internacionalización que les permita constituir un referente en el ámbito internacional. *El horizonte establecido fue el año 2015. Este era el plazo para reforzar el papel de la universidad como motor de la nueva sociedad basada en el conocimiento, por medio de una estrategia que se hacía eco de las clasificaciones de universidades como herramienta de benchmarking*, de modo que se reconocía como objetivo fundamental mejorar la posición de las universidades españolas en los rankings internacionales.

Alcanzado dicho horizonte, y a pesar de las vicisitudes de la iniciativa, vinculadas a una época de profunda crisis económica, cabe preguntarse si las universidades españolas han mejorado efectivamente sus posiciones en los rankings internacionales y si, en particular, lo han hecho y hasta qué punto las vinculadas a los 32 proyectos de Campus de Excelencia Internacional, en los que han participado casi todas las universidades públicas españolas y algunas de las privadas. A esta cuestión intenta responder el presente trabajo, en el que analizamos en qué medida las universidades que lideran proyectos de Campus de Excelencia Internacional han mejorado sus posiciones en el contexto internacional. Hemos tratado también de analizar si esta variación en el ranking podría estar relacionada con el carácter de ser universidad pública o privada, ser coordinadora de proyecto CEI o no, o bien con el hecho de formar parte de proyectos CEI en los que participen una o más universidades.

Dado que los Campus de Excelencia Internacional han sido decididos en función de la capacidad investigadora de las universidades integrantes y de la calidad del proyecto investigador presentado, hemos elegido entre las clasificaciones globales aquella que está basada casi fundamentalmente en la producción investigadora y el reconocimiento científico internacional por medio de los niveles de citas, el *Academic Ranking of World Universities* (ARWU), también conocido como ranking de Shanghái, que ya ha sido utilizado con anterioridad en el análisis de los sistemas universitarios y sus retos (Docampo, 2011; Erkkilä, 2013).

Hasta el año 2013 se creía que los resultados del ranking de Shanghái eran irreproducibles, por lo que hubiese sido imposible obtener la información necesaria para el análisis planteado en el trabajo. Gracias a la aportación de uno de

los autores (Docampo, 2013) se ha desarrollado una completa metodología que permite estimar con un elevado nivel de precisión los resultados de ARWU en cualquiera de sus ediciones desde 2003 hasta 2015. Utilizando dicha metodología podemos obtener las puntuaciones de 3500 instituciones de educación superior a nivel global: prácticamente todas las universidades con una producción científica anual de al menos 100 contribuciones indexadas. Esos cálculos nos permitirán evaluar la posición en la que las universidades se encontraban antes de la iniciativa CEI y la que ocupan en los tres últimos años.

El trabajo se estructura del siguiente modo. En primer lugar revisamos algunos antecedentes de mejora en sistemas universitarios a lo largo del mundo. Luego exponemos las principales características del ranking de Shanghái, para, a continuación, explicitar detalladamente la metodología que hemos seguido en nuestro trabajo para la recogida de los datos de las universidades españolas. La sección central de nuestro trabajo está centrada en el análisis de las cuestiones que fueron formuladas en la introducción. Una sección en la que damos cuenta de las limitaciones de nuestro estudio al tiempo que incluimos nuestras conclusiones y recomendaciones da remate al trabajo.

2. INICIATIVAS DE MEJORA EN LAS UNIVERSIDADES

La universidad del futuro debe plantearse si es una institución en la sociedad o para la sociedad, como sugiere Barnett (2012), lo que implica no solamente constituirse en un reflejo de la sociedad sino erigirse en un agente activo para mejorarla, en definitiva, para construir una buena sociedad. Para ello se precisan universidades que desarrollen acertadamente sus misiones, siendo en general las universidades que ostentan buenos niveles de investigación y docencia (que suelen ir de la mano) las que también presentan buenos resultados en su conexión con la sociedad, al extender su influencia al conjunto de la misma (Shattock, 2010).

Estos retos se han abordado en los sistemas universitarios de todo el mundo tratando de organizar la búsqueda o el camino hacia la excelencia en las universidades, como se recoge en el análisis realizado por Altbach y Salmi (2011), en el que estudian casos concretos de universidades de todos los continentes que aspiran a ser calificadas como universidades de clase mundial. En ellas hay tres factores clave: una alta atracción de talento de profesores y estudiantes, abundantes recursos para ofrecer un aprendizaje adecuado y llevar a cabo investigaciones avanzadas, y características favorables de

governabilidad (Salmi, 2009). Paralelamente, ha crecido la necesidad de medir y cuantificar si se ha alcanzado esa categoría y qué posición se ocupa en la misma, para lo cual se han desarrollado diferentes procedimientos y rankings, no exentos de críticas como indica Hazelkorn (2011).

Precisamente, entre los planteamientos de la iniciativa CEI, además de la búsqueda de la excelencia en docencia e investigación y de la internacionalización, se encuentra el reconocimiento de la necesidad de una buena conexión universidad-entorno. Esto implica el reconocimiento de las consecuencias que tiene la investigación en el territorio (Kretz y Sá, 2013) y, en definitiva, la asunción de los postulados de la triple hélice de Etzkowitz y Leydesdorff (1998), añadiendo un cuarto factor (medios, cultura y sociedad civil), incluso un quinto (énfasis en el entorno natural) según Carayannis y Campbell (2011). En suma, la iniciativa CEI pretendía servir de estímulo para las misiones básicas de la universidad y también para dinamizar el desarrollo social y económico del territorio. De alguna forma constituye un antecedente que coincide con los planteamientos de la estrategia de la Unión Europea de especialización de los territorios denominada *Research and Innovation Strategy for Smart Specialisation* o RIS3 (2013).

Iniciativas similares han tenido lugar por todo el mundo desde la década de los años 1980. Salmi (2015a) analiza las experiencias existentes en todos los continentes (véase un resumen en la tabla I), que pueden afectar a las universidades en su conjunto o a centros de excelencia especializados, prestando atención a su grado de financiación, que ha sido desigual según los países. Este autor destaca cuantitativamente la apuesta de países como China, Francia, Singapur o Taiwán, en los que reconoce la influencia que los rankings internacionales han tenido tanto en la promoción de las iniciativas como en la determinación de los gobiernos para llevarlas a cabo. Es por ello que advierte acerca del

riesgo de sobre-enfatizar la investigación, debido a la influencia de los rankings internacionales, con la posible consecuencia de fomentar el elitismo en lugar de la mejora institucional, provocando con ello una mayor desigualdad en la educación superior, ignorando de facto posibles restricciones académicas y democráticas en algunos países.

Por otra parte, Delgado (2015) analiza las iniciativas de excelencia en 9 países de Europa y Asia, con programas que reciben financiación desde 65 a 2.700 millones de euros. Aun señalando que no hay un modelo común de universidad de excelencia y que cada programa presenta sus peculiaridades, este autor encuentra aspectos comunes relativos a la promoción de la investigación de clase mundial y sinergias interdisciplinarias e interinstitucionales, así como a la mejora de la visibilidad de las universidades en el contexto mundial mediante la activación de los procesos de cooperación internacionales.

En la *6th International Conference on World-Class Universities* (Salmi, 2015b), celebrada en Shanghái y organizada por la propia Universidad Jiao Tong promotora del ARWU, ha ocupado un lugar especial el análisis de las iniciativas de excelencia. El propio Salmi (2015b) hace un exhaustivo y actualizado análisis de ellas para los diferentes países en los que se pusieron en marcha. Además se analizaron experiencias específicas de países como como Rusia (Yudkevich, 2015), Singapur (Ramakrishna, 2015), Francia (Buisson, 2015) o Alemania (Huber, 2015), entre otros, y siempre con vistas al progreso en los rankings. Sin embargo, en ningún caso los análisis de las iniciativas de excelencia establecen una comparación descendiendo a la puntuación o posición concreta de cada una de las universidades que forman el sistema universitario de un país como es nuestro objetivo para el caso de España.

Con respecto a programas en los que el catalizador ha sido específicamente la mejora de las posiciones en el ranking de Shanghái podemos señalar

Tabla I. Distribución geográfica de las iniciativas de excelencia.

Región	Período 1989-2004	2005-2014
África	--	Nigeria
Asia-Pacífico	Australia, China, Hong Kong, Japón, Nueva Zelanda, Corea del Sur	China, India, Japón, Malasia, Singapur, Corea del Sur, Taiwán, Tailandia
Europa	Dinamarca, Finlandia, Irlanda, Noruega	Dinamarca, Francia, Alemania, Luxemburgo, Noruega, Polonia, Rusia, España, Suecia
Oriente Medio	--	Israel, Arabia Saudí
Norte América	Canadá	Canadá

Fuente: Salmi, 2015a.

el caso de Francia, en el que han participado en el programa de excelencia el 40% de sus instituciones de educación superior y Alemania, en el que lo han participado, bien en proyectos concretos o de mejora institucional, el 60% de las universidades (Grau, 2015). La mejora en el sistema francés se ha buscado mediante un plan de agregación de universidades en nuevas instituciones (COMUES). De un total de más de 70 instituciones de educación superior se pretende en Francia pasar a un número de COMUES por debajo de 20. En Docampo y otros (2015) se ha realizado un trabajo de prospectiva para indicar cuál sería la posición de las COMUES si fuesen analizadas dentro de ARWU. El trabajo constata que Francia ganaría visibilidad entre las 100 primeras universidades, manteniendo intacto el total de instituciones listadas entre las 500 universidades clasificadas en ARWU. Para dar una idea del interés público por esta iniciativa en Francia podemos mencionar el hecho de que este trabajo de prospectiva fue recogido y comentado en una de las publicaciones diarias de mayor impacto (Laroussiere, 2015).

La iniciativa CEI de España cuenta por tanto con numerosos antecedentes, que de seguro han contribuido a la definición de sus objetivos y prioridades. El contexto económico de la universidad no es neutro (Luque-Martínez, 2015) ni tampoco lo es el de estas iniciativas que parten de situaciones muy diferentes en lo económico y en lo político, con financiación diferente además de desigual orientación y vinculación de los diferentes colectivos implicados. Hay que constatar que existe algún antecedente del interés por evaluar los efectos de los CEI como se ha puesto de relieve por Docampo y otros (2012), al analizar la posición que ocuparían los CEI en ARWU bajo el supuesto de fusión de las universidades que los integran o por De la Torre

y otros (2014) que estudian las iniciativas CEI en la comunidad de Madrid. Pretendemos con el presente trabajo buscar evidencias que nos permitan comprobar si la iniciativa ha podido tener algún efecto medible en la mejora de las universidades españolas en los rankings internacionales.

3. ACADEMIC RANKING OF WORD UNIVERSITIES (ARWU)

El ranking de Shanghái surge en 2003 y ha alcanzado desde entonces notoriedad internacional. La metodología ha sido expuesta por sus autores (Liu y Cheng, 2005), aunque durante años han persistido dificultades insalvables en su interpretación con el fin de replicar los resultados de ARWU. Ese objetivo se ha conseguido finalmente con la publicación de Docampo (2013).

ARWU clasifica instituciones de acuerdo con la suma ponderada de las puntuaciones en seis indicadores, recogidos en la Tabla II, que se obtienen a partir de datos en bruto sobre premios internacionales, listas de investigadores relevantes, producción científica total y publicaciones, en dos revistas concretas de ámbito muy general en la ciencia, *Science* y *Nature*.

Para replicar con precisión los datos de ARWU son necesarios los siguientes pasos (Docampo y Cram, 2014):

1. Hacer acopio de los nuevos datos en bruto

Analizar las fuentes de datos relevantes con el fin de obtener información actualizada sobre nuevos premios Nobel (cada año) y medallas Fields (cada cuatro años), los nuevos miembros de las listas de HCR (highly cited researchers) de Thompson ISI

Tabla II. Criterios, indicadores y ponderación en el ARWU

Criterio	Indicador	Código	Valor
Calidad de la Docencia	Antiguos alumnos de una institución con premios Nobel y medallas Fields	Alumni	10%
Calidad del Profesorado	Profesores de una institución que han obtenido premios Nobel y medallas Fields	Award	20%
	Investigadores con alto índice de citación en diversas materias	HiCi	20%
Producción Investigadora	Artículos publicados en <i>Nature</i> y <i>Science</i>	N&S*	20%
	Artículos indexados en Science Citation Index - Expanded y Social Science Citation Index	PUB	20%
Rendimiento per Cápita	Rendimiento académico per cápita de una institución	PCP	10%

* Para instituciones especializadas en Humanidades y Ciencias Sociales, no se considera el criterio N&S y el valor se redistribuye entre los demás indicadores de forma proporcional.

Fuente: <http://www.shanghairanking.com/es/ARWU-Methodology-2015.html>.

y la filiación de los autores en *Science* o *Nature* en los pasados cinco años y en las publicaciones ISI del año anterior. En los países para los que ARWU utiliza el dato de FTE (*full time equivalent staff*), buscar una fuente oficial que proporcione datos de número y categoría de profesores para cada universidad. Todas las fuentes necesarias para la obtención de los datos anteriores son accesibles y se relacionan en el apéndice bibliográfico dentro de la sección de recursos electrónicos.

2. Combinar datos en bruto y aplicar escalado

Sumar y/o promediar las puntuaciones resultantes de los datos recogidos en la etapa anterior en las ventanas de tiempo de acuerdo con los algoritmos publicados en la web de ARWU. Para cada indicador (salvo PCP) multiplicar el valor de la puntuación en bruto por un factor de escala fija de modo que la Universidad de Harvard reciba una puntuación de 10.000. Calcular un paso intermedio necesario para el cómputo del indicador PCP mediante la división de la suma ponderada de las puntuaciones brutas escaladas por

el FTE del California Institute of Technology (CALTECH), institución con mejor puntuación PCP, y aplicar un factor de escala de modo que dicha puntuación intermedia para Caltech sea de 10.000.

3. Comprimir los datos en bruto a escala

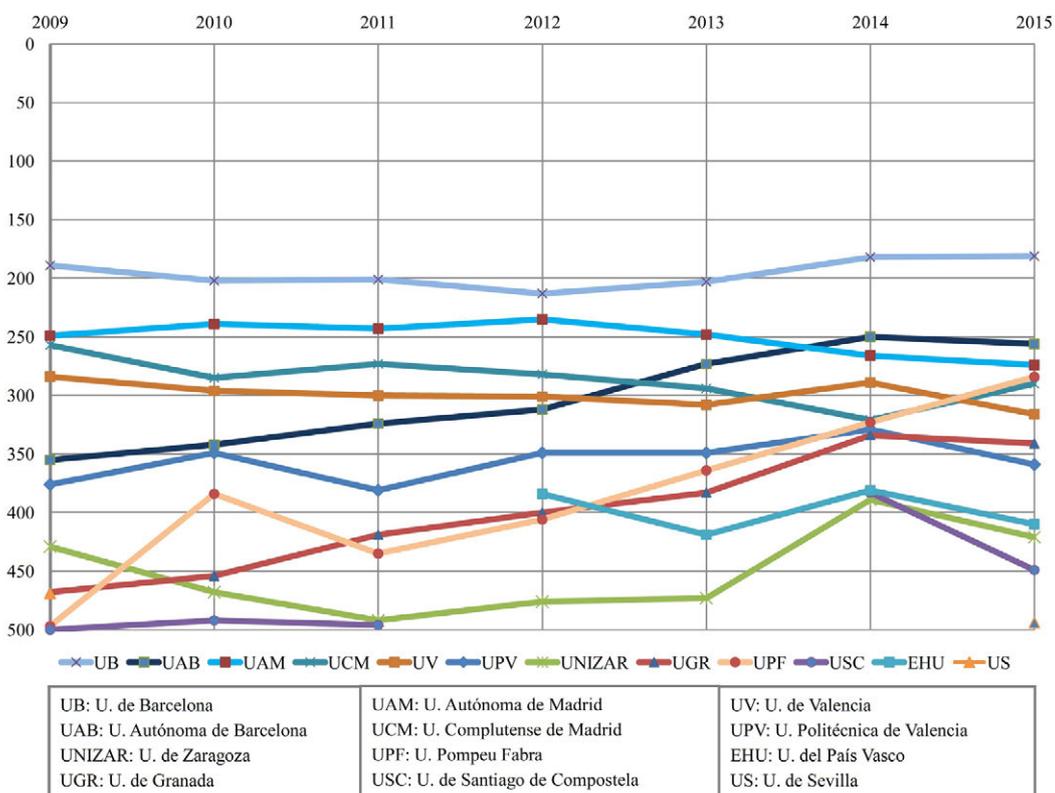
Para cada indicador, comprimir el rango dinámico de las puntuaciones en bruto mediante el cálculo de su raíz cuadrada para obtener la puntuación final.

4. Calcular la puntuación total

Finalmente, combinar las puntuaciones finales de los indicadores utilizando los pesos indicados en la web de ARWU, y reescalar linealmente la puntuación total de modo que la de la primera clasificada sea de 100.

En la Figura 1 representamos la posición de las universidades españolas que han aparecido al principio y/o al final de nuestro periodo de análisis (2009-2015) entre las 500 universidades recogidas en la lista oficial de ARWU, aunque algunas universidades no salen todos los años entre las 500 primeras, como la USC, EHU y US.

Figura 1. Evolución de las universidades españolas entre las 500 primeras en ARWU 2009-2015



4. ASPECTOS METODOLÓGICOS

Para cumplir los objetivos del presente trabajo, se ha procedido de la siguiente manera.

1. Obtención de Indicadores:

De acuerdo con la metodología desarrollada por Docampo (2013), se han reproducido los resultados de ARWU para un total de 3500 universidades ente 2008 y 2015, lo que permite obtener no solamente las puntuaciones de las universidades españolas en ARWU, sino también sus posiciones en el mundo en el período que estamos analizando. Los márgenes de error de dos de los indicadores (PUB y N&S, en los demás no se comete ningún error en las aproximaciones), debido sobre todo a la identificación precisa de las instituciones en los motores de búsqueda de la Web of Knowledge, se mantienen por debajo del 3%, por lo que inducen un error en la puntuación final del orden del 1%, lo que permite realizar adecuadamente las comparaciones. Baste señalar a ese efecto que el error de redondeo de ARWU (que solamente muestra una cifra decimal en sus resultados oficiales, puede llegar a ser cercano a ese 1% en las instituciones con baja puntuación).

2. Agregación de indicadores

Ha habido en total tres convocatorias de Campus de Excelencia Internacional (2009, 2010 y 2011). Ese hecho, junto a la conveniencia de alejar la comparación interanual de la posible influencia de oscilaciones puntuales en los valores de los indicadores, hace recomendable tomar como referencia de comparación la media de tres años, lo que permite suavizar las variaciones alejadas de la tendencia. Es decir, a efectos de medir la variación en las posiciones en ARWU se contemplan dos períodos: el momento inicial es el resultado de la puntuación media alcanzada por las universidades en los años 2009, 2010 y 2011, siendo el momento final de la comparación la media también de tres años, en este caso 2013, 2014 y 2015.

3. Identificación de los proyectos de Campus de Excelencia Internacional:

Se ha confeccionado una base de datos que recoge para cada universidad su carácter público o privado, si participa en algún proyecto o no, si ese proyecto tenía una o más universidades y si era la universidad coordinadora o no. Se han incluido además en la base de datos las puntuaciones en ARWU y

las posiciones globales ocupadas, obtenidas en las etapas 1 y 2.

4. Tratamiento estadístico de la información:

Analizando los datos de los dos períodos anteriormente mencionados, se pretende observar si hay diferencias en las puntuaciones medias del conjunto de universidades españolas, de las públicas o privadas y de cada una de ellas, así como en las posiciones netas (resultado de la diferencia neta entre subidas y bajadas de posición en la clasificación mundial durante ese período). Esto permite asimismo comprobar si hay diferencias, y en su caso si son positivas o negativas, en las puntuaciones medias de las universidades coordinadoras o que están en un CEI con una o más universidades.

5. Clasificación de las universidades participantes en proyectos CEI:

A partir del análisis estadístico, se pretende clasificar a las universidades de acuerdo con la diferencia que generan sus características y las del CEI al que pertenecen en la puntuación media en el ranking ARWU.

5. ANÁLISIS

Como base para la comparación se ha tomado la puntuación media de los tres años del comienzo del período correspondientes a las convocatorias CEI (2009, 2010 y 2011). La puntuación media final es la referida a los tres últimos años (2013, 2014 y 2015). Los datos relativos a cada universidad en las puntuaciones promedio, clasificación global y pertenencia a proyectos CEI se incluyen en las tablas del anexo 1 y 2.

Considerando la puntuación total de cada universidad estimada según ARWU, en el comienzo del período analizado, la UB alcanzaba la máxima puntuación con 17,20 puntos, siendo la media de todas las españolas de 4,84 puntos. Al final del período analizado se ha producido un aumento considerable en las puntuaciones de las universidades españolas. La UB sigue en cabeza, ahora con 18 puntos, y la puntuación media asciende a 5,77 puntos. Un aumento importante cercano al 20%. Conviene recordar que la dinámica de conversión de los datos brutos en puntuaciones finales se produce mediante un proceso de compresión (véase Docampo y Cram, 2014) que de hecho acerca a todas las universidades entre sí y con respecto a la que encabeza la lista. Una consecuencia que se deriva de este hecho es que si se quiere doblar la puntuación en ARWU han de multiplicarse por 4 los indicadores en bruto. Cuando hablamos, por tan-

to, de una mejora del 19% estamos implícitamente hablando de una mejora real superior al 40% en las cifras de producción científica en cantidad (artículos indexados en las revistas JCR) y en calidad (artículos en *Science* y *Nature*).

Otro aspecto destacable en este período es que se ha reducido la variabilidad. La dispersión en la puntuación global media del ranking de las universidades españolas era mayor en 2009 y, se ha reducido en 2015, como lo muestra el descenso del coeficiente de variación (0,81 y 0,70, respectivamente).

En cuanto a los criterios que utiliza ARWU, reflejados en la Tabla II, las universidades españolas no tienen puntuación en AWARDS y solamente una la tiene en ALUMNI, por lo que no tiene sentido analizar su variabilidad. El criterio de autores altamente citados (HiCi) es el que presenta mayor variabilidad puesto que afecta a un número reducido de universidades. Los criterios de publicaciones (PUB) y de rendimiento (PCP) son los más estables y de menor variabilidad, como se puede ver en la tabla III, en la que se recogen las puntuaciones (años 2009 y 2015) de las universidades españolas.

Por otro lado, la evolución de la puntuación media de las universidades españolas no es homogénea, dependiendo de la naturaleza pública o privada de la universidad. Así, las universidades públicas españolas alcanzan mayor puntuación que las privadas en todos los años, con diferencias importantes. La puntuación media de las universidades públicas también aumenta más en términos absolutos que la de las universidades privadas, puesto que se pasa de una media de 6,99 en el período 2009-10-11 a una media de 7,86 en el período 2013-14-15. Mientras que las universidades privadas pasan de una media de 0,98 a 1,73, mayor porcentualmente dado que parten de unos resultados muy bajos en 2009.

Tabla III. Estadísticas descriptivas de los criterios del ranking ARWU (año 2015) y puntuación total (años 2009 y 2015) de las universidades españolas

	N	Media	Desv. típ.	Coef. variación
HiCi 2015	74	1,03	2,70	2,624
N&S 2015	74	2,38	3,45	1,447
PUB 2015	74	16,2	11,67	0,719
PCP 2015	74	7,65	5,63	0,736
Puntuación total 2015	74	5,77	4,06	0,703
Puntuación total 2009	74	4,84	3,94	0,814

Si se analiza la posición de las universidades españolas en el mundo, en 2009 había 33 instituciones españolas entre las 1.000 primeras, mientras que en el año 2015 esa cifra ya es de 39, como resultado de que se incorporan 7 nuevas y sale una. Circunscribiendo el análisis a las 32 que permanecieron entre las 100 primeras a lo largo del intervalo temporal 2009-2015, las universidades españolas han ganado 444 posiciones netas (se ganan 1.118 por 16 instituciones y se pierden 674 en las otras 16) en el contexto de las 1.000 primeras universidades del ranking.

Considerando las universidades públicas y las posiciones de la media de los tres años iniciales y finales, se han incorporado 5 instituciones a las 1.000 primeras del mundo. Por otro lado, hay 31 universidades públicas que mejoran sus posiciones entre un momento y otro del período (ascendiendo un total de 2.783 posiciones) frente a 19 universidades que bajan posiciones (un total de 681) resultando un saldo positivo de 2.102 posiciones mejoradas por las universidades españolas.

a. Análisis según participación en proyectos CEI

Dado el número reducido de universidades privadas que participan en la iniciativa CEI, vinculadas solamente a dos proyectos, nos ha parecido razonable limitar el análisis a las universidades públicas. Como ya se ha comentado ha habido una mejora medible tanto en las posiciones como en las puntuaciones. En ambos casos con una correlación muy elevada, de 0,965 para las puntuaciones en el ranking y una correlación ordinal de 0,964 para el caso de las posiciones.

1. Universidades coordinadoras vs no coordinadoras de proyectos CEI

Los proyectos CEI surgen en instituciones que presentan iniciativas para incentivar y mejorar los resultados de investigación. Por otra parte, la evaluación de dichos proyectos mediante una comisión internacional primó el impacto científico de la institución coordinadora, un indicador muy relacionado con el posicionamiento en rankings (Grau, 2015). No es por tanto extraño que las universidades coordinadoras de un proyecto de Campus de Excelencia Internacional (ver tabla IV) presentan siempre puntuaciones globales medias muy superiores a las universidades que no son coordinadoras. Esto entra dentro de la lógica puesto que asumir la coordinación del proyecto supone una capacidad de liderazgo probada y, por tanto, capacidad reconocida de coordinación de las diferentes universidades agregadas.

Debido a que el punto de partida de las coordinadoras es sensiblemente más alto, los márgenes de mejora de las universidades coordinadas son mayores. Efectivamente, se comprueba que aunque para ambos tipos de universidad, coordinadora o no, se produce un claro incremento de puntuación media entre los años 2009-11 y 2013-15, el incremento en el caso de las coordinadoras está en torno al 9% mientras supera el 21% en las no coordinadoras.

Como resultado de la cooperación bajo el liderazgo de la universidad coordinadora se constata la reducción de las diferencias de puntuación entre las universidades. Así, al comienzo del período las universidades coordinadoras superaban en un 84% la puntuación media de las no coordinadoras mientras que al final del período esa diferencia se reduce al 66,5%. Podríamos concluir que el establecimiento del programa ha resultado en una mayor homogeneidad en el sistema universitario español.

Tabla IV. Comparación de la puntuación global media de las universidades españolas en ARWU, según sean o no coordinadoras de proyectos CEI

	Universidad Coordinadora	N	Media	Desv. típica
Puntuación total 2013-2015	Si	29	9,33	3,56
	No	19	5,61	1,16
Puntuación total 2009-2011	Si	29	8,54	3,21
	No	19	4,54	1,19

2. CEI con una universidad vs CEI con más de una

En los proyectos CEI con una sola institución, las universidades presentan puntuaciones medias muy superiores a la media de las puntuaciones de las universidades que forman parte de proyectos CEI con varias universidades. Ambas categorías incrementan su puntuación media entre el inicio y el final, como muestra la Tabla V. Así, las universidades que forman parte de un CEI con una sola universidad han incrementado alrededor de un 13% su puntuación media, mientras que las que integran un CEI con varias universidades aumentan algo menos (un 11%). Esto hace que la diferencia entre unas y otras se acreciente, pasando de una diferencia del 43% al principio a otra del 46% al final del periodo, siempre con puntuación mayor para las que forman en solitario un proyecto CEI. De todas formas, la diferencia en este caso no es sustantiva, dada la similitud de las cifras.

Tabla V. Comparación de la puntuación global media de las universidades españolas en ARWU, según proyectos CEI con una o varias universidades

	Nº univ. en CEI	N	Media	Desv. típica
Puntuación total 2013-2015	1	16	10,01	3,50
	Más de 1	30	6,86	2,63
Puntuación total 2009-2011	1	16	8,85	3,31
	Más de 1	30	6,18	2,85

3. Análisis de las universidades que forman parte de proyectos CEI

Tomando en consideración las posiciones de las universidades públicas españolas en ARWU, se puede contrastar si existe correlación entre la posición alcanzada y determinadas características referidas a los proyectos CEI (Tabla VI). Para ello hacemos uso de los test tau de Kendall y rho de Spearman. Según los valores de ambos estadísticos, el hecho de ser o no universidad coordinadora de un proyecto CEI está asociado a la posición ocupada en ARWU: las coordinadoras ocupan posiciones mejores en el ranking al principio, aunque la intensidad de la correlación se reduce entre el principio y el final tanto para el coeficiente de Kendall como el de Spearman, como efecto del proceso de homogeneización al que anteriormente hemos hecho referencia.

Por otro lado, también es de interés comprobar si el hecho de que el proyecto CEI esté formado por una o más universidades está correlacionado de manera significativa con la posición que ocupen dichas universidades en ARWU. El resultado obtenido para los dos estadísticos empleados conduce a una misma interpretación: efectivamente se observa una correlación positiva entre la mejor posición en el ranking y el hecho de figurar como única universidad en un CEI, derivada del hecho de que las universidades que han elegido proyectos mono-institucionales partían de una situación de ventaja al principio del periodo analizado y la han mantenido. No obstante, la participación en proyectos CEI parece haber ido acompañada de una mejora en las puntuaciones de las demás universidades, por lo que al final del periodo analizado las diferencias se han reducido.

Tabla VI. Correlación entre la posición de las universidades españolas en ARWU y ser o no coordinadora de proyecto CEI o pertenecer a un CEI con una o más universidades

	Posición Coordinación		Posición Nº universidades en CEI	
	2009-2011	2013-2015	2009-10-11	2013-14-15
Tau de Kendall	0,566	0,509	0,375	0,398
Rho de Spearman	0,685	0,616	0,454	0,483

4. Clasificación de las universidades según tipo de CEI

Con el fin de clasificar las universidades en función de las características que se están analizando, es decir, carácter público o privado, ser o no coordinadora de un proyectos CEI y tomar parte en un proyecto CEI con una o más universidades, se ha aplicado el método CHAID (Luque-Martínez y Muñoz-Leiva, 2012). Para ello se han establecido los siguientes criterios, el nivel máximo de profundidad del árbol establecido es de 4, mínimo de casos para la partición de un nodo 10, mínimo de casos en un nodo 5. El resultado obtenido se muestra en la figura 2 con dos niveles. Está claro que el criterio con mayor capacidad discriminante en cuanto a la variable dependiente, que es la puntuación en ARWU 2015, es el que distingue por número de universidades en el proyecto CEI. Dentro del grupo de los proyectos CEI conformados por varias universidades, le sigue el criterio de ser coordinadora o no. El hecho de ser pública o privada no es un criterio que marque diferencias en este contexto. Al final, se tienen cuatro grupos de universidades:

Grupo del nodo 2, formado por 17 universidades (23% del total) que participan en proyectos CEI de una sola universidad, que alcanzan una puntuación media próxima a 10. El coeficiente de variación nos indica que se trata del segundo grupo menos heterogéneo en cuanto a las puntuaciones obtenidas.

El nodo 3 no es un nodo final, reúne a los CEI con más de una universidad y alcanzan una puntuación media de 6,5, situada entre las puntuaciones extremas del nodo 1, la mayor, y nodo 2, la menor. Este nodo incluye 33 universidades, el 44,6% del total, y se divide en los nodos 4 y 5.

Grupo del nodo 5, el menos numeroso, con 14 universidades (cerca del 20%) y una puntuación media un 22% inferior al anterior. Forman parte de proyectos CEI con varias universidades, en los que actúan como institución coordinadora, y constituyen un grupo más heterogéneo que el anterior en cuanto a las puntuaciones medias.

Grupo del nodo 4, formado por 19 universidades (un poco más del 25% del total) con una puntuación sensiblemente inferior al primero (casi un 90% menor). Constituye, sin duda, el grupo más homogéneo de los analizados. Reúne a las instituciones que forman parte de proyectos CEI con varias universidades en los que no son coordinadoras.

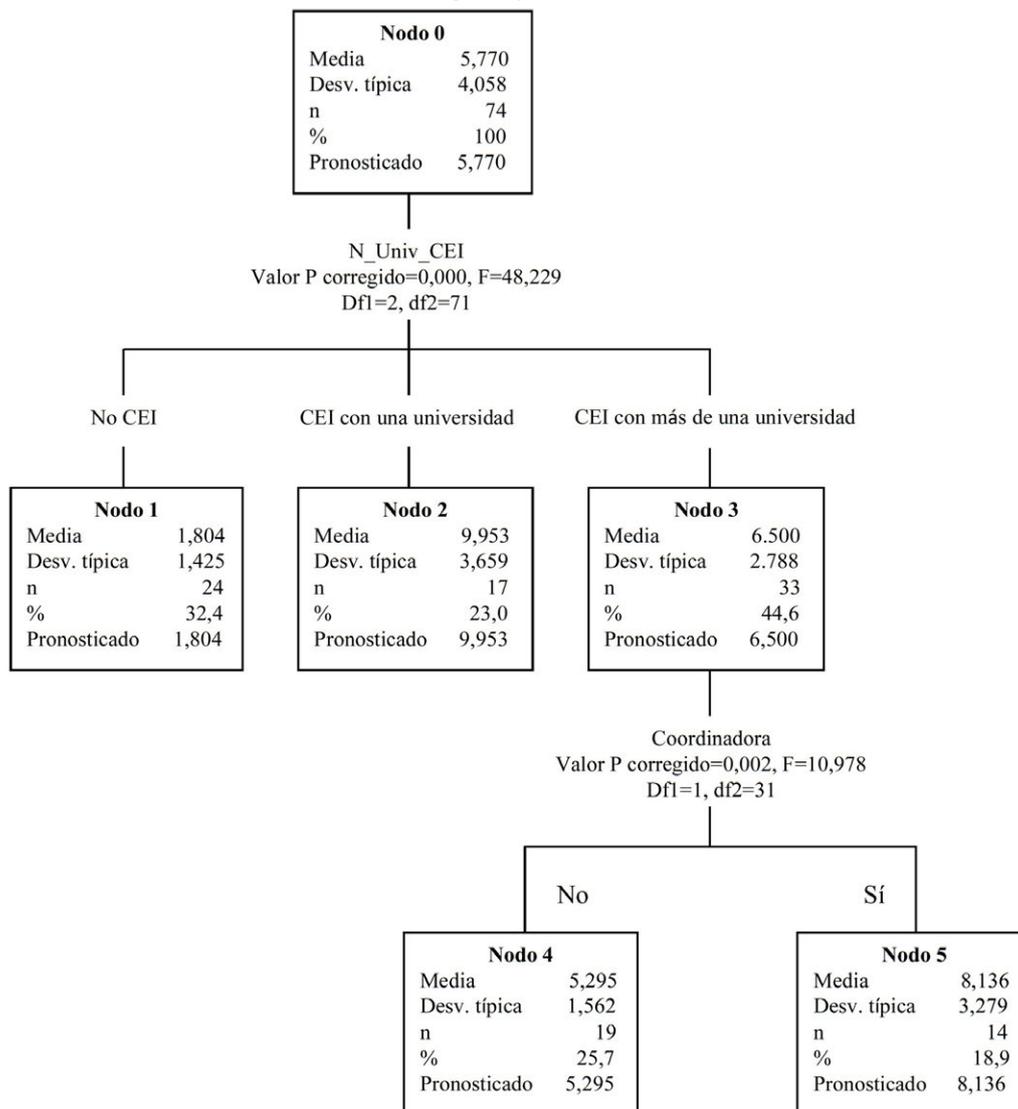
Finalmente, el grupo del nodo 1 es el de menor puntuación, llegando solamente un 18% de la puntuación media del grupo primero. Es el más numeroso, con 24 universidades (32,4% del total), y el más heterogéneo en su puntuación media al tener el mayor coeficiente de variación. Está integrado por universidades que no forman parte de ningún proyecto CEI, la mayor parte de ellas instituciones privadas.

6. LIMITACIONES, CONCLUSIONES Y RECOMENDACIONES

En el presente trabajo hemos analizado la mejora de la visibilidad internacional de las universidades españolas, derivada de su mejora en la prestación investigadora durante el periodo de duración de los proyectos de Campus de Excelencia Internacional. Con el fin de juzgar esa mayor o menor visibilidad, hemos elegido entre las clasificaciones globales la que nos ha parecido más ajustada al objetivo de centrarse en las mejoras en investigación, el ranking de Shanghái. Entre las limitaciones de este trabajo hay que asumir las propias del ranking ARWU tanto en la selección de los criterios como en la ponderación de los mismos, que han sido ampliamente tratadas en la literatura (Marginson y van der Wende, 2007; Hazelkorn, 2011; Sanz-Casado y otros, 2013).

No obstante, nos parece importante señalar que ARWU, pese a la proliferación de otros rankings de universidades y a las duras críticas recibidas, se ha convertido en el referente básico para identificar a las universidades de investigación a nivel global. De hecho, "la accesibilidad de sus fuentes y el hecho de que la jerarquía de universidades

Figura 2. Clasificación de las universidades según tipo de CEI



que ARWU reconoce se alinee con las percepciones acerca de la calidad de las universidades de élite a lo largo de los últimos tiempos han contribuido poderosamente a la primacía del ranking de Shanghái entre las clasificaciones académicas de ámbito global” (Docampo y Cram, 2014). Por otra parte, conviene destacar el tratamiento favorable en ARWU a la producción científica en Ciencias Sociales y Humanidades: como se ha puesto de manifiesto en Docampo (2013) ARWU asigna una puntuación doble a las publicaciones recogidas en el Social Science Citation Index, en el que además de las publicaciones indexadas en Ciencias Sociales se incluye en el caso de las universidades españolas más de un tercio de las indexadas en el ámbito de las Artes y las Humanidades.

Otra de las limitaciones con las que nos hemos encontrado es la posible volatilidad anual de los indicadores de productividad investigadora. Con el fin de evitar incrementos espurios debido a las posibles oscilaciones anuales, y en consonancia con el hecho de que ha habido tres convocatorias CEI, hemos optado por reducir el riesgo trabajando con las puntuaciones medias de tres años tanto al inicio como al final del período. De todos modos los autores son plenamente conscientes de que correlación no implica causación, por lo que debemos señalar que está fuera de la capacidad de nuestro análisis identificar estrictamente los efectos de los CEI con independencia de las condiciones de contorno en los que la iniciativa se ha desarrollado. Entre otros factores de incremento de la producción científica y por ende de las posiciones en clasificaciones internacionales, con-

viene no olvidar los incentivos a la producción investigadora y la puesta en marcha de los sistemas de acreditación para el profesorado universitario.

A lo largo del trabajo hemos documentado el hecho de que, entre 2009 y 2015, el sistema universitario español ha mejorado su prestación investigadora a juzgar por el incremento de su puntuación media en ARWU, superior al 19%. En dicho período se ha reducido la dispersión, derivando en un sistema universitario más homogéneo respecto a la puntuación media alcanzada en ARWU. Es importante remarcar que, debido al proceso de compresión de las puntuaciones en el ranking de Shanghái, esta mejora supone un incremento real superior al 40% en el conjunto de los indicadores que componen el ranking en los que las universidades españolas tienen puntuaciones positivas.

Al final del período de análisis se constata que hay más universidades españolas entre las 1.000 primeras en la clasificación mundial con los criterios de ARWU, de 33 se pasa a 39, es decir un 18% más. Casi todas las universidades que estaban al principio y al final de este período han conseguido una perceptible mejora en sus posiciones.

Atendiendo a las características de los proyectos CEI, las universidades coordinadoras tenían y tienen más puntuación media y ocupaban posiciones más elevadas en el ranking, en consonancia con su reconocida capacidad de liderazgo para la coordinación. Las universidades coordinadoras, al partir de niveles más altos en sus puntuaciones, presentan una evolución más estable que las no coordinadoras.

Las universidades que están solas en un proyecto CEI parecen haber sacado mayor provecho de la iniciativa, puesto que no solamente partían de una media mayor que la media de las universidades que están en proyectos con varias universidades sino que además al final del período se amplía tal diferencia, aunque en pequeña escala. El que un proyecto se presente por una sola universidad no es sinónimo de no colaboración con otras universidades, precisamente son universidades que presentan elevados indicadores de colaboración (por ejemplo en producción científica). Una posible explicación es que la financiación del proyecto CEI no se tiene que repartir con otras universidades, además todo el proceso es más fácil de coordinar, más eficaz en el desarrollo y más orientado a los resultados concretos de una sola universidad. Estas conclusiones son revalidadas al analizar los datos ordinales, relativos a las posiciones de las universidades.

Ser universidad coordinadora está correlacionado con ocupar mejores posiciones, aunque más al principio que al final del período, debido la positiva contribución que la coordinadora parece haber ejercido sobre las instituciones coordinadas. Las universidades que coordinan CEI mono-institucionales ocupan, como ya se ha indicado, mejores posiciones en el ranking, un fenómeno que se acentúa al final del período.

Las evidencias encontradas apuntan, *ceteris paribus*, a un cierto efecto iniciativa CEI en las puntuaciones medias y en las posiciones de las universidades públicas españolas en ARWU. Se produce al mismo tiempo una mejora y una reducción de las diferencias entre las universidades que no han coordinado un proyecto CEI y las que sí lo han hecho. Esto puede aconsejar desarrollos policéntricos en futuras convocatorias, en los que se intente, en primer lugar, promover proyectos con una universidad y en segundo lugar proyectos con pocas universidades y próximas, con el fin de mejorar la visibilidad internacional de las universidades no coordinadoras.

Aunque los proyectos CEI fueron concedidos en años diferentes, en la práctica se gestaron al mismo tiempo. La iniciativa CEI se ha desarrollado en un plazo demasiado corto y solamente en tres años se ha dispuesto de cierta financiación, que además ha sido escasa en relación a las expectativas y los objetivos perseguidos en los proyectos. A esto hay que añadir una postura titubeante con la continuidad de la iniciativa. No obstante, la prestación investigadora de las universidades españolas, analizada a través del prisma de los resultados en el ranking de Shanghái, se ha revelado sumamente positiva, incluso en un contexto de crisis económica que ha mermado en gran medida los recursos disponibles para la investigación. Es difícil pensar, de todos modos, que dicha situación se repita si se continúa trabajando con escasez de medios y difuso horizonte de futuro.

El reto de la mejora de la visibilidad internacional del sistema universitario español precisa de más tiempo y una apuesta más decidida y sostenida no solamente en términos de financiación sino con otras medidas complementarias y de apoyo que puedan reforzarla, como las acciones en los planes nacionales de investigación, de movilidad internacional y otras medidas de internacionalización, de creación de empresas y emprendimiento, de acciones en posgrado y, en particular, en doctorado.

7. REFERENCIAS

- Altbach, P. G.; Salmi, J. (Eds.). (2011). *The road to academic excellence: The making of world-class research universities*. Washington DC; World Bank Publications. <http://dx.doi.org/10.1596/978-0-8213-8805-1>
- Barnett, R. (Ed.). (2012). *The future university: Ideas and possibilities*. New York; Routledge.
- Buisson L. (2015). Trading between visibility and performance at global and local levels: The example of research universities in France. 6th International Conference on World-Class Universities 1-4 November, Shanghai (China). <http://www.shanghairanking.com/wcu/wcu6/20.pdf>
- Carayannis, E.G.; Campbell, D. F. J. (2011). Open Innovation Diplomacy and a 21st Century Fractal Research, Education and Innovation (FREIE) Ecosystem: Building on the Quadruple and Quintuple Helix Innovation Concepts and the "Mode 3" Knowledge Production System. *Journal of the Knowledge Economy*, vol. 2 (3), 257-326. <http://dx.doi.org/10.1007/s13132-011-0058-3>
- De La Torre, E.; Perez-Esparrells, C.; Casani, F. (2014). El programa campus de excelencia internacional en España y algunas estrategias de universidades de la Comunidad de Madrid para mejorar su posición en los rankings globales. *Investigaciones de Economía de la Educación*, vol. 9, 351-369. <http://EconPapers.repec.org/RePEc:aec:ieed09:09-18>
- Delgado, L. (2015). Análisis comparado de programas de excelencia. *Jornadas sobre la Iniciativa de Campus de Excelencia Internacional "Análisis de las experiencias en diferentes países y propuestas de futuro"*. Granada 9 abril 2015. <http://biotic.ugr.es/pages/jornada-cei-9-de-abril>
- Docampo, D. (2011). On using the Shanghai ranking to assess the research performance of university systems. *Scientometrics*, vol. 86 (1), 77-92. <http://dx.doi.org/10.1007/s11192-010-0280-y>
- Docampo, D.; Luque-Martínez, T.; Torres-Salinas, D.; Herrera, F. (2012). Efecto de la agregación de universidades españolas en el Ranking de Shanghai (ARWU): caso de las comunidades autónomas y los campus de excelencia. *El profesional de la información*, vol. 21 (4), 428-432. <http://dx.doi.org/10.3989/redc.2011.2.797>
- Docampo, D. (2013). Reproducibility of the Shanghai academic ranking of world universities results. *Scientometrics*, vol. 94 (2), 567-587. <http://dx.doi.org/10.1007/s11192-012-0801-y>
- Docampo, D.; Cram, L (2014). On the Dynamics of the Shanghai ranking. *Scientometrics*, vol. 98 (2), 1347-1366. <http://dx.doi.org/10.1007/s11192-013-1143-y>
- Docampo, D.; Egret, D.; Cram, L (2015). The effect of university mergers on the Shanghai ranking. *Scientometrics*, vol. 104 (1), 175-191. <http://dx.doi.org/10.1007/s11192-015-1587-5>
- Erkkilä, T. (2013). *Global University Rankings: Challenges for European Higher Education*. New York; Palgrave Macmillan. <http://dx.doi.org/10.1057/9781137296870>
- Etzkowitz, H.; Leydesdorff, L. (1998). The endless transition: A "triple helix" of university-industry-government relations. *Minerva*, vol. 36 (3), 203-208. <http://dx.doi.org/10.1023/A:1017159001649>
- Hazelkorn, E. (2011). *Rankings and the reshaping of higher education: The battle for world-class excellence*. New York; Palgrave Macmillan. <http://dx.doi.org/10.1057/9780230306394>
- Huber B. (2015). The role of universities in society: Challenges ahead. 6th International Conference on World-Class Universities 1-4 November, Shanghai (China). <http://www.shanghairanking.com/wcu/wcu6/14.pdf>
- Grau Vidal, F.J. (2015). Rankings, impacto científico y sistemas universitarios, Monografías CRUE, 2015. ISBN: 978-84-608-3021-4. http://www.crue.org/Publicaciones/Documents/MONO_RANKING_WEB.pdf
- Kretz, A.; Sá, C. (2013). Third stream, fourth mission: Perspectives on university engagement with economic relevance. *Higher Education Policy*, vol. 26 (4), 497-506. <http://dx.doi.org/10.1057/hep.2013.32>
- Larousserie, D 2015. La racine carrée du classement de Shanghai. *LE MONDE SCIENCE ET TECHNO*, http://www.lemonde.fr/sciences/article/2015/04/06/la-racine-carree-du-classement-de-shanghai_4610477_1650684.html
- Liu, N. C.; Cheng, Y. (2005). Academic ranking of world universities: Methodologies and problems. *Higher education in Europe*, vol. 30 (2), 127-136. <http://dx.doi.org/10.1080/03797720500260116>
- Luque-Martínez, T.; Muñoz-Leiva, F. (2012): *Segmentación jerárquica*. En Luque-Martínez, T. (2012). *Técnicas de análisis de datos en investigación de mercados*. Madrid, Pirámide.
- Luque-Martínez, T. (ed.) (2015). *Horizon 2031. The University of Granada in Light of its V Centenary. "Reflections on the future of the University"*. Granada; Editorial Universidad de Granada. <http://biotic.ugr.es/pages/horizon-2031>
- Luque-Martínez, T. (2015). Actividad investigadora y contexto económico. El caso de las universidades públicas españolas. *Revista Española de Documentación Científica*, vol. 38 (1), e076. <http://dx.doi.org/10.3989/redc.2015.1.1135>
- Marginson, S.; van der Wende, M. (2007). Globalization and Higher Education. *OCDE Education Working Papers* No. 8, OCDE, Paris. <http://dx.doi.org/10.1787/173831738240>
- Ministerio de Educación (2011). *Estrategia Universidad 2015. Contribución de las universidades al progreso socioeconómico español 2010-2015* pdf [consulta: 14 abril 2011] <http://www.educacion>

es/dctm/eu2015/2011-estrategia-2015-espanol.pdf?documentId=0901e72b80910099

Ramakrishna, S. (2015). Strategies to be globally visible and locally engaged. 6th International Conference on World-Class Universities 1-4 November, Shanghai (China). <http://www.shanghairanking.com/wcu/wcu6/22.pdf>

RIS 3 (2013). *Guide to Research and Innovation Strategies for Smart Specialisation (RIS 3) 2013*. http://s3platform.jrc.ec.europa.eu/en/c/document_library/get_file?uuid=e50397e3-f2b1-4086-8608-7b86e69e8553

Salmi, J. (2015a). *How effective are excellence initiatives at building world-class universities?* En Luque-Martínez, T. (ed.) (2015) *Horizon 2031. The University of Granada in Light of its C Centenary. "Reflections on the future of the University"*. Granada; Editorial Universidad de Granada. <http://biotic.ugr.es/pages/horizon-2031>

Salmi, J. (2015b). Excellence strategies and the creation of world-class universities. 6th International Conference on World-Class Universities 1-4 November,

Shanghai (China). <http://www.shanghairanking.com/wcu/wcu6/3.pdf>

Salmi, J. (2009). The challenge of establishing world-class universities. Washington, DC: World Bank Publications. <http://dx.doi.org/10.1596/978-0-8213-7865-6>

Sanz-Casado, E.; García-Zorita, C.; Serrano-López, A.E.; Efraín-García, P.; De Filippo, D. (2013). Rankings nacionales elaborados a partir de múltiples indicadores frente a los de índices sintéticos. *Revista Española de Documentación Científica*, vol. 36 (3), 1-18. <http://dx.doi.org/10.3989/redc.2013.3.1.023>

Shattock, M. (2010). *Managing successful universities*. Maidenhead; McGraw-Hill Education (UK), p.27.

Yudkevich, M. (2015). Global visibility and locally engagement: Can they go together? The case of National Research University Higher School of Economics. 6th International Conference on World-Class Universities 1-4 November, Shanghai (China). <http://www.shanghairanking.com/wcu/wcu6/21.pdf>.

ANEXO 1: Relación de Campus de Excelencia Internacional, universidades coordinadoras y participantes

Proyecto CEI	Universidad coordinadora	Universidades participantes
BKC Barcelona Knowledge Campus	Barcelona	Politécnica de Catalunya
Campus Moncloa. La energía de la diversidad	Complutense de Madrid	Politécnica de Madrid
Campus Carlos III	Carlos III de Madrid	
UAB CEI: apuesta por el conocimiento y la innovación	Autònoma de Barcelona	
CEI UAM + CSIC	Autónoma de Madrid	
Andalucía TECH	Sevilla	Málaga
Campus Energía: energía para la excelencia	Politécnica de Catalunya	
Campus Iberus: CEI del Valle del Ebro	Zaragoza	Pública Navarra, Lleida, Rioja
Campus de Montegancedo	Politécnica de Madrid	
Campus UPF - Icària Internacional	Pompeu Fabra	
HUBc: Health University Barcelona Campus	Barcelona	
EUSKAMPUS. Una Universidad, un País, un Campus	UPV/EHU	
VLC/CAMPUS. Valencia, International Campus of Excellence	Valencia	Politécnica de Valencia
Campus Vida	Santiago de Compostela	
Campus Habitat 5U	Politécnica de Valencia	Valencia, Jaume I, Alicante, Miguel Hernández
Proyecto Horizonte 2015. Donde talento y progreso se unen	Navarra	
CEI BioTic Granada	Granada	

Proyecto CEI	Universidad coordinadora	Universidades participantes
Ad Futurum	Oviedo	
Campus de Excelencia Agroalimentario (CEI-A3)	Córdoba	Almería, Cádiz, Huelva, Jaén
Cantabria Campus Internacional	Cantabria	
Campus de Excelencia Internacional Catalunya Sud	Rovira i Virgili	
Campus do Mar Knowledge in depth	Vigo	Coruña, Santiago
Campus Mare Nostrum	Murcia	Politécnica de Cartagena
Campus Studii Salamantini	Salamanca	
CEI CANARIAS: Campus Atlántico Tricontinental	Universidad de Las Palmas de Gran Canaria	La Laguna
ARISTOS Campus Mundus 2015	Ramon Llull	Deusto, Comillas
CEI Triangular-E3 Los Horizontes del Hombre	Valladolid	León, Burgos
Campus Euromediterráneo del Turismo y el Agua	les Illes Balears	Girona
CEI-MAR. CEI del Mar	Cádiz	Almería, Granada, Huelva, Málaga
CYTEMA. Campus Científico y Tecnológico de la Energía y el Medioambiente de la UCLM	Castilla-la Mancha	
HIDRANATURA. CEI en Gestión Eficiente de Recursos Hidro-naturales	Extremadura	
CEI Energía Inteligente	Alcalá	Rey Juan Carlos, Murcia, UPCT, Extremadura

ANEXO 2: Relación de Universidades Públicas españolas con su puntuación y posición en ARWU

Universidad	Puntuación		Posición ARWU		Posición concreta ARWU	
	2009-11	2013-15	2009-11	2013-15	2009-11	2013-15
Barcelona	17,1	17,5	150-200	150-200	192	189
Autónoma de Barcelona	11,6	14,7	301-400	201-300	354	258
Autónoma de Madrid	15,0	14,4	201-300	201-300	245	262
Complutense de Madrid	14,1	13,5	201-300	201-300	272	297
Valencia	13,3	13,4	201-300	301-400	295	303
Pompeu i Fabra	9,3	12,9	401-500	301-400	483	319
Politécnica de Valencia	11,2	12,4	301-400	301-400	368	345
Granada	9,7	12,3	401-500	301-400	445	348
País Vasco	10,2	10,8	401-500	401-500	424	416
Zaragoza	9,8	10,4	401-500	401-500	444	441
Santiago de Compostela	8,6	9,9	401-500	401-500	518	475
Politécnica de Catalunya	8,1	9,8	401-500	401-500	557	483
Sevilla	9,2	9,3	401-500	501-600	480	513
Oviedo	7,3	8,9	601-700	501-600	627	539
Vigo	7,8	8,8	501-600	501-600	588	554
Rovira i Virgili	6,7	8,5	601-700	501-600	680	569
Politécnica de Madrid	7,0	8,3	601-700	501-600	651	590
La Laguna	7,3	8,0	601-700	601-700	618	612
Alacant	6,5	7,3	701-800	601-700	716	677

Universidad	Puntuación		Posición ARWU		Posición concreta ARWU	
	2009-11	2013-15	2009-11	2013-15	2009-11	2013-15
Illes Balears	6,7	7,2	601-700	601-700	691	693
Cantabria	6,4	7,1	701-800	601-700	723	699
Salamanca	6,8	6,9	601-700	701-800	677	729
Castilla La Mancha	6,7	6,8	601-700	701-800	683	741
Murcia	6,3	6,7	701-800	701-800	741	744
Miguel Hernández de Elche	6,3	6,7	701-800	701-800	741	750
Málaga	5,6	6,7	801-900	701-800	840	759
Valladolid	7,1	6,6	601-700	701-800	644	764
Jaén	4,7	6,3	1001-1100	801-900	1051	808
Alcalá	5,9	6,3	701-800	801-900	796	822
Navarra	5,9	6,2	701-800	801-900	786	831
Carlos III de Madrid	5,4	6,2	801-900	801-900	863	832
Rey Juan Carlos I	5,3	6,2	801-900	801-900	885	840
Girona	5,0	6,1	901-1000	801-900	973	836
Jaume I de Castellón	4,5	6,1	1001-1100	801-900	1077	839
Córdoba	5,8	6,1	801-900	801-900	808	844
Pablo Olavide	4,5	5,9	1001-1100	801-900	1077	871
Lleida	4,3	5,6	1001-1100	901-1000	1125	916
A Coruña	4,7	5,4	1001-1100	901-1000	1036	949
Nacional de Educación a Distancia	4,2	5,3	1001-1100	901-1000	1147	968
Cádiz	4,8	5,2	1001-1100	901-1000	1009	967
Extremadura	5,7	5,2	801-900	901-1000	834	980
Huelva	3,7	5,2	1201-1300	901-1000	1258	988
Pública de Navarra	4,2	5,0	1101-1200	1001-1100	1142	1035
Almería	4,2	4,5	1101-1200	1101-1200	1150	1149
Las Palmas de Gran Canaria	4,0	4,5	1101-1200	1101-1200	1192	1159
León	3,9	4,3	1201-1300	1201-1300	1208	1223
Politécnica de Cartagena	3,9	4,1	1201-1300	1201-1300	1200	1286

ESTUDIOS / RESEARCH STUDIES

Organización y puesta en valor del fondo documental del *Instituto Español de Entomología* (archivo histórico del Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas)

María Cruz Osuna Arias*, Juana Molina Nortes**, Carolina Martín Albaladejo*

*Museo Nacional de Ciencias Naturales (CSIC)

**Unidad de Recursos de Información Científica para la Investigación (CSIC)

Correos-e: osunacruz@hotmail.com, juana.molina@bib.csic.es carolina.martin@mncn.csic.es

Recibido: 25-03-2015; 2ª versión: 03-06-2015; Aceptado: 17-01-2016.

Cómo citar este artículo/Citation: Osuna Arias, M. C.; Molina Nortes, J.; Martín Albaladejo, C. (2016). Organización y puesta en valor del fondo documental del *Instituto Español de Entomología* (archivo histórico del Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas). *Revista Española de Documentación Científica*, 39(3): e144. doi: <http://dx.doi.org/10.3989/redc.2016.3.1287a>

Resumen: El presente artículo tiene el propósito de ofrecer una síntesis del trabajo de organización realizado sobre el fondo documental del extinto Instituto Español de Entomología (IEE) del Consejo Superior de Investigaciones Científicas (CSIC) cuyo tratamiento ha constituido el proyecto piloto para la aplicación del cuadro de clasificación normalizado para los Archivos Históricos del CSIC. Toda la labor ha sido llevada a cabo por investigadores y profesionales de la documentación, desde la perspectiva del proyecto de investigación donde se enmarca y con la intención de ofrecer el resultado al servicio de la comunidad científica que lo requiera. El proyecto ha incluido la mejora de la conservación física de la documentación y la descripción en el catálogo automatizado de la Red de Bibliotecas y Archivos del CSIC. El fondo forma parte del Archivo del Museo Nacional de Ciencias Naturales (MNCN) y su consulta puede realizarse a través del catálogo de Archivos del CSIC.

Palabras clave: Archivos del CSIC; cuadro de clasificación para los archivos históricos del CSIC; descripción archivística; series documentales de archivos; Instituto Español de Entomología (CSIC); Historia de la Ciencia.

Organization and value of archival documents of the Spanish Institute of Entomology (historical archive of National Museum of Natural Sciences, Spanish National Research Council)

Abstract: This article provides an overview of the work carried out for organizing the archival collection of the defunct Spanish Institute of Entomology, attached to the Spanish National Research Council (CSIC). This work has been used as a pilot project for implementing the standard classification scheme for CSIC's historical archives. All the work was carried out in collaboration between researchers and documentation professionals, from the standpoint of the research project to which it belongs and with the intention of making the results available for the scientific community as required. The physical preservation of the documents was improved and their description was included in the automated catalog of the CSIC Libraries and Archives Network. This collection is part of the archives of the National Museum of Natural History (MNCN) and can be consulted through the CSIC archive catalog.

Keywords: CSIC archives; classification scheme for CSIC historical archives; archival description; archival documentary series; Spanish Institute of Entomology (CSIC), History of science.

Copyright: © 2016 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Attribution (CC BY) España 3.0.

1. INTRODUCCIÓN

A mediados del siglo XIX tuvo lugar en España el despegue de la Entomología, disciplina que se ocupa del estudio científico de los insectos artrópodos. Sus inicios estuvieron liderados por la figura de Mariano de la Paz Graells (1809-1898)¹, pero fue en el siglo XX cuando se consolidó una nueva etapa y puede hablarse de Entomología moderna (Bach y Compte, 1997). Muchos científicos contribuyeron en este sentido, como Laureano Pérez Arcas, Ignacio Bolívar y Urrutia, Ricardo García Mercet, Manuel Martínez de la Escalera, Cándido Bolívar Pieltáin y José María Dusmet, entre otros (Martín Albaladejo, 2005).

Son también varias las instituciones que desempeñaron un papel relevante, como la Sociedad Española de Historia Natural, la Escuela Superior de Ingenieros Agrónomos y sobre todo el Museo Nacional de Ciencias Naturales (MNCN). En el Museo, uno de los establecimientos científicos que recibió el apoyo y el impulso de la Junta de Ampliación de Estudios (JAE), los trabajos sobre Entomología tuvieron una relevancia significativa, al tiempo que varios de sus entomólogos más distinguidos ocuparon cargos de responsabilidad en las principales entidades relacionadas con las Ciencias Naturales de la época, destacando Ignacio Bolívar y Urrutia, a cuyo prestigio como científico internacionalmente reconocido se sumó su labor institucional en la dirección del MNCN, del Real Jardín Botánico y también de la JAE (Gomis, 1988).

La Sección de Entomología del MNCN existía desde 1901 y en el primer tercio del siglo XX ya había alcanzado un grado de madurez destacable. Contaba con una apreciable colección de insectos, material científico, laboratorios y una biblioteca especializada; también eran numerosas las publicaciones que generaban sus miembros, las relaciones institucionales y la importancia de la revista sobre Entomología que editaban (Revista *EOS*, nacida en 1925). Expresión de reconocimiento internacional a la labor de esta Sección fue la celebración en Madrid del *VI Congreso Internacional de Entomología*, en septiembre de 1935, siendo Ignacio Bolívar presidente del mismo.

Con la creación del CSIC, por Ley de 24 de noviembre de 1939² todos los centros de investigación antes dependientes de la JAE pasaron a formar parte del mismo, convirtiéndose así en la nueva institución encargada de "*fomentar, orientar y coordinar la investigación científica nacional*". Su estructura se configuró a través de Patronatos³ que representaban las diferentes ramas del saber, donde se incluyeron los institutos, encargados, en último extremo, de llevar a cabo la labor científica

encomendada. Uno de ellos fue el Instituto Español de Entomología (IEE) que se creó por Decreto de 10 de marzo de 1941⁴ y fue incluido, en principio, dentro del Patronato "Alonso de Herrera" dedicado a las Ciencias de la Agricultura y Ganadería, base de la riqueza de la España de la época⁵. Dos años más tarde, en 1943, se adscribió al Patronato "Ramón y Cajal", dedicado este a las Ciencias Biológicas y del que dependió hasta el año 1968 en el que vuelve al "Alonso de Herrera"⁶.

El IEE incorporó las colecciones, materiales, biblioteca y personal de la antigua Sección de Entomología del MNCN (Gomis, 2014), así como el espacio físico que ocupaba esta Sección. El decreto de creación dispuso además que "*serán funciones del Instituto la recolección, conservación y determinación de la fauna entomológica, especialmente la de España, así como de toda clase de investigaciones que redunden en beneficio de la economía e interés nacionales*"⁷. La creación del nuevo instituto se justificó tanto por la proyección internacional y avance científico que ya se había alcanzado en esta materia, como por el interés social y la oportunidad económica que representaba, en concreto en los sectores agrícola, sanitario, forestal y pecuario.

Gonzalo Ceballos y Fernández de Córdoba (1989-1967) fue el primer director. Su vinculación al Museo se remontaba a 1918, cuando terminó sus estudios como ingeniero de Montes y comenzó a colaborar con la Sección de Entomología del MNCN; fue a partir de 1929 cuando Ignacio Bolívar le promovió en relación a los estudios de Entomología aplicada (Otero y López, 2012), materias que se deseaba reforzar en relación con los ya bien avanzados trabajos en Taxonomía y Sistemática que se hacían en la Sección. En 1934 Ceballos⁸ obtuvo una plaza como catedrático de la Escuela Superior de Ingenieros de Montes de Madrid y, terminada ya la Guerra Civil, fue el último responsable de la Sección de Entomología, sustituyendo a Cándido Bolívar que había partido al exilio en México.

El instituto se dividió inicialmente en tres secciones: Entomología General, Entomología Agrícola y Entomología Forestal. Las funciones que tenía encomendadas abarcaban, por un lado, la investigación pura en relación con la determinación y clasificación de especies, incluyendo el incremento y conservación de una colección de insectos compuesta por millones de ejemplares; para ello se realizaban salidas al campo para muestrear y se llevaban a cabo labores destinadas a la preservación y cuidado del material. También se adquirían, cuando el presupuesto y la oportunidad lo permitían, colecciones enteras y se incorporaba a los fondos el nuevo material recolectado por colaboradores que el instituto tenía en distintos lugares.

Por otro lado, se estimulaba la investigación que aportara beneficio directo a otros sectores (sanidad, agricultura, medio ambiente, etc.); en este campo se pretendía ofrecer un servicio de referencia y consultoría a nivel nacional⁹. Los primeros trabajos de Entomología aplicada estuvieron orientados a la conservación forestal¹⁰. Dado que desde una perspectiva económica existía gran interés en las repoblaciones forestales y en la protección de las cuencas hidrográficas, pronto se va a contar con el apoyo del Patronato "Juan de la Cierva".

El problema de la falta de personal científico y técnico suponía un reto, pues a consecuencia de la Guerra y del exilio se había perdido la mayor parte de la plantilla con la que contaba la antigua sección. Entre las diferentes estrategias que se sucedieron a lo largo del tiempo en cuanto a la incorporación de efectivos humanos, resalta la figura del colaborador científico, de larga tradición dentro del CSIC, personal cuya vinculación con la Institución permitía una flexibilidad adecuada a la falta de presupuesto.

El instituto mantuvo durante toda su trayectoria interés por la divulgación científica; en este sentido se organizaron cursos y conferencias y se dispuso de unas salas de exposición permanente que mostraban una selección significativa de insectos. Lo recaudado como entrada a esta exposición suponía una pequeña contribución¹¹ que revertía en el presupuesto del instituto y que ayudaba a la continua reclamación de medios, personal y dotación general que desde la dirección del IEE se realizaba en diferentes momentos¹². Con una disposición también divulgativa se publicaba desde 1943 la revista *GRAELLSIA* (Martín Albaladejo y otros, 2014), dirigida tanto a los aficionados a la Entomología como a otros profesionales con intereses aplicados: médicos, veterinarios, ingenieros, agricultores, profesores, etc.

Se pueden distinguir tres etapas en relación a los directores que se sucedieron en el tiempo. La primera y más extensa fue la de Gonzalo Ceballos, que tuvo como secretarios a Eduardo Zarco (1941-1958) y Eugenio Morales Agacino (1958-1967). Le sucedió Ramón Agenjo (Selga y Templado, 1984), quien fue director desde 1967 a 1978 y secretaria de Dolores Selga Serra (1967-1984). Un cambio a destacar fue el producido en 1967 con la inclusión en el IEE del Departamento de Fauna del Suelo y Entomología Aplicada, del Instituto de Edafología y Fisiología Vegetal, también del CSIC, centro que tuvo una relación constante con el IEE por la coincidencia en temas de investigación similares entre sus profesionales. En esta época las secciones se ampliaron: Sección de Sistemática y Distribución, Sección de Entomología Experimental y Aplicada,

Sección de Zoología del Suelo, Sección de Nematología y Laboratorio de Patología de Insectos. La dirección de la última etapa, de tan solo siete años de duración (1978-1984), estuvo a cargo de Salvador V. Peris (González Mora, 2008), quien organizó el IEE en tres departamentos: Departamento de Entomología Sistemática y Distribución, Departamento de Entomología Ecológica y Experimental y Departamento de Zoología del Suelo. Tanto las diferentes etapas como los diferentes estilos personales de la dirección del IEE tienen su reflejo en las series documentales conservadas.

En los años ochenta el CSIC llevó a cabo diversas reestructuraciones en sus centros, transformaciones a las que no fue ajeno el IEE. En 1984, por acuerdo de Junta de Gobierno de 2 de diciembre, este instituto pasó a ser uno de los cinco departamentos que componían la nueva estructura del Museo¹³; unos años más tarde la Colección de Entomología pasó al Departamento de Colecciones y el Departamento de Entomología quedó incluido en el de Biodiversidad y Biología Evolutiva.

2. PROYECTO PILOTO PARA LA NORMALIZACIÓN DE LA CLASIFICACIÓN DE LOS FONDOS ARCHIVÍSTICOS DEL CSIC

Los trabajos de organización y descripción del fondo del IEE coinciden en el tiempo con el "Plan de Actuación de Archivos del CSIC" redactado en 2012 por la Unidad de Recursos de Información Científica para la Investigación (URICI) en el ámbito de lo establecido por la normativa del CSIC. Esta Unidad (antigua Red de Bibliotecas) fue creada por resolución de la Presidencia del CSIC del 2 de junio de 2011 y asumió, entre otras funciones, la recuperación del patrimonio documental histórico. En el año 2013 se puso en marcha uno de los programas del plan dedicado a la identificación de fondos, en el que se abordó la necesidad de establecer un modelo único de clasificación documental para la estructura y funciones de todos los centros del CSIC. El objetivo era ofrecer a los bibliotecarios y archiveros un "cuadro de clasificación normalizado" para ser aplicado en la descripción multinivel del catálogo de archivos del CSIC. Aunque este catálogo de archivos comenzó a funcionar en 1998 aplicando la *Norma Internacional de Descripción Archivística*, ISAD(G) a los registros descritos bajo el formato MARC, la búsqueda por cuadros de clasificación era una nueva funcionalidad que se inauguró en junio del 2012 con motivo del "Día Internacional de los Archivos". Con este nuevo acceso a la descripción multinivel se ofrece al usuario la información del contexto de los fondos de archivo, se describen los fondos, divisiones de fondo en series documentales y por último las unidades compuestas o simples.

El "cuadro de clasificación normalizado" va dirigido a aquellos centros e institutos del CSIC con fondos históricos desde 1940 a 1980. En el Anexo 1 se ofrece una selección de la bibliografía utilizada para la realización de este cuadro de clasificación.

La primera versión del cuadro se presentó desde la Coordinación de Archivos de la URICI y fue revisada y modificada en diferentes reuniones durante ese año por el grupo de trabajo de la Red de Bibliotecas y Archivos del CSIC¹⁴. Se decidió comenzar entonces por el fondo del IEE como "proyecto piloto" para la aplicación del cuadro de clasificación normalizado (Molina Nortes, 2014). Dos fueron los motivos fundamentales por los que se escogió este fondo:

- Por una parte, se trataba de un fondo con un volumen razonable de documentación (99 unidades de instalación) en el que estaban representadas las actividades más típicas de una institución científica de este periodo, lo que le hacía ser un buen modelo para la aplicación del sistema de clasificación propuesto. Además, el tema del fondo, relacionado con el área de ciencias naturales y agrícolas, era el idóneo por ser similar a otros conservados en la Red de Bibliotecas y Archivos del CSIC, como los del Instituto de Edafología y Fisiología Vegetal (Madrid), el de la Misión Biológica a Galicia (Pontevedra) o el del Instituto Pirenaico de Ecología (Jaca).
- Por otra, como integrantes del equipo del proyecto de investigación se contaba con especialistas en la organización de archivos del CSIC, miembros que conocían las características de estos fondos y la aplicación de la ISAD(G) aplicada al Formato MARC 21 dentro del catálogo de archivos del CSIC. Este aspecto facilitaba enormemente la comunicación en todas las tareas de gestión documental (Molina Nortes, 2013).

El trabajo con el fondo del IEE permitió la identificación de las funciones y actividades reflejadas en la documentación de acuerdo al modelo propuesto en el cuadro de clasificación, adscribiendo las diferentes agrupaciones documentales descritas en el catálogo a su correspondiente código de clasificación, código que reflejaba la categoría funcional a la que pertenecía. Por otra parte, en el proceso de descripción de títulos, tanto de series como de agrupaciones inferiores, se tuvo en cuenta como elemento prioritario la denominación de funciones más que de contenidos.

El fondo del IEE queda entonces estructurado en las cinco grandes agrupaciones del cuadro:

1. Dirección y Gobierno
2. Administración General y Secretaría
3. Investigación
4. Actividad académica y cultural
5. Servicios de apoyo a la investigación

En estas agrupaciones de primer nivel están recogidas las actividades más representativas de cualquier centro de investigación, grupos que a su vez se subdividen en ocasiones hasta en tres y cuatro niveles más, quedando las series documentales adscritas a ellos. Las series de las funciones más generales están reflejadas en los puntos 1 y 2 y las series de funciones específicas desde el punto 3 hasta el 5. La propuesta va acompañada de un sencillo sistema de codificación compuesto por 18 códigos alfanuméricos significativos con los que clasificar fácilmente las series en el nivel jerárquicamente superior que corresponda (Ver Anexo 2).

3. ORGANIZACIÓN Y DESCRIPCIÓN DEL FONDO DEL IEE

3.1. Estado de conservación, instalación y descripción general

El 20 de diciembre de 1984 la Junta de Gobierno del CSIC aprobó la supresión del IEE y la integración del personal, de la colección entomológica y de la biblioteca, en el MNCN. La desaparición del instituto generó una nueva realidad administrativa. La ausencia de política archivística del CSIC, que por esa época aún no contaba con un plan integral de archivos, llevó al olvido los documentos que ya no eran de utilidad inmediata y práctica para la gestión, de forma que estos permanecieron abandonados en las dependencias del antiguo instituto, ahora museo. A partir de los años noventa la investigadora Dra. Isabel Izquierdo Moya¹⁵, va recuperando y remitiendo en sucesivas transferencias al Archivo del MNCN la documentación del desaparecido IEE. Su contenido puede ser conocido de forma somera a través de las relaciones de entrega, aunque debido a la falta de personal no se culminaron todas las fases de organización y descripción pertinentes para la conservación final de la documentación.

El fondo documental tratado en este proyecto constaba de 99 cajas de archivo definitivo y sus fechas extremas comprenden los años 1923 hasta 1985. En el mes de mayo del año 2012 se realizó una revisión del fondo para tener una primera aproximación a su estructura y conocer sus características más relevantes. Se hicieron unos primeros exámenes y se analizó la información de los instru-

mentos de descripción disponibles. A partir de estos datos, se elaboró un informe en el que se identificaban a grandes rasgos las series documentales más representativas, así como otras agrupaciones más generales de las funciones del instituto; también se indicaba el tipo de ordenación, volumen, firmas y fechas extremas de estas agrupaciones y se redactó una primera propuesta de actuación archivística, que finalmente no se llevó a cabo por falta de recursos humanos destinados al proyecto.

Es necesario aclarar que el fondo documental tratado no recoge la totalidad de la documentación producida por el instituto. Una parte del fondo se disgregó llegando al archivo en momentos diferentes y, en algunos casos, se encontró catalogada al comienzo de este proyecto. Se trata de documentación de carácter iconográfico, sobre todo dibujos y fotografías, que están incluidas en fondos especiales y cuyo contenido puede ser consultado en el Catálogo de Archivos del CSIC. Por otra parte, los papeles más vinculados a los investigadores han constituido, tradicionalmente, el fondo de "Personal Científico" dentro del archivo histórico del museo, por lo que allí también se encuentra documentación de algunos de los científicos que formaron parte del IEE.

Tanto para avanzar en el conocimiento del fondo a describir como para el estudio científico del propio proyecto de investigación, se realizó una copia en formato digital de los Libros de Registro de entrada y salida, libros que recogen los apuntes de los asuntos tratados en el centro desde 1941 hasta 1985. Este documento ofrece una información general reunida en una única unidad documental, de manera que permite abordar, en una primera aproximación, la investigación sobre las actividades diarias de la institución. La información de los Libros de Registro se vertió en una base de datos en Excel, disponible también para su consulta.

Se elaboró un somero inventario en Excel como instrumento de descripción en el que se consignaron los topográficos de las unidades de instalación. Por otra parte, las etiquetas de algunas cajas también aportaron información adicional sobre tipos documentales.

Respecto al contenido del fondo, tras la primera revisión se constató la falta de orden en la mayor parte de la documentación. Dentro de las mismas cajas contenedoras se encontraron mezcladas series de actividades diversas que abarcaban rangos cronológicos muy amplios; en apariencia parecía tratarse de un fondo con un volumen alto de documentos de apoyo informativo. Todo ello puede estar relacionado con la ausencia de un protocolo de archivo en el momento de la realización de

las transferencias. Se encontraron también casos excepcionales en los que la documentación guardaba un cierto orden, cronológico o alfabético, como en la correspondencia mantenida por la dirección o la secretaría del centro, que alcanza un volumen aproximado de 27 cajas. Por una parte, se conserva una correspondencia administrativa, con información muy variada, que trata de todos los asuntos de gobierno y política científica de la institución, formada por oficios y minutas con sellos y registros de validación. Por otra parte, existe además otra correspondencia con un marcado carácter personal, relacionado con la labor propia de la investigación y conservación de la colección entomológica, cuya información es muy variada y de elevado interés para la historia de las actividades puramente científicas, intercambio de publicaciones, movimiento de ejemplares etc.

El bloque documental identificado como administración y gestión económica es el que apareció más desordenado, con gran número de fotocopias formando parte de los expedientes. Se comprobó que su contenido era diverso: gestión de las relaciones externas con otros organismos, informes de proyectos de investigación, expedientes económicos de su financiación, libros de contabilidad, gestión del patrimonio mueble e inmueble, gestión de los recursos humanos, etc. En ocasiones aparecían mezclados documentos de carácter científico, como los cuadernos del laboratorio de Entomología, diarios de actividades, etc.

En cuanto a su instalación, la documentación se conservaba en cajas de plástico, con agrupaciones de documentos en carpetas de diversos formatos y, en ocasiones incluso sin ellas (Figura 1); en general, se constató la presencia de un gran número de elementos metálicos, grapas, clips, etc. que fueron eliminados.

3.2. Organización de las series y acceso del fondo a través del catálogo de archivos y bibliotecas del CSIC

La serie documental permite abordar la aparente complicada estructura administrativa de un centro, al traducir y hacer accesibles las funciones, competencias y actividades de la institución reflejadas en los documentos conservados. En nuestro caso, la tarea de identificación de las series se ha visto determinada por las características del fondo ya comentadas, su carácter irregular e incompleto. El volumen de documentación varía de unas series a otras, circunstancia que no ha impedido reflejar en el cuadro de clasificación una estructura general más cercana a la verosimilitud administrativa que cabría encontrar en el archivo histórico de la institución.

Figura 1. Estado en que se encontró la documentación del fondo del Instituto Español de Entomología.



Se han descrito un total de 31 series documentales. En todo momento se ha intentado ofrecer una descripción coherente y adecuada a su nivel, tratando de adaptar los recursos disponibles al volumen de trabajo y al siempre determinante factor tiempo. Los niveles inferiores de descripción elegidos se han adaptado a las peculiaridades y necesidades de cada serie, siendo mayoritario el nivel de unidad documental compuesta (expediente) y en ocasiones el de unidad de instalación (caja). Se han respetado los criterios de ordenación (cronológica, onomástica, alfabética o por materias) establecidos originariamente. En aquellos casos en los que se ha encontrado una evidente alteración del orden original, debido al paso del tiempo o a otros problemas de conservación, se ha tratado de reconstruir.

Cada una de las series ha sido descrita en un registro que ofrece los datos básicos: ordenación, data, contenido, firmas, etc. Todos los registros de serie contienen un campo que le relaciona con el fondo y cada uno de los registros de niveles inferiores está relacionado, de la misma forma, con su correspondiente serie. Este campo dentro de los registros descriptivos del catálogo de archivos es el que permite establecer la descripción multinivel que marca la ISAD (G).

Al analizar el contenido de algunas de las series propuestas llama la atención su desigual pervivencia a lo largo del tiempo de vida del instituto. Quizá en algunos casos esto se deba a los problemas de conservación y transferencias detectados, pero también hay que destacar la impronta personal que han dejado los diferentes responsables de la

institución, que adaptaron los recursos de personal disponibles y los diferentes estilos de trabajo a las circunstancias de cada momento.

Si bien el primer objetivo del tratamiento del fondo ha sido proporcionar una visión lógica del conjunto, la labor descriptiva no puede darse aún por finalizada. En este sentido, el proyecto es susceptible de ampliarse si así se valora en un futuro. Las series descritas conforme al cuadro de clasificación normalizado para centros del CSIC pueden ser consultadas en el catálogo de archivos del CSIC (Aleph) desde el registro descriptivo del fondo¹⁶ (Anexo 3).

Entre las actividades que quedan reflejadas en la documentación descrita se distinguen unas funciones generales y otras específicas. Las funciones generales son las que van a quedar plasmadas en la documentación producida tanto por los órganos de gobierno, que se materializan en la correspondencia de los directores y secretarios, y en las actas de sesiones de sus órganos colegiados (Juntas de centro). También se han considerado funciones generales las producidas por los servicios de administración, como por ejemplo es la gestión de los recursos económicos, reflejada en la serie "Expedientes de gastos", la serie de "Nóminas", gracias a la cual se puede analizar la evolución de los recursos humanos o la serie "Expedientes informativos". En las funciones generales también se incluyen las relaciones institucionales, que en general, tratan de la relación entre el IEE y los órganos centrales de dirección del CSIC, de los que depende para el desarrollo de las funciones que tiene encomendadas (serie "Correspondencia administrativa" que

se compone de las comunicaciones recibidas y enviadas por el instituto que pasan por registro y llevan su correspondiente validación oficial), y los recursos patrimoniales (de memorias, convenios de colaboración, proyectos de obras, expedientes de adquisición de material, inventarios) que se recogen en las series "Memorias", "Convenios de colaboración", "Proyectos de obras", "Expedientes de adquisición de material", "Inventarios", etc. Estas series aportan datos de interés para el análisis de actividades no sólo de tipo administrativo, sino también científico.

Entre las funciones de tipo específico se encuentra la actividad relacionada con la gestión de la colección entomológica y las relaciones profesionales de colaboración e intercambio científico representan fundamentalmente las funciones específicas de la institución y pueden analizarse a través de la riquísima serie de "Correspondencia científica" que mantiene orden y estructura durante casi treinta años seguidos (1941-1967), y a través de la cual pueden ser estudiadas las

relaciones profesionales de colaboración e intercambio científico. Los avatares en la investigación de este área disciplinar también pueden ser analizados consultando las series "Estudios e informes técnicos" y la de "Proyectos de investigación". También son interesantes los "Diarios de actividades" (Fig. 2) del laboratorio de Entomología anteriores a la creación del IEE, los "Expedientes de congresos", "Informes técnicos" o los "Proyectos de investigación". La gestión editorial de las revistas del instituto EOS y GRAELLSIA se puede estudiar a través de los "Expedientes de edición". Las actividades culturales y de difusión del instituto se reflejan fundamentalmente en los "Expedientes de gestión de la exposición permanente", en la correspondencia del registro de entrada y salida y en la serie exposiciones y divulgación cultural. Este fondo incluye por el momento sólo la documentación textual del instituto. Resta por incorporar la abundante documentación gráfica que se conserva, compuesta en su mayoría por dibujos científicos.

Figura 2. Diarios de actividades, cuadernos de laboratorio, libros de registro del fondo del Instituto Español de Entomología.



3.3. Principales series documentales

Se describen a continuación dos de las series más relevantes del fondo. La serie de "*Correspondencia científica*" se corresponde con el periodo que ocupa la dirección del instituto de Gonzalo Ceballos, desde su creación en 1941 hasta 1967, año de su fallecimiento. Son sus productores, por tanto, el director y los dos secretarios. Destaca la participación muy activa de estos últimos, Eduardo Zarco y Morales Agacino, quienes en relación directa con el director atendieron todos los asuntos de la dirección científica: consultas sobre el contenido de la colección entomológica del instituto, identificación de ejemplares, intercambios con otras instituciones en relación a viajes de muestreo, congresos y publicaciones. En su tipología documental la serie se compone de las comunicaciones recibidas y enviadas: abundan las cartas, tarjetas postales, telegramas, y las minutas enviadas desde el instituto. En ocasiones se conservan los documentos adjuntos a la correspondencia: informes, memorias, etc.

Para la descripción de esta serie se ha tenido en cuenta el orden en que fue conservada. Sorprende su disposición y custodia en un único orden alfabético de correspondencias, continuado durante casi 26 años, aunque con el paso del tiempo el orden se volvió irregular, difícil de reconstruir e inadecuado para un archivo definitivo. Sucede que unas veces el orden alfabético se refiere a las personas y otras a las instituciones, acusándose muchas incidencias tanto de apellidos como de nomenclatura. Por falta de tiempo y dado su elevado volumen (16 unidades de instalación), se optó por agrupar individualmente e incorporar al catálogo únicamente aquellas correspondencias cuyo volumen fuera representativo, utilizando para ello los descriptores de personas e instituciones. Este recurso descriptivo no altera el orden original al tiempo que hace más accesible su consulta y permite en un futuro ampliar el catálogo que en la actualidad cuenta con más de doscientos registros de esta serie.

En cuanto a su contenido, como se ha señalado, recoge las principales actividades del instituto, su relación con otras instituciones e investigadores y la gestión de la colección entomológica, sobre todo intercambios, préstamos y donaciones. La temática tratada va desde campañas de muestreo al intercambio de ejemplares de insectos pasando por todos los diversos temas de colaboración científica en general y colaboración institucional. También se tratan asuntos relativos a las publicaciones de trabajos y al intercambio de información científica. Son abundantes los intercambios postales con otros países, entre otros

Alemania, Austria, Gran Bretaña, Bélgica, Rusia, Chipre, Canadá, Finlandia y Portugal. Es, por tanto, recurrente el uso de otros idiomas, aunque, dado que el instituto contaba con un servicio de traducción, la mayoría de los documentos, sobre todo en lengua alemana e inglesa, conservan su versión en español.

No se considera necesario ofrecer la larga lista de científicos que integran esta correspondencia, pero sí destacar algunos nombres. Por una parte los entomólogos colaboradores del instituto que repartidos por diversos lugares establecen una relación fluida y fructífera, entre ellos: Anselmo Pardo (Melilla), Joaquín Matéu (Barcelona), Francisco Español (Barcelona) y Luis Báguena. Este último merece una mención aparte ya que se cuenta con más de doscientas cartas suyas que ofrecen una amplia visión de su extensa trayectoria y de la estrecha colaboración que tuvo con el instituto, tanto destinado en Guinea, Honduras o Valencia o trabajando en la sede de Madrid. Como entomólogo español sin la categoría de colaborador se puede mencionar a Ignacio Docavo.

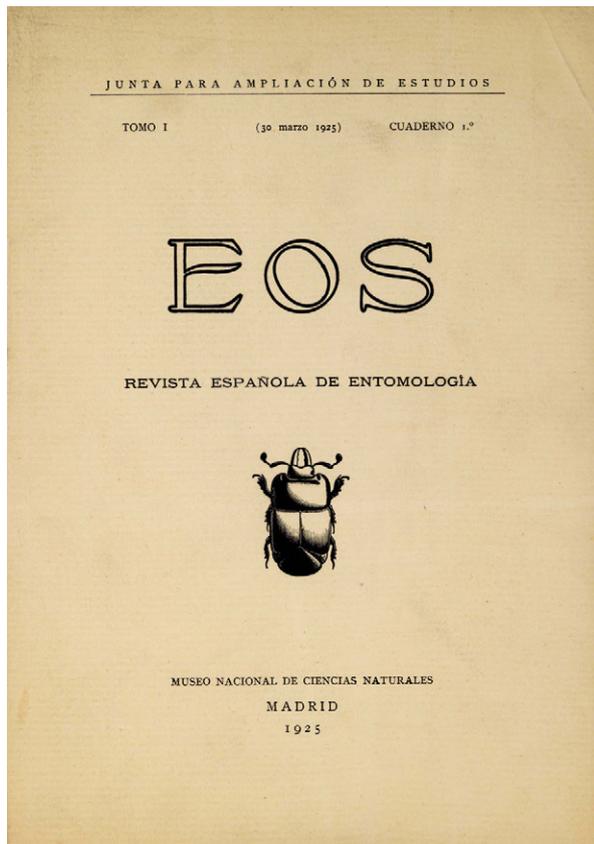
Entre los entomólogos extranjeros con abundante correspondencia con el IEE destaca el ruso Boris Uvarov, los austriacos Max Beier, Wilhelm Goetsch y Stephan von Breuning, el finlandés Keith McE. Kevan y el portugués Antonio Barros Machado.

Ejemplo de relación institucional destaca la correspondencia con Gregorio Marañón en su calidad de Director del Instituto de Cultura Hispánica.

A partir de la dirección de Ramón Agenjo, y después de Salvador V. Peris, el carácter de la correspondencia cambia, constituyendo la serie denominada "*Correspondencia de Dirección*" y que incluye también la correspondiente a la secretaría de Dolores Selga, mantenida esta aparte. Su carácter es testimonio de las funciones de dirección y administración del instituto y su contenido incluye la gestión de actividades como relaciones institucionales, investigaciones, tesis, informes, convenios, cooperación, asuntos de personal y publicaciones del instituto. La dirección y la secretaría de estos dos periodos mantienen una correspondencia separada, pero se considera que tanto por su temática como por su función institucional, constituyen la misma serie.

La segunda serie que se desea destacar es la de "*Edición de las revistas EOS y GRAELLSIA*", actividad relevante en casi todos los institutos del CSIC. Las dos revistas que el IEE gestionó fueron EOS y GRAELLSIA (Figura 3). La primera comenzó a editarse en tiempos de la JAE, en el año 1925. La segunda, GRAELLSIA, fue fundada en 1943 por Gonzalo Ceballos (Martín Albaladejo y otros, 2014).

Figura 3. Portadas de las revistas EOS y GRAELLSIA



Dada la falta de recursos y de personal especializado, la gestión y edición de las revistas EOS y GRAELLSIA fue asumida como una función más del director y el secretario. La serie se compone de las comunicaciones recibidas y enviadas por el instituto en relación a la actividad editorial. Por el momento sólo está recogida esta actividad en relación a la dirección de Ceballos.

Los tipos documentales más abundantes son las cartas y las minutas y entre los asuntos tratados se encuentran antecedentes sobre el envío, intercambio y solicitud de volúmenes o separatas, envío de artículos para corrección de pruebas de imprenta, gestión con distribuidores y otros editores, etc. Una serie complementaria a esta sería la "Correspondencia científica", dado que en ocasiones puede haber antecedentes relativos a la gestión editorial.

Tanto la serie de "Correspondencia científica" como la de "Edición de las revistas EOS y GRAELLSIA", se ha ordenado y organizado propiciando la elaboración de registros donde se ha recogido el asunto en el campo título y los oportunos índices onomásticos y de institución que facilitan la consulta.

4. CONCLUSIONES

Sin duda lo más novedoso de este proyecto archivístico es que ha servido de "proyecto piloto" para aplicar el nuevo *Cuadro de Clasificación de Fondos de Centros del CSIC 1940-1980*, una iniciativa que ha partido de la Unidad de Recursos de Información Científica para la Investigación dentro del programa de identificación del Plan de Actuación de Archivos del CSIC (2012). Se ha trabajado en la organización y descripción de un volumen de 99 unidades de instalación que suponen más de 35.000 documentos. Se ha realizado además, un trabajo de acondicionamiento físico de los documentos y resignaturización, utilizando para ello el material normalizado del Archivo Histórico del MNCN.

En este fondo han sido identificadas 31 series cuyos expedientes y agrupaciones han quedado descritos, hasta el momento, en 610 registros, accesibles a través del catálogo de archivos de la Red de Bibliotecas y Archivos del CSIC.

Con este proyecto se pone en valor la importancia que tienen las series documentales en la des-

cripción archivística, aspecto que no está implantado de forma sistematizada en todos los fondos de archivo del CSIC y que servirá para comenzar a desbrozar ese "bosque documental" con el que siempre nos encontramos a la hora de organizar la memoria histórica escrita de este complejo organismo de investigación que es el CSIC.

La labor realizada no podría entenderse sin la colaboración entre archiveros y científicos, una cooperación que ha sido tan útil y que sobre todo ha conducido a establecer objetivos concretos, factibles y prácticos, valorar la consecución de resultados a corto plazo y ha permitido alejarse de la dispersión que suele tentar cuando es amplio el volumen documental a tratar. El diálogo y la cercanía continua ha enriquecido el proyecto, los expertos en Entomología han ayudado a entender las funciones específicas de la institución, con las dificultades añadidas de conservar, acrecentar y disponer de manera sistemática un volumen ingente de esos pequeños seres que son los insectos. Por otra parte los investigadores se han beneficiado de inmediato del trabajo realizado. Además de hacer saber cuáles eran sus prioridades e intereses desde el mismo momento que comenzó la catalogación documental, ésta podía ser consultada y por tanto utilizada como materia prima para la historia de la ciencia, tarea que se ha materializado en la publicación de varios artículos de investigación con información obtenida del fondo archivístico del IEE y otros más que están en preparación.

El proyecto se ha desarrollado acorde al principio de procedencia y orden original, utilizando para la descripción la norma ISAD(G). Las fases más importantes se han articulado partiendo del estudio de la institución productora, el análisis de la documentación existente y sus necesidades descriptivas y el estudio de las principales series conservadas. Los instrumentos de descripción resultantes, el catálogo accesible a través de Internet y el cuadro de clasificación desarrollado, dan la posibilidad de conocer de forma detallada la organización, competencias y funcionamiento de esta institución.

En definitiva, el proyecto descriptivo ha servido para contextualizar el acervo documental del IEE de acuerdo a las actividades generales propias de cualquier centro de investigación de este periodo de la historia del CSIC. La identificación y descripción de series ofrece al historiador de la ciencia una herramienta adecuada para decidir líneas de investigación sobre un aspecto determinado susceptible de ser estudiado.

Es de desear que la descripción del fondo se concluya en su totalidad, incorporando la documentación iconográfica del instituto y descendiendo a niveles de descripción inferiores en aquellas series que por su importancia lo requieran. De esta forma se com-

pletará el acceso a los documentos, ampliándose el conocimiento que tenemos actualmente del IEE. Así, los objetivos marcados han sido llevados a término pues la documentación es hoy más accesible y se encuentra disponible para ser utilizada por todos los usuarios interesados, a la vez que se ha construido una estructura de ordenación y descripción flexible, que ayudará a acometer trabajos futuros.

5. AGRADECIMIENTOS

Agradecemos al personal del Archivo del MNCN su dedicación a los fondos, especialmente a Manuel Parejo Paulino, por la inestimable ayuda que siempre nos prestó. A los alumnos de la Universidad Autónoma de Madrid, Ana Díaz Sal y Diego Márquez Ganuza y de la Universidad de Alcalá de Henares, Beatriz Balcells Gutiérrez, su colaboración y buena disposición en las tareas encomendadas. A Raquel Ibáñez González del CCHS, por su buena disposición para las consultas en las que nos ha atendido. Damos igualmente las gracias al personal de Biblioteca del MNCN por su apoyo y buen hacer en todas las consultas que les hemos realizado. A María Sánchez Luque la lectura crítica del manuscrito.

Este trabajo se enmarca dentro del proyecto de investigación titulado *El Instituto Español de Entomología, CSIC (1941-1985): una etapa en la historia de una ciencia* (ref. HAR2011-28621) que incluye entre sus objetivos la puesta en valor del fondo documental de esta institución. La realización de este estudio se efectuó entre febrero y agosto de 2013.

ACKNOWLEDGEMENTS

We are grateful to the staff of the Archives of MNCN for their commitment to the collection, specially to Manuel Parejo Paulino for the invaluable help he continuously provided us. To the students Ana Díaz Sal and Diego Márquez Ganuza (Universidad Autónoma de Madrid) and Beatriz Balcells Gutiérrez (University of Alcalá de Henares) for their cooperation and willingness to carry out the assigned tasks. To Raquel Gonzalez Ibanez CCHS, for her willingness for consultations in which she has treated us. To the MNCN Library staff for their support and excellent work in all queries we have made them. We would also like to thank María Sánchez Luque, for her critical reading of the article.

This work is part of the research project entitled *The Spanish Institute of Entomology, CSIC (1941-1985): a stage in the history of science* (Ref. HAR2011-28621) which includes among its objectives the enhancement of documentary background of this institution. The realization of this study was conducted between February and August 2013.

6. NOTAS

- Véase por ejemplo Cervantes (2009).
- Ley de 24 de Noviembre de 1939 creando el Consejo Superior de Investigaciones Científicas. *Boletín Oficial del Estado*, 28 de noviembre de 1939, núm. 322, pp. 6668-6671.
- Ministerio de Educación Nacional. Decreto de 10 de febrero de 1940 regulando el funcionamiento del Consejo Superior de Investigaciones Científicas. *Boletín Oficial del Estado*, 17 de febrero de 1940, núm. 48, pp. 1201-1203.
- Ministerio de Educación Nacional. Decreto de 10 de marzo de 1941 por el que se crea en el Consejo Superior de Investigaciones Científicas, el Instituto Español de Entomología. *Boletín Oficial del Estado*, 22 de marzo de 1941, núm. 81, pp. 1932-1933.
- CSIC, Memoria de 1940-1941, pág. 46.
- Esta alternancia de adscripción en los dos Patronatos, el "Alonso de Herrera" y el "Santiago Ramón y Cajal" es evidente no solo en el caso del IEE, sino también en otros institutos del CSIC dedicados a las ciencias biológicas y de la vida. Así lo reflejan las memorias históricas anuales generales del CSIC consultables en línea: <http://www.csic.es/web/guest/memorias>
- Ministerio de Educación Nacional. Decreto de 10 de marzo de 1941 por el que se crea en el Consejo Superior de Investigaciones Científicas, el Instituto Español de Entomología. *Boletín Oficial del Estado*, 22 de marzo de 1941, núm. 81, pp. 1932-1933.
- Más información sobre Gonzalo Ceballos en Agenjo (1967).
- De todo esto dan buena cuenta las Memorias anuales del instituto que pueden consultarse, por un lado las impresas del CSIC y por otro las anuales o parciales que se conservan en el fondo del IEE.
- Véase el *Informe de directrices de 1955* en ACN0457 y Martín Albaladejo y otros (2016).
- Véase el *Informe general de 1958* en ACN0457. También está muy bien conservada, y puede consultarse, toda la documentación contable relacionada con la exposición, venta de entradas, material expositivo, gratificaciones al personal que custodia las salas, etc.
- Véase el *Informe directrices de 1963* en ACN0458.
- Archivo Central del CSIC. Actas de la Junta de Gobierno del C.S.I.C. Acta de 2 de diciembre de 1984.
- El grupo de trabajo está formado por bibliotecarios y archiveros de los siguientes centros: Real Jardín Botánico (Madrid), Museo Nacional de Ciencias Naturales (Madrid), Centro de Ciencias Humanas y Sociales (Madrid), Instituto de Ciencias Agrarias (Madrid), Instituto de Física Teórica y Matemáticas (Madrid), Centro de Física Miguel Antonio Catalán (Madrid), Ciencias de la Construcción Eduardo Torroja (Madrid), Escuela de Estudios Árabes (Granada), Escuela de Estudios Hispanoamericanos (Sevilla), Misión Biológica de Galicia (Pontevedra), Instituto Pirenaico de Ecología (Jaca), Institución Milá y Fontanalls (Barcelona) e Instituto Botánico (Barcelona).
- Isabel Izquierdo, conservadora de la Colección de Entomología desde 1987 a 2012, asume la responsabilidad de remitir al Archivo legajos de documentos que pertenecieron al antiguo instituto.
- El fondo IEE ocupa el puesto nº 5 en el cuadro de organización del Archivo del Museo Nacional de Ciencias Naturales. Desde el registro de fondo nº 100674 se enlaza con todas las series descritas.

7. REFERENCIAS

- Agenjo, R. (1967). El Excmo e Ilmo Sr. Prof. Dr. D. Gonzalo Ceballos y Fernández de Córdoba, 1895-1967. *Eos*, 43: 319-343.
- Bach, C.; Compte, A. (1997). La Entomología moderna en España. Su desarrollo: de los orígenes a 1960. *Boletín de la Sociedad Entomológica Aragonesa*, 20: 367-392.
- Cervantes, E. (coord.) (2009). *El naturalista en su siglo: homenaje a Mariano de la Paz Graells en el CC aniversario de su nacimiento*, Logroño, Instituto de Estudios Riojanos - Museo Nacional de Ciencias Naturales, CSIC, p. 279.
- Gomis, A. (1988). Presentación, IX-XV. En: *Ignacio Bolívar y las ciencias naturales en España*. Estudios sobre la Ciencia nº 4. Madrid, Consejo Superior de Investigaciones Científicas, p. 224.
- Gomis, A. (2014). Mimbres para otro cesto: De la Sección de Entomología del Museo Nacional de Ciencias Naturales al Instituto Español de Entomología. *Boletín de la Real Sociedad Española de Historia Natural, Sección Biología*, 108: 37-47.
- González Mora, D. (2008). In memoriam Excmo. Sr. Prof. Dr. Salvador Vicente Peris Torres (1922-2007). *Boletín de la Asociación Española de Entomología*, 32(1-2): 13-17.
- Martín Albaladejo, C. (2005). *Bibliografía entomológica de autores españoles (1758-2000)*. CD-ROM, Museo Nacional de Ciencias Naturales, CSIC.
- Martín Albaladejo, C.; Camacho, A.I.; Ramos, M.A.; Nieves-Aldrey, J.L.; Fernández J. (2014). GRAELLSIA 70 años después: la era digital. *Graellsia*, 70(2): e014. <http://dx.doi.org/10.3989/graellsia.2014.v70.123>.

- Martín Albaladejo, C.; Carrascosa, A.V.; Notario, A. (2016). El Instituto Español de Entomología (CSIC) y la multitud molesta. *Asclepio* 68(1): p125. <http://dx.doi.org/10.3989/asclepio.2016.04>.
- Molina Nortes, J. (2013). Metodología y técnicas para la gestión de archivos científicos. Disponible en: <http://digital.csic.es/handle/10261/75302>.
- Molina Nortes, J. (2014). Cuadro de Clasificación para centros del CSIC (1940-1980): hacia una normalización en la organización de la documentación de instituciones científicas. *Enredadera*, 23. Disponible en: <http://bibliotecas.csic.es/web/enredadera/numero-23> [consultado en junio de 2016].
- Otero Carvajal, L.E.; López Sánchez, J.M. (2012). *La lucha por la modernidad. Las Ciencias Naturales y la Junta para Ampliación de Estudios*, Madrid, CSIC - Publicaciones de la Residencia de Estudiantes, p. 1310.
- Selga, D.; Templado, J. (1984). In Memoriam. Prof. D. Ramón Agenjo Cecilia (1908-1984). *Eos*, 60: 7-21.

ANEXO 1. BIBLIOGRAFÍA

Esta selección bibliográfica, utilizada para la realización del cuadro de clasificación de los archivos históricos (1940-1980) del CSIC, ha sido compendiada por Juana Molina Nortes, coordinadora de Archivos del CSIC (2011-2013).

Archivo de la Universidad de Alicante. Cuadro de Clasificación de la Documentación Administrativa. Disponible en: <http://sar.ua.es/es/archivo/documentos/gest-doc/cuadro-de-clasificacion/cuadro-de-clasificacion.pdf> [consultado en 20 diciembre 2012]

Archivo de la Universidad Pública de Navarra. Cuadro de Clasificación. versión 13: Disponible en: http://www.unavarra.es/digitalAssets/174/174987_Cuadro-enero-2013.pdf [consultado en 15 de enero de 2013]

Archivo Universitario de Granada. Cuadro de Clasificación de fondos de la UGR. Disponible en: <http://archivo.ugr.es/pages/fondos/cuadros/pdf/version13> [consultado el 20 de septiembre de 2012]

Archivo Universitario de la Universidad de Las Palmas. Cuadro de Clasificación. Disponible en: http://biblioteca.ulpgc.es/conocenos_archivo#3 [consultado el 15 de septiembre de 2012]

Archivo Universitario de la Universidad de Sevilla. Cuadro de Clasificación. Disponible en: <http://servicio.us.es/archivos/HTM/Cuadro.htm> [consultado el 21 septiembre de 2012]

Arovelius, R. y otros (2010). *Management and Preservation of Scientific Records and Data*. International Council on Archives. Section on University and Research Institution Archives Committee on Scientific and Research Data, p. 41. Disponible en: http://www.ltu.se/cms_fs/1.95314!/file/Handbook%202010%20final%20SUV%20logo.pdf [consultado el 3 agosto 2012]

Ayuntamiento de Barcelona. Quadre de Classificació Uniforme de Documents. Disponible en: <http://w110.bcn.cat/ArxiuMunicipal/Continguts/Docu->

<ments/Fitxers/Quadre%20de%20Classificacio.pdf> [consultado el 28 de octubre de 2012]

Barbadillo Alonso, J. (2007). Apuntes de clasificación archivística. *Legajos: cuadernos de investigación archivística y gestión documental*, 10: 27-50.

Barbadillo Alonso, J. (2010). Clasificaciones y relaciones funcionales de los documentos de archivo. *Tabula, Revista de Archivos de Castilla y León*, 13: 95-104.

Boscarini, F. (2010). La clasificación de documentos basada en funciones: comparación de la teoría y la práctica. *Tabula Revista de Archivos de Castilla y León*, 13: 41-58.

Chica Blanco, G. de la; Icaza Ruiz, P. M^a; Osuna Arias, M. C.; Poves Pérez, E. (2011). La memoria de los archivos científicos en el Centro de Ciencias Humanas y Sociales del CSIC. Comunicación presentada en las Quintas Jornadas de Archivo y Memoria. *Extraordinarios y fuera de serie: formación, conservación y gestión de archivos personales*. 17-18. Madrid. Disponible en: http://digital.csic.es/bitstream/10261/48635/1/Memoria_de_los_archivos.pdf [consultado el 29 de abril de 2012]

CNEDA. Comisión de Normas Españolas de Descripción Archivística (2011). *Modelo Conceptual de Descripción Archivística y Requisitos de Datos Básicos de las Descripciones de Documentos de Archivo, Agentes y Funciones. Parte 1: Tipos de Entidad. Parte 2: Relaciones*. Borrador final de la CNEDA (09-06-2011), 266 p. Disponible en: http://www.mecd.gob.es/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/documentacion/NEDA_MCDA_P1_P2_20120618-pdf/NEDA_MCDA_P1_P2_20120618.pdf [consultado en junio de 2016].

CONSEJO INTERNACIONAL DE ARCHIVOS. ISDF. Norma internacional para la descripción de funciones Primera edición Elaborada por el Comité de Buenas Prácticas y Normas Profesionales Dresde, Alema-

- nia, 2-4 mayo 2007. Disponible en: http://www.mecd.gob.es/cultura-mecd/dms/mecd/cultura-mecd/areas-cultura/archivos/recursos-profesionales/normas-archivisticas/ISDF_ESP_definitiva.pdf [consultado el 23 de enero de 2012]
- CRUE. Conferencia de Archiveros de las Universidades Españolas. (2013). *Informe de la situación de los archivos universitarios españoles 2012*. p. 36. Disponible en: http://cau.crue.org/Documents/Jornadas/2013-Madrid/INFORME_ENCUESTA_2012.PDF [consultado en junio de 2016]
- Duchein, M. (1985). El respeto de los fondos en archivística: principios teóricos y problemas prácticos. En: Walne, P. *La administración moderna de archivos y la gestión de documentos: el prontuario RAMP*, Unesco. 69-92. Disponible en: <http://www.unesco.org/webworld/ramp/html/r8532f/r8532f00.htm>
- Franco Espiño, B. La normalización en el ámbito del CIA y el trabajo del Grupo de Expertos en Descripción Archivística (EGAD). *Jornada normalización de la descripción archivística: hacia un modelo conceptual Subdirección General de los Archivos Estatales*. Madrid, 12 de abril de 2013 CNEDA Beatriz Franco Espiño. Disponible en: http://www.mecd.gob.es/cultura-mecd/ca/dms/mecd/cultura-mecd/areas-cultura/archivos/novedades/normalizacion-de-la-descripcion-archivistica/JORNADACNEDA_Normalizacion.pdf [consultado en abril de 2013]
- Gobierno de Canarias. Cuadro de clasificación funcional de la Administración Pública de la Comunidad Autónoma de Canarias. Disponible en: <http://www.gobiernodecanarias.org/cpj/igs/temas/archivos/cuadroclasificacionfuncional.html> [consultado el 22 de diciembre de 2012]
- Gómez Díaz, R.; Bringas González, R. (2005). Normalización y requisitos funcionales de la descripción archivística: una propuesta metodológica. *Scire*, 11(1): 103-112. Disponible en: <http://ibersid.eu/ojs/index.php/scire/article/view/1510/1488> [consultado el 14 de marzo 2012]
- Grupo de Trabajo de Archivos Personales (2012). Propuesta de cuadro de clasificación común de fondos personales. V 2.0. *XVIII Jornadas de la Conferencia de Archiveros de la Universidades Españolas. Universidad de Cádiz 23-25 de mayo*. Disponible en: http://cau.crue.org/Documents/GT/GT-AP/XVIII_Jornadas_CAU_UCA_Cadiz_GT_Fondos_Personales_25_05_2012.pdf [consultado el 22/11/2012]
- Haas, J.K., y otros, (1985). *Appraising the records of modern science and technology : a Guide*. Massachusetts Institute of Technology. Disponible en: <http://babel.hathitrust.org/cgi/pt?id=mdp.39015011296897;view=1up;seq=96> [consultado el 3 de marzo de 2012]
- Harper, P. (2004). Preserving scientific archives: the work of the National Cataloguing Unit for the Archives of Contemporary Scientists. *Notes and records of the Royal Society of London*, 58 (2): 227-232.
- Harper, P. (2005). *Preservation and access issues in contemporary scientific scientific archives: a general overview*. Museu de Astronomia e Ciências Afins, Rio de Janeiro. p. 15. Disponible en: http://www.mast.br/pdf/harper_peter_preservation_and_access_issues_in_contemporary_scientific_archives_a_general_overview.pdf [consultado el 13 de febrero de 2012]
- Heredia, A. (2010). Clasificación, cuadros de clasificación y e-gestión documental. *Tabula Revista de Archivos de Castilla y León*, 13: 139-152.
- Heredia, A. (2010). La presencia de las relaciones en el vocabulario archivístico. *Revista d'arxius*, 9: 349-363.
- Heredia, A. (2011). *Lenguaje y vocabulario archivístico: algo más que un diccionario*. Andalucía, Consejería de Cultura, p. 221.
- Molina Nortes, J. (2013). *Metodología y Técnicas para la Gestión de Archivos Científicos del CSIC*, CSIC. Unidad de Recursos de Información Científica para la Investigación, 171. Disponible en: http://digital.csic.es/bitstream/10261/75302/1/curso_archivos_CSIC_2013.pdf
- Molina Nortes, J. (2014). Cuadro de Clasificación para centros del CSIC (1940-1980): hacia una normalización en la organización de la documentación de instituciones científicas. *Enredadera*, 23. Disponible en: <http://bibliotecas.csic.es/web/enredadera/numero-23> [consultado en junio de 2016]
- Müller, B. (2006) À la recherche des archives de la recherche. Problèmes de sens et enjeux scientifiques. *Genèses*, 2006/2, n° 63: 4-24.
- National Cataloguing Unit for the Archives of Contemporary Scientists. 2004. *Notes and records of the Royal Society of London*, 58(2): 227-232.
- Páez García, M.A. (2004), El cuadro de clasificación integrado: normalización de la clasificación archivística, *Revista PH*, 47, p. 1-12. Disponible en: <http://www.iaph.es/revistaph/index.php/revistaph/article/view/1691#.VGSELTSG8ao> [consultado el 16 de septiembre de 2012]
- Pérez Montes, C.M.; Caso Neyra, M. (2003). La importancia de un patrimonio documental: los archivos científicos. Disponible en: <http://digital.csic.es/handle/10261/2957> [consultado el 25 de enero de 2012]

- Puig-Samper Mulero, M.A. (coord.) (2007). *Tiempos de Investigación: JAE-CSIC cien años de ciencia en España*, Madrid, CSIC. p. 488.
- Romero, A.; Santesmeses, M.J. (eds.) (2008). *Cien años de política científica en España*. Madrid, Fundación BBVA. p. 415.
- Santamaría Gallo, A. (2006). La Norma Española de Descripción Archivística (NEDA): Análisis y Propuesta de Desarrollo. Madrid, Subdirección General de los Archivos Estatales, 27-07-2006. Disponible en: <http://www.mecd.gob.es/cultura-mecd/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/documentacion/NedaAnalisis-pdf/NedaAnalisis.pdf> [consultado en junio de 2016].
- Santos, P.R.E. (2008). A ciência, os cientistas e os seus arquivos. *Arquivo & Administração*, 7(1): 21-33. Disponible en: <http://157.86.8.70:8080/certifica/handle/icict/5280> [consultado en junio de 2016].
- Santos, P.R.E. (2009). Arquivologia no laboratório: teoria, métodos e práticas de uma disciplina. En: Conferencia da Suv e IV Encontro de Archivos Científicos, *A Natureza dos arquivos universitários e de instituições de pesquisa: perspectiva internacional*, 118-126. Rio de Janeiro.
- Vivas, A. (2000). Clasificación de fondos documentales de archivos históricos universitarios: El modelo del Archivo Histórico de la Universidad de Salamanca, *Cuadernos de documentación multimedia*, 10: 293-316.
- Welfelé, O. (1995). Les archives scientifiques en France. *Janus: Revue Archivistique*, 2: 52-54.
- Welfelé, O. (1999). Quels matériaux pour l'historien d'après-demain?: le devenir des Archives Scientifiques. En: *Sciences et Archives Contemporaines : actes de la table ronde organisée par l'Ecole Nationale du Patrimoine les 20,21 et 22 novembre 1996*. 103-126. Paris, Ecole Nationale du Patrimoine.
- Welfelé, O. (2004). A proveta arquivada. Reflexões sobre os arquivos e os documentos oriundos da prática científica contemporânea. *Revista da SBHC, Rio de Janeiro*, 2(1): 65-72.

ANEXO 2. CÓDIGOS DE FUNCIONES

1. Dirección y Gobierno

- DR1 para series propias de funciones directivas de un centro.
- JU1 para series propias de los órganos consultivos de un centro.

2. Administración General y Secretaría

- AD2 series de asuntos de administración general.
- PE2 series de recursos humanos.
- EC2 series de gestión económica.
- PA2 series de gestión patrimonial.
- EX2 series de gestión de relaciones exteriores y asuntos jurídicos.

3. Investigación

- IN3 documentación de gestión de la investigación.
- IN31 series de secciones, departamentos o grupos de investigación.
- IN32 series producidas por investigadores concretos.
- IN33 para colecciones especiales de documentos u objetos.

4. Actividad académica y cultural

- AC4 documentación de gestión de las actividades académico culturales de un centro.
- AC41 series de gestión de la actividad docente.
- AC42 series de gestión de congresos y conferencias.
- AC43 series de gestión de otras actividades culturales.

5. Servicios de apoyo a la investigación

- SP5 series de gestión de las publicaciones científicas.
- SI5 series de servicios de información bibliográfica y documental.
- SC5 series de servicios científico-técnicos.

ANEXO 3. CUADRO DE CLASIFICACIÓN

Instituto Español de Entomología

1. Dirección y Gobierno

1.1. Dirección (DR1)

Actividad normativa

Correspondencia de dirección

Informes de planificación y organización

1.2. Órganos Colegiados (JU1)

Expedientes de sesiones

2. Administración General y Secretaría

2.1 Comunicaciones y organización administrativa (AD2)

Correspondencia administrativa

Diarios de actividades

Libros de registro

Libros de visitas

Memorias

2.2. Recursos Humanos (PE2)

Expedientes de personal investigador

Expedientes informativos

2.3. Recursos Económicos (EC2)

Expedientes de elaboración del presupuesto

Expedientes de gastos

Expedientes de gastos de la exposición entomológica

Gestión de la caja pagadora

Registro de seguimiento contable

Nóminas

2.4. Recursos Patrimoniales (PA2)

2.4.1 Bienes muebles

Inventario de bienes muebles

2.4.2 Bienes inmuebles

Proyectos de obras

Expedientes de mantenimiento bienes inmuebles

2.5 Relaciones Externas y Asuntos Jurídicos (EX2)

Convenios

Expedientes de gestión de actividades externas

3. Investigación

Departamentos/Secciones/Grupos de Investigación (IN31)

Correspondencia científica

Diarios de actividades

Estudios e informes técnicos

Proyectos de investigación

Viajes y excursiones

4. Actividad Académica y Cultural

4.2. Congresos y Conferencias (AC42)

Congresos

4.3. Actividades Culturales (AC43)

Exposiciones y divulgación cultural

5. Servicios de Apoyo a la Investigación

5.1. Gestión de las Publicaciones Científicas (SP5)

Edición de las revistas EOS y GRAELLSIA

5.2. Servicios de Información (SI5)

Gestión de la biblioteca

ESTUDIOS / RESEARCH STUDIES

Gestión de fondos de archivos con datos enlazados y consultas federadas

Yusniel Hidalgo-Delgado*, José A. Senso**, Amed Leiva-Mederos***, Pedro Hípola**

* Departamento de Programación, Universidad de las Ciencias Informáticas, La Habana, Cuba

** Departamento de Información y Comunicación, Universidad de Granada, Granada

*** Universidad Central «Marta Abreu» de Las Villas, Cuba

Correos-e: yhidalgo@uci.cu; jsenso@ugr.es; amed@uclv.edu.cu; phipola@ugr.es

Recibido: 24-04-2015; 2ª versión: 11-09-2015; Aceptado: 26-10-2015

Cómo citar este artículo/Citation: Hidalgo-Delgado, Y.; Senso, J. A.; Leiva-Mederos, A.; Hípola, P. (2016). Gestión de fondos de archivos con datos enlazados y consultas federadas. *Revista Española de Documentación Científica*, 39(3): e145. doi: <http://dx.doi.org/10.3989/redc.2016.3.1299>

Resumen: En este trabajo se presentan las principales tecnologías de la Web Semántica que pueden ser de utilidad para la gestión de fondos archivísticos. Se examinan diversos proyectos de ámbito internacional y local que parten de descripciones normalizadas ISAD-G para generar ontologías, así como la disponibilidad de LIAM (Linked Archival Metadata), que facilita la transformación de datos de archivo a formato RDF (Resource Description Framework). Por otra parte, se analiza cómo la gestión de datos enlazados permite la interoperabilidad entre sistemas de información y la búsqueda facetada a partir de fondos documentales almacenados, descritos en OWL (Ontology Web Language), SKOS (Simple Knowledge Organization System) y Dublin Core. Los autores proponen la utilización de un CMS (Content Management System) que gestione fondos de archivo, compatible con SIOC (Semantically-Interlinked Online Communities) y OAI-PMH (Open Archives Initiative - Protocol Metadata Harvesting), para facilitar el intercambio y la recuperación de información. En concreto, se detallan las tecnologías que se han utilizado para desarrollar CoroArchivo, sistema que además se evalúa con un experimento que realiza la creación automática de ontologías a partir de descripciones ISAD-G almacenadas en DSpace. La herramienta desarrollada permite realizar consultas federadas sustentadas en las clases de exclusión e igualdad del vocabulario OWL.

Palabras clave: Datos enlazados; ontologías; archivos; repositorios; servicios de información; Drupal; búsqueda federada; DSpace.

Management of archival materials with Linked Data and federated queries

Abstract: In this paper the major technologies of the Semantic Web which may be useful for archives management are summarized. Several local and international projects that generate ontologies from standardized descriptions based on ISAD-G are examined. It is also discussed LIAM (Linked Archival Metadata), that facilitates the transformation of archive records into RFD (Resource Description Framework) format. Furthermore, we analyze how Linked Data enables interoperability between information systems and faceted search of OWL (Ontology Web Language), SKOS (Simple Knowledge Organization System) and Dublin Core records. The authors propose the use of a CMS (Content Management System) compatible with SIOC (Semantically-Interlinked Online Communities) and OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting) for archive records to improve the exchange and retrieval of information. We specifically describe the technologies used for developing CoroArchivo, system assessed by an experiment that automatically generates ontologies from ISAD-G records stored in DSpace. The evaluation tool lets users perform federated queries based on the OWL vocabulary disjointness and equivalent classes.

Keywords: Linked Data; ontologies; archives; repositories; information services; Drupal; federated queries; DSpace.

Copyright: © 2016 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Attribution (CC BY) España 3.0.

1. INTRODUCCIÓN

La Web Semántica (Berners-Lee y otros, 2001) se ha ido implantando en diferentes ámbitos de la actividad documental, como el procesamiento automatizado de la información, el control de autoridades, la recuperación de datos enlazados, los servicios de resumen... Y parece claro que diversas iniciativas de la Web Semántica pueden ser de gran utilidad en la gestión archivística (Mena, 2006; Hernández, 2007). Una de las iniciativas que más recorrido tiene dentro del ámbito de la gestión de información es Linked Data, o datos enlazados, entendido como el conjunto de buenas prácticas para la publicación de datos con el fin de facilitar su reutilización. Esta idea poco a poco se va asentando en la comunidad archivística.

Tanto es así que se han realizado aproximaciones desde distintos puntos de partida. Sin pretender realizar aquí un recorrido exhaustivo, ya que no es el objetivo principal de este trabajo, podemos agrupar en varias corrientes las diferentes formas de afrontar la publicación de datos enlazados en el terreno de los archivos. Por un lado, aquellos proyectos que intentan introducir elementos semánticos en los registros para su posterior conversión a RDF. Por otro, aquellos que se centran principalmente en el control de autoridades y materias mediante diversos mecanismos de transformación.

Dentro del primer grupo, uno de los trabajos más destacados es el de Vasallo que, de forma teórica, muestra cómo realizar la transformación de la norma Topic Maps en RDF (Vasallo, 2010). Este proceso, dentro del ámbito archivístico, transforma elementos de la norma ISAAR (CPF) en relaciones Scope y asociaciones Topic Map. Aunque se trata de un caso muy puntual y difícilmente extrapolable a otras realidades, es interesante porque pone de manifiesto algunos puntos de contacto entre la semántica y la descripción archivística. Evidentemente sus limitaciones en el ámbito de la gestión de autoridades y de entidades externas hacen que sea difícil desarrollar la propuesta. Gracy, por su parte, plantea una metodología que permita la transformación del formato EAD en Linked Data, estableciendo un modelo de datos para que el formato archivístico pueda interoperar con clases y propiedades de ontologías específicas (Gracy, 2014). La autora propone alinear etiquetas EAD con la DBpedia, detectando aquéllas que puedan ser utilizadas en los procesos de descripción archivística, con el fin de enriquecer los registros. En una segunda fase se establecen coincidencias con los vocabularios FOAF, el formato LOD (Linking Open Description of Events) y un conjunto de metadatos geográficos. Los resultados de esta metodología muestran las enormes posibilidades que ofrece EAD en combi-

nación con otros vocabularios ontológicos y su salida Linked Data. Para gestionar el vocabulario se emplea la herramienta de procesamiento de lenguaje natural Open Calais, actualmente propiedad de Thomson Reuters. El principal problema de la arquitectura propuesta por Gracy reside en que genera constantes inconsistencias en los registros, a causa de sus escasos mecanismos de limpieza de datos, con lo que se complica mucho el proceso de reutilización posterior. Además, no se ha verificado la efectividad del proceso por medio de ningún mecanismo de evaluación. Algo similar es lo que sucede con el trabajo de Nam-Park, que transforma los registros del Archivo Nacional de Corea en datos enlazados usando SKOS y Dublin Core para facilitar la conversión en RDF, lo que favorece la realización de búsquedas flexibles (Nam-Park, 2015). El gran logro de este proyecto, desde nuestro punto de vista, reside en lograr la interoperabilidad entre aplicaciones al ofrecer servicios de información con datos extraídos tanto de una biblioteca, un museo y los archivos, a través de la conexión de varias bases de datos. Sin embargo, el resultado genera registros que no pueden ser enlazados, bien por estar incompletos, bien por falta de enlaces internos.

Consideramos que el trabajo más valioso en materia de transformación de datos es el de Rademaker, que propone una arquitectura funcional para realizar la transformación a Linked Data empleando para ello los registros de la Fundación Getulio Vargas (Rademaker y otros, 2015). La clave del proceso está en la migración de datos, que comienza con la transformación de una base de datos relacional a OAI-PMH, permitiendo la exportación a RDF mediante la herramienta D2RQ. Esto facilita el mapeo de los datos y la construcción de un grafo que, después de alinearlos con varios vocabularios (FOAF, SKOS, Dublin Core y Prov), se almacena en una tripleta y es indexado con Apache Solr (servidor open source en Java para realizar búsquedas). Los datos pueden ser consultados mediante SPARQL.

La segunda aproximación se centra en proponer un cambio estructural para la sintaxis del registro de archivo. En ese sentido el trabajo que en 2010 realizó el EACWG (Encoded Archival Context Working Group) con EAC-CPF (Encoded Archival Context – Corporate Bodies, Persons, and Families) marca un punto de inflexión, al corregir los errores relacionados con el control de autoridades dentro de EAD. Esta nueva norma facilita la interoperabilidad al proporcionar un mecanismo que permite desarrollar la norma ISAAR (CPF) en formato de ontología. Además de controlar autoridades, esta pauta permite generar enlaces a recursos de autoridad externos, empleando para ello un subconjunto de etiquetas de la recomendación Xlink.

Muy relacionado con EAC-CPF se encuentra el proyecto SNAC (Social Networks and Archival Context), desarrollado por Daniel Pitti, que perfecciona los mecanismos para el control de autoridades y gestión de puntos de acceso (Lynch, 2014). Para ello extrae las etiquetas EAC-CPF de los registros EAD, y se mapean con diferentes esquemas (VIAF, Library of Congress Name Authority File y Union List of Artist Names). De esta manera las entradas de autoridad no solo son controladas por EAC-CPF, sino que existirán entradas alternativas para los títulos a través de VIAF (Virtual International Authority File), y para datos históricos mediante los otros dos sistemas. SNAC crea además un prototipo para acceder a recursos de archivo con carácter histórico a partir de los datos de redes sociales y profesionales, enlazando recursos de archivos, bibliotecas y museos.

Más allá de la simple conversión de datos en RDF, han aparecido estudios que analizan la gestión de vocabularios controlados para buscar la interoperabilidad y mejorar la búsqueda y la recuperación de la información. En este sentido destaca el trabajo de Grimoüard, que estudia las posibilidades para la gestión de materias mediante endpoints de SPARQL en archivos franceses (Grimoüard, 2014).

Hay que destacar, para finalizar, el desarrollo de LIAM (Linked Archival Metadata), un formato de metadatos para la transformación de descripciones archivísticas en RDF, que facilita las transformaciones que se hacían sobre EAD para la conversión de datos de archivo, así como los esfuerzos de EGAD (Expert Group on Archival Description) por lograr una nueva versión de la norma ISAD-G más relacional.

La Web Semántica enriquece a la archivística no solo aumentando el valor de las descripciones, como se refleja en la mayoría de trabajos aquí descritos, o en proyectos como Europeana o Archive Hub. Estas iniciativas dibujan también una realidad: los datos de los archivos cobran más relevancia que un simple registro descriptivo, puesto que se aporta la posibilidad de la reutilización, revalorizando el trabajo que los profesionales realizan.

La principal ventaja de LOD está en la reutilización, por lo que la forma en la que se consiga llegar a ella tampoco ha de ser tan determinante, aunque queda claro que este proceso debería evitar pérdidas de información y mantener, en la medida de lo posible, la descripción realizada y las relaciones del objeto descrito con su entorno. Y de ahí parte nuestro trabajo: sobre la base de lo explicado anteriormente, hemos desarrollado una aplicación, CoroArchivo, que facilita la descripción de documentos de archivo en universidades y la

gestión de la información mediante Linked Data. El sistema emplea Drupal como CMS, y DSpace como proveedor de datos. Se trata de una herramienta de trabajo en entorno web capaz de gestionar grandes conjuntos de datos transformados en descripciones RDF, y que tiene como principal objetivo el procesamiento y el intercambio de información entre archivos de universidades.

2. TECNOLOGÍAS DISPONIBLES

2.1. OAI-PMH

Existen diferentes modelos para el tratamiento de datos enlazados dentro de repositorios OAI-PMH. En la mayoría de los casos las plataformas empleadas, el software soportado o los proyectos de aplicación, suelen determinar cómo proceder con cada uno de ellos. Resultaría realmente complejo categorizar todas las propuestas que existen, ya que cada una de ellas se adapta a la realidad tecnológica con la que tiene que trabajar, así como a la naturaleza de los datos. Pero sí destacaremos alguno que nos ha servido a modo de guía o patrón, como por ejemplo el trabajo de Coppens, que aporta una forma muy novedosa de gestionar fondos de archivo empleando datos enlazados por medio de los estándares de descripción EAD e ISAD-G (Coppens y otros, 2009). En el mismo sentido, la forma en la que Europeana ha integrado los materiales de archivos, bibliotecas y museos (Haslhofer y otros, 2011; Doerr y otros, 2011), o Nestor, una aplicación construida a partir de la experiencia práctica de varios archiveros, y que emplea EAD para los metadatos (Ferro y Silvello, 2013), son ejemplos clave.

Existe un grupo de iniciativas que, para preservar datos digitales, ha optado por OAI-PMH por las ventajas que presenta este modelo. Nos hemos centrado en el estudio de Preso, que busca la creación de espacios para preservar materiales fílmicos y fotográficos. Por su ámbito de actuación, la Unión Europea, se orienta a ofrecer soporte para Europeana (Addis y otros, 2010).

Aunque el uso de OAI-PMH está bastante generalizado, también se han estudiado aquellos proyectos que proponen complementar este protocolo por medio de diferentes mecanismos. Así OAI2LOD es un servidor, compatible con protocolos de interoperabilidad, que permite consultar los metadatos con SPARQL (Sparql Protocol and RDF Query Language). Realmente esto supone un salto cualitativo con respecto al resto de proyectos, ya que resulta posible sacar más partido a las relaciones existentes entre los datos. En concreto el trabajo con las relaciones "same as" entre dos instancias

del servidor aporta una mayor riqueza en las búsquedas (Haslhofer y Schnadl, 2008, 2010). Para lograr esto, el software servidor compara los valores de un conjunto de atributos seleccionados manualmente según su similaridad léxica usando la distancia Levensthein. Es un sistema modular, ya que para la transformación de los datos en RDF/XML se emplea otro servidor que trabaja de forma paralela, D2RQ Server (Bizer y Seaborne, 2004). Uno de los principales problemas detectados en OAI2LOD, la imposibilidad de compaginar varios proveedores de datos, se resolvió con importantes mejoras realizadas por Coppens, que permitieron la importación de datos procedentes de diversos repositorios OAI-PMH (Coppens y otros, 2013).

2.2. Búsqueda federada

Una de las áreas más investigadas dentro del ámbito de la Web Semántica ha sido la recuperación de información. Entre las primeras aproximaciones destaca Síndice, que embebe datos RDF y los microformatos que los describen en un servidor que se consulta por medio de un SPARQL Endpoint, permitiendo realizar consultas semánticas sobre los recursos almacenados (Tummarello y otros, 2007). Otras aportaciones que han servido como base para el trabajo que se describe en este texto es el método de construcción de índices en tiempo real que emplea Squin (Hartig y otros, 2009), y las consultas basadas en reglas que emplea FedX, que por medio del framework Sesame permite la optimización de consultas federadas (Schwarte y otros, 2011).

En esa misma línea, pero empleando técnicas diferentes, se encuentra el motor Darq, que complementa ese tipo de búsquedas con un sistema de toma de decisiones para refinar las búsquedas federadas (Quilitz y Leser, 2008). Existen otros sistemas que se valen de información estadística para realizar procesos similares, como Splendid (Görlitz y Staab, 2011). El problema es que ambos métodos se centran demasiado en técnicas complementarias y no exploran lo suficiente las facilidades semánticas de recuperación de información a través del lenguaje OWL. Precisamente por ese motivo consideramos tan destacada la aportación de Coppens, que gestiona datos enlazados mediante un índice estructurado para optimizar las consultas *OWL:same as* en los Endpoints locales y remotos de SPARQL de forma simultánea (Coppens y otros, 2013).

También ha servido de base para el proyecto CoroArchivo el motor de consultas Elite, que permite mapear con gran facilidad las relaciones entre los datos por medio de estructuras R-Tree,

integrando información de diversa índole (incluidas tripletas RDF) con bases de datos relacionales (Nolle y Nemirovski, 2013).

La combinación de estas tecnologías ha servido para desarrollar el sistema de búsqueda federada de nuestro proyecto, enriqueciendo las consultas del tipo *OWL:same as*. Se ha optado por mapear las clases que se obtienen de la búsqueda distribuida para filtrar mejor la información. La similitud de los datos se obtiene mediante el cálculo de la distancia Manhattan. Los resultados se complementan con las etiquetas *OWL:equivalentClass*, *OWL:equivalentProperty*, *OWL:differentFrom* y *OWL:AllDifferent*, combinadas con el algoritmo de OWL2tips (Domínguez-Velasco, 2013a). Al igual que en Elite, nuestra propuesta emplea un razonador para formular las consultas y un módulo de indexación distribuida que conecta el sistema con ontologías, tanto locales como externas, a partir de mapas de referencias.

3. COMPONENTES DEL SISTEMA

3.1. Herramientas DSpace para el trabajo con Linked Data

Hemos utilizado el software open source DSpace, que gestiona repositorios compatibles OAI-PMH usando un solo data center.

A pesar de las deficiencias que se le atribuyen a la filosofía OAI-PMH, porque, al solo contar con el protocolo de metadatos OAI-DC (Schöpfel y otros, 2012), gestiona con austeridad las descripciones bibliográficas, y porque a veces forma parte de la web invisible y resulta poco accesible para los servicios de búsqueda (Merlino-Santesteban, 2012), DSpace ha ganado popularidad cuando gestiona datos semánticos sobre todo gracias a:

- el uso de gramáticas de contexto entre repositorios
- la gestión y distribución de todo tipo de metadatos
- la posibilidad de que se recupere información de varios repositorios con una misma interfaz
- su capacidad para generar búsquedas estructuradas y consultas remotas de datos enlazados
- sus funcionalidades para publicar datos enlazados
- las posibilidades que ofrece para construir léxicos, tesauros y ontologías.

3.2. Diseño del sistema

3.2.1. Transformación y desarrollo de ontologías para DSpace

Hemos partido de tres ontologías:

- una en SKOS (Solomou y Papatheodorou, 2010) con funciones de tesaurus
- otra para FOAF (Brickley y Miller, 2010), que sirve de herramienta de control de autoridades
- una tercera especial para la norma ISAD-G en OWL, que enriquece la descripción archivística y aprovecha las propiedades de la estructura semántica en la búsqueda federada.

A continuación se explica el proceso de generación de una ontología en ISAD-G y sus vocabularios.

La descripción archivística ISAD-G gira en torno a normas de procesamiento documental que podríamos considerar algo alejadas de otros formatos que se utilizan habitualmente en la Web Semántica. Para construir una ontología que procesara documentos de archivo, partiendo de registros ISAD-G, se conservaron todos los elementos de la descripción transformándolos a formato OWL. Para ello se siguió el mismo proceso planteado en anteriores proyectos (Koutsomitropoulos y otros, 2008; Baker, 2012). Se utilizaron las propiedades de los objetos y de los datos junto a sus especificaciones de cardinalidad. La ontología generada en OWL fue convertida a XML, facilitando la transformación de los elementos de la descripción archivística en el entorno del repositorio DSpace, como en otros proyectos similares (Koutsomitropoulos y otros, 2008). Lo primero que

se realiza en este tipo de conversión es la transformación de la sintaxis. A continuación la semántica.

- Transformación de la sintaxis: se partió de la estructura sintáctica de OWL, de SKOS y de Dublin Core para ajustar los datos a la estructura de DSpace. En esta etapa es necesario mostrar todos los datos de OWL en XML, sin que se pierda la riqueza de las descripciones al integrarlas en Protégé, el software gestor de ontologías elegido en nuestro caso.

Lo primero que se hizo fue construir tablas de convergencia de datos para que las clases principales de los vocabularios no se perdieran. Se decidió utilizar Sampras (Domínguez-Velasco, 2013b), sistema capaz de conservar las propiedades de todos los "datatypes" usados en el diseño del modelo ontológico, estructurar los "syntax encoding schemes" y "cosificar" los datos de Protégé. Con SKOS se mejoró el control del vocabulario; de FOAF para propiciar el control de autoridades. OWL sirvió para buscar relaciones de inclusión y exclusión a través de elementos útiles en la recuperación de la información. El proceso permitió asignar los valores exactos de cada campo en el modelo de OAI-PMH, lo cual se logra con indicadores de relación, reglas de construcción o constructores y refinamiento de datos. Concluido el proceso de cosificación de datos, se importó el resultado dentro de DSpace, que genera un archivo para su visualización en formato XML donde subyacen las estructuras de los tres lenguajes (tabla I).

Tabla I. Equivalencia de la ontología en SKOS

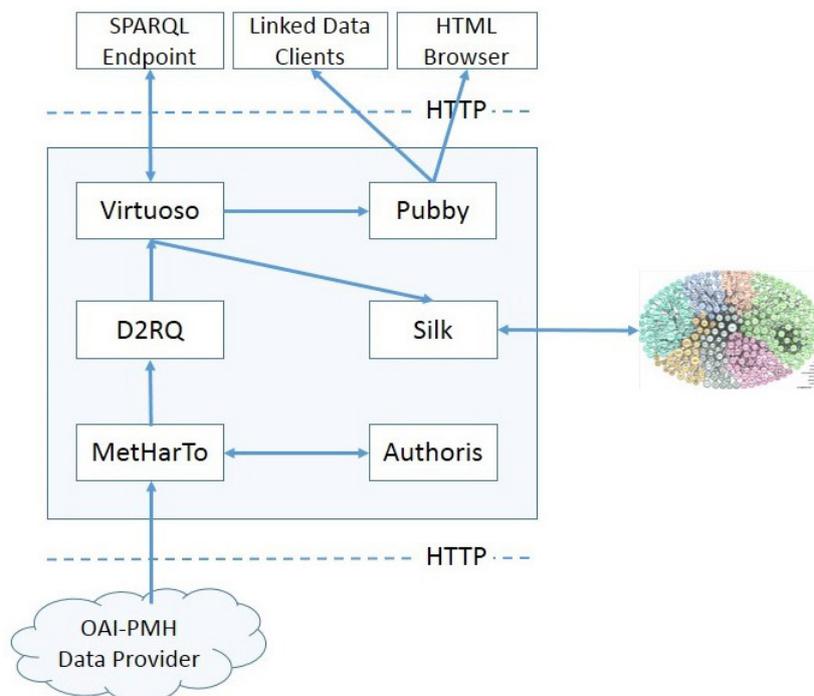
Elemento de XML	Función	Notación de SKOS
TC	Describe un término descrito bajo entrada principal	<skos:Concept>
ET	Traducción del Término al inglés	<skos:prefLabel lang="en">
RT	Traducción del Término al Ruso	<skos:prefLabel lang="ru">
PT	Traducción del Término al portugués	<skos:prefLabel lang="po">
ET ¹	Traducción alternativa al inglés	<skos:altLabel lang="en">
RT ¹	Traducción alternativa al ruso	<skos:altLabel lang="ru">
PT ¹	Traducción alternativa al portugués	<skos:altLabel lang="po">
TR	Término relacionado	<skos:related>
NA	Pequeña descripción	<skos:definition>
DC	Clasificación Dewey	<skos:notation>
USE	Remite de un término no autorizado a uno autorizado	<skos:altLabel lang="el">

- Silk (Jentzsch y otros, 2010) para la construcción de enlaces de datos en recursos disímiles, uniendo los enlaces RDF de las fuentes particulares con otras fuentes de datos. Dispone de una consola para realizar enlaces entre dos datasets y un servidor http capaz de recibir datos e introducirlos en el flujo RDF a través de "data ítems". La flexibilidad y calidad de los datos que gestiona el lenguaje Silk-LSL permite adaptarlos a la filosofía de trabajo Linked Data.
- Metharto (Hidalgo-Delgado y otros, 2013), capaz de extraer los metadatos de archivo y los elementos de SKOS e ISAD-G del proveedor de datos soportado en OAI-PMH.
- OpenLink Virtuoso (Haslhofer y Schnadl, 2008), servidor universal que combina las funcionalidades de los tradicionales gestores de bases de datos relacionales, bases de datos virtuales, RDF, XML, y que facilita el uso de texto libre en aplicaciones web.
- D2RQ (Bizer y Seaborne, 2004), lenguaje de mapeo que lleva a cabo la relación entre bases de datos y lenguajes semánticos RDF y vocabularios OWL, capaz de transformar un documento RDF en la estructura sintáctica de turtle. Los mapas que se generan definen un grafo en RDF donde se incluye toda la infor-

mación de la base de datos al igual que en SQL, solo que los datos generados en la estructura del RDF siempre van a exigir tablas relacionales virtuales. Se puede acceder a la plataforma RDF de varias formas según las necesidades de implementación. Por ejemplo: acceso a SPARQL, al servidor para datos enlazados, a un almacén de datos RDF, a una interfaz simple para RDF que convierte los datos bibliográficos a formato LIAM, y a una API de Jena que interopera bases de datos DR2Q.

El flujo de trabajo en el tratamiento de las ontologías y la implementación de Linked Data ha sido el siguiente. Metharto extrae del proveedor de datos las descripciones bibliográficas necesarias y se comunica con Authoris para determinar las entradas de autoridad a partir de reglas de decisión (Leiva-Mederos y otros, 2013). D2RQ genera el mapa del archivo con todos los campos y tablas organizados en un esquema relacional. Las tablas son la equivalencia de las clases, subclases e instancias previamente diseñadas en las ontologías y exportadas a DSpace. Con el script para RDF denominado Dump-rdf se genera un grafo que se aloja en Virtuoso, servidor que provee triple stores y permite consultas SPARQL. Teniendo los datos en Virtuoso se puede configurar Silk para obtener enlaces en lenguaje OWL a través de todos los grafos del sistema, de VIAF, de DLBP (Digital Bibliography & Library Project) y de Europeana (figura 2).

Figura 2. CoroArchivo y la tecnología OAI2LOD (Haslhofer y Schnadl, 2008)



3.3. Funcionamiento de la plataforma

El sistema está operativo en la Universidad Central de las Villas (UCLV), donde el convenio de universidades flamencas VLIR tiene instaladas las infraestructuras tecnológicas necesarias para su funcionamiento (tabla II).

A continuación se detalla el proceso de transformación de los datos:

- Paso 1. Extracción de metadatos: haciendo uso de Metharto se localizan los SPARQL Endpoints de los fondos de archivo y se generan las tablas relacionales con los datos para los diversos niveles de descripción ISAD-G y los elementos de SKOS.
- Paso 2. Pre-procesamiento: los datos que se encuentran en DSpace, independientemente de si ya han sido transformados o no, se procesan para llevar a cabo el control de autoridades y que tengan equivalencia con la estructura OAI-PMH. Por medio de reglas preestablecidas se detectan de forma automática los autores duplicados, se asignan títulos uniformes para documentos y publicaciones, y descriptores de materia dentro del entorno de Linked Data.

Los datos se modelaron utilizando ontologías específicas para la archivística. En el caso de registros bibliográficos de otro tipo de materiales es posible usar Fabio (Peroni y Shotton, 2012) o Bibo (Giasson y D'arcus, 2009), ambas de aplicación dentro del ámbito bibliotecario. En nuestro caso se recurrió a una ontología en formato OWL creada ad hoc por Álvarez (Ledesma, 2013) siguiendo las propuestas de terceros (Sánchez Alonso y otros, 2008). Los elementos de materia se estructuraron con SKOS (Solomou y Papatheodorou, 2010), y las entradas de autoridad de acuerdo con FOAF (Brickley y Miller, 2010).

- Paso 3. Transformación: con el editor de ontologías Protégé se exportaron los datos a DSpace, y con 2RQ se transformaron las bases de datos relacionales en grafos RDF usando un script en lenguaje bash que permite incorporarlos a Virtuoso. En 2RQ existe una herramienta denominada D2R que facilita la transformación de las consultas SPARQL en SQL, y su posterior ejecución sobre las bases de datos relacionales, lo cual es posible cargando y creando un grafo RDF externo.
- Paso 4. Enlace de datos: empleando Silk se establecieron enlaces RDF que apuntaban a fuentes de datos de archivo. De este modo cada conjunto de datos externos quedó accesible para las aplicaciones en forma de triple store. Los datasets VIAF, DLBP y Europea se descargan en Virtuoso. Disponibles ya en formato SPARQL se configuró Silk para generar enlaces de exclusión e inclusión, realizando el control de autoridades en los datasets con reglas de decisión. Es decir, de forma diferente a otros proyectos comentados en el apartado 2.2, que usan distancias geométricas para establecer clústers de coincidencia o similitud. Este proceso finaliza con la transformación de los datos a formato LIAM.
- Paso 5. Publicación: los grafos RDF quedan disponibles en la web para los usuarios y para cualquier otro sistema de información. Se decidió usar Pubby (Cyganiak y Bizer, 2008) con SPARQL, ya que rescribe la URI base, a diferencia de SPARQL, que lanza las consultas contra los RDF subyacentes. Como interfaz para la recuperación de información se escogió Drupal, que da soporte a toda esta tecnología.

Tabla II. Conjunto de repositorios de la UCLV

Nombre del Repositorio	OAI-PMH Endpoint
Fondo Coronado Fotos	http://fotos.coroando.uclv.edu.cu/oai
Fondo Coronado Cartogramas	http://cartogramas.uclv.edu.cu/oai
Manuscritos Coronado	http://manuscritos.uclv.edu.cu/oai
Carteles Coronado	http://carteles.uclv.edu.cu/oai
Revista Coronado	http://revistas.uclv.edu.cu/oai
Archivo Histórico Universitario	http://archiuniv.uclv.edu.cu/oai
Archivo de Central de la Universidad	http://centralarchv.uclv.edu.cu/oai

3.4. Módulos de búsqueda semántica y facetada

Drupal, desarrollado a partir de 1999 por Dries Buytaert, y publicado bajo licencia GNU dos años más tarde, genera sitios web combinando varios "bloques" para adaptar sus funcionalidades a las necesidades específicas. Es además un "content management framework" (Byron y otros, 2012). La información se almacena en una base de datos relacional (MySQL, PostgreSQL, SQLite...) por medio del lenguaje de programación PHP. Permite la publicación de datos en formato RDF, además de soportar otros, como n-triples, JSON, XML, RSS y turtle. Gestiona las URIs de los materiales RDF publicados y un SPARQL Endpoint para la consulta. Se pueden personalizar los campos RDF y los namespaces (Alonso-Sierra y otros, 2012).

La búsqueda facetada y las consultas en SPARQL se realizan por medio de los siguientes módulos de Drupal:

- RDF Extensions 7.x-2.0-alpha1
- SPARQL 7.x-2.0-alpha1
- Views 7.x-3.0-beta3
- SPARQL Views 7.x-2.0-alpha2
- Entity API 7.x-1.0-beta8
- CTools 7.x-1.0-alpha4
- SPARQL Views
- RDF UI
- JSONP SPARQL
- OSF (Open Semantic Framework)
- Views UI modules
- OWL2Tips
- DSpace

Instalados los módulos necesarios, OWL2tips (Domínguez-Velasco, 2013a) facilita la visualización en formato OWL y la transformación de las descripciones archivísticas a estructura LIAM, capaz de operar con grandes conjuntos de datos.

3.4.1. Búsqueda federada

Para generar las búsquedas facetadas de ontologías en Drupal se utiliza una versión modificada de la propuesta de Coppens (Coppens y otros, 2013) junto a OWL2tips (Domínguez-Velasco, 2013a). El gran problema de Drupal para la gestión de ontologías es su apego a vocabularios específicos de tratamiento de datos. Lenguajes de ontologías como SKOS, Dublin Core o SIOC son ineficientes cuando se trata de documentación archivística y

de búsqueda de información federada. Hay muchos elementos semánticos que provienen del lenguaje OWL que quedan fuera de las consultas clásicas de SPARQL, lo que impide la elaboración de consultas más sofisticadas que exploten las posibilidades de la transitividad y la simetría cuando se trata de localizar información en varios recursos.

El módulo de búsqueda utiliza los elementos OWL y los enlaces de igualdad y de exclusión del lenguaje para establecer nuevas consultas dirigidas a puntos específicos de la estructura OWL, de manera que se pueda acceder a diferentes recursos de información en bases de datos remotas. OWL2tips usa SPARQL Endpoints remotos capaces de enlazar las consultas con cada dependencia del archivo de la Universidad, de la DBLP, VIAF y Europeana. Existen muchos problemas cuando se intenta realizar búsquedas federadas a partir de vocabularios específicos y en repositorios diferentes, pues:

- Cada recurso necesita un SPARQL Endpoint que permita construir relaciones dentro de él.
- Cuando mapeamos un vocabulario no se puede usar una misma consulta para recuperar información en recursos diferentes.
- Es imprescindible trabajar con la especificidad en las consultas y no utilizar elementos genéricos en la descripción, por ejemplo foaf:name en vez de RDFs.
- Las consultas no son efectivas sino rígidas, y obligan a que los usuarios empleen lenguajes de consulta que no conocen y que les hacen complejo localizar la información.
- Los usuarios no pueden recuperar toda la información que requieren, y sus consultas se hacen ineficientes porque no abarcan elementos que la búsqueda clásica de SPARQL es incapaz de gestionar.

Como solución hemos utilizado los enlaces que proceden del lenguaje OWL y constructores que generan un índice de búsqueda que indexa cada elemento de las ontologías. Para esto se almacenan los fondos documentales en los Service Endpoint, y las propiedades de mapeo de los elementos junto a los mapas de cada clase. La información se extrae de las consultas realizadas en lenguaje SPARQL con ayuda de los constructores utilizados en la definición de la ontología, combinados con las propiedades *OWL:same as*, *OWL:equivalentClass*, *OWL:equivalentProperty*, *OWL:differentFrom* y *OWL:AllDifferent* como enlaces. El proceso de indexación está precoordinado para facilitar la recuperación de información, con funcionalidades preestablecidas que enlazan los recursos. Gracias al procesador distribuido de SPARQL es posible

utilizar información pre-indexada para las nuevas consultas que entran en la plataforma y distribuir búsquedas derivadas de consultas generales de acuerdo con las necesidades de los usuarios potenciales del sistema. Esto se consigue recurriendo a la extensión ARQ del módulo JSONP SPARQL de Drupal, con tres procedimientos que permiten consultas federadas: 1) la generación del índice de distribución para procesar los enlaces que describen igualdad y desigualdad de clases; 2) la distribución de las consultas (usando un index building) que facilita la búsqueda federada y el mapeo de las búsquedas a través de índices distribuidos; y 3) la generación de asociaciones de clases e instancias a través de mecanismos de aprendizaje difusos insertados en la herramienta OWL2tips (figura 3).

Figura 3. Segmento de una sentencia SPARQL en CoroArchivo

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX arch: <http://archive.uclv.edu.cu/localontology/namespace#>
CONSTRUCT {?resource owl:sameAs ?remotesource}
WHERE {
  ?resource a owl:thing.
  ?resource arch:type "person".
  ?resource arch:name ?concept.
  ?concept skos:prefLabel ?name.
  ?concept skos:Concept ?name.
  SERVICE <http://archiveuclv.org/sparql>
  {
    ?remotesource a foaf:Person.
    ?remotesource foaf:name ?remotename.
    FILTER (str(?remotename) = ?name)
  }
}
```

OWL2tips trabaja con las relaciones de diferenciación e igualdad descritas en las ontologías para distribuir las búsquedas y extraer los mapas RDF de cada estructura dentro del módulo JSONP SPARQL de Drupal. Se hacen tres usos concretos de OWL2tips en esta fase:

- Para enriquecer la indexación junto a SPARQL: de esa manera es posible generar consultas que representan el origen de las relaciones de tipo *OWL:same as*, *OWL:equivalentClass*, *OWL:equivalentProperty*, *OWL:differentFrom* y *OWL:AllDifferent*, usadas como medio para hacer enlaces en el dataset. Esto lo logramos mediante constructores para facilitar enlaces internos dentro del

dataset. Las consultas son descritas y procesadas con el álgebra SPARQL, y expresadas mediante el algoritmo que maneja OWL2tips que se ha diseñado para trabajar con Drupal específicamente en este proyecto.

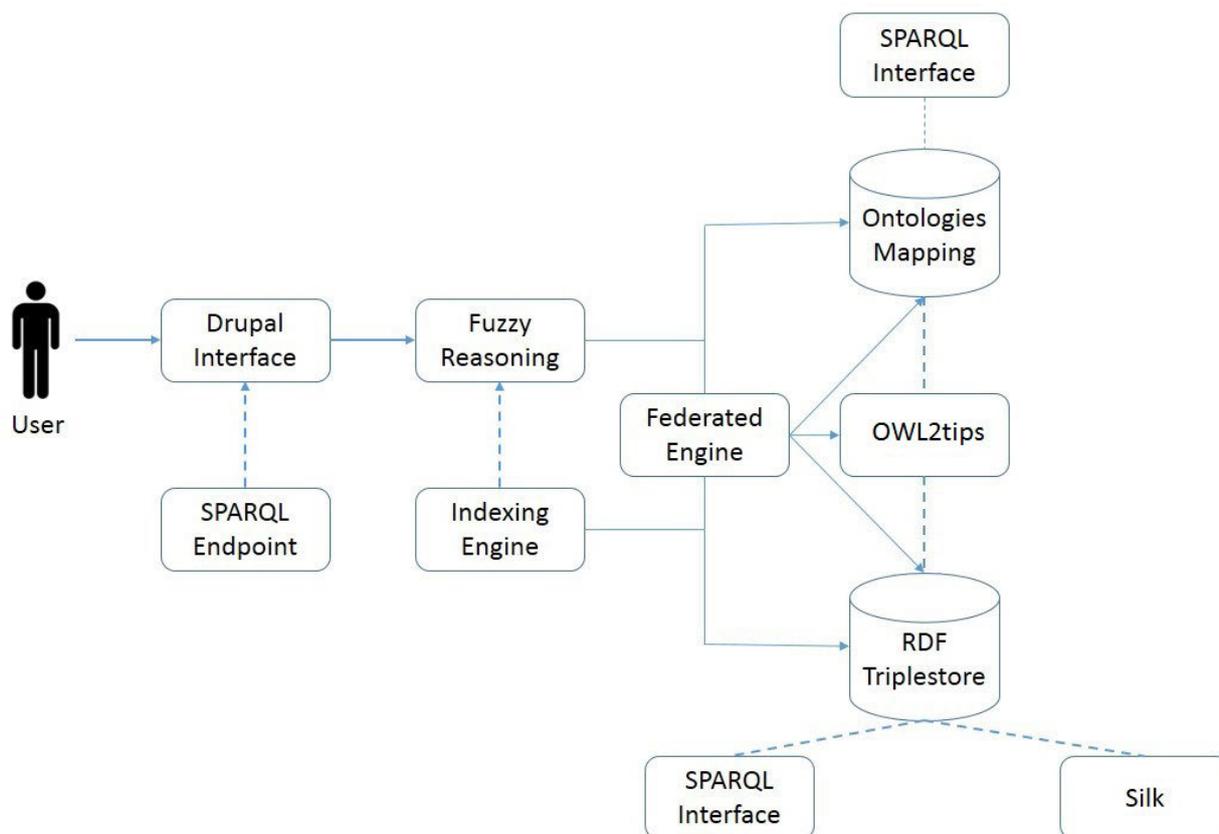
- Para complementar los índices de OWL2tips con Silk: creando un archivo con las especificaciones para generar enlaces dentro de Silk y empleando un sistema de indexación capaz de reconocer el origen de los enlaces SPARQL por medio de reglas o constructores de consultas. El sistema de indexación de Silk lleva a cabo un parsing del archivo de configuración de la herramienta Silk y crea los índices mediante los Endpoints de SPARQL, los mapas de clases y las propiedades específicas de los SPARQL Endpoints disponibles directamente en el archivo de configuración para cada una de las propiedades.
- Para aproximar las consultas valiéndose de las posibilidades que ofrecen los enlaces de las propiedades *OWL:same as*, *OWL:equivalentClass*, *OWL:equivalentProperty*, *OWL:differentFrom* y *OWL:AllDifferent*, utilizando algoritmos de aprendizaje basados en lógica difusa, con lo que se logra recuperar con más precisión los llamados conjuntos de datos intermedios, que muchas veces se solapan cuando se recupera la información (figura 4).

4. PROCESO DE EVALUACIÓN

4.1. El dataset de entrada

Coroimagen UCLV es un conjunto de datos que posee 1.422 instancias, más de 68 datasets y 5 servicios de archivo descargados de VIAF, Europeana, DLBP y de los archivos de la propia institución donde se realiza la prueba. La mayor complejidad para manejar estos recursos es establecer la correspondencia entre los datasets con las fuentes de información, como sucede con DSnotify (Popitsch y Haslhofer, 2011).

Se creó, por ese motivo, un archivo que establece la relación entre los elementos utilizados en la descripción que hace posible medir el nivel de eficiencia de CoroArchivo a partir de los eventos que registra la aplicación y su influencia en f-measure. Dicho archivo sirvió también para detectar que existía un elevado número de instancias, 98, que estaban mal escritas. Para subsanar este error se crearon indicadores de exclusión que aislaran del experimento aquellas instancias cuyo nivel de ocurrencia afectaba al proceso de evaluación. Así fue posible lograr una distribución homogénea de eventos en un tiempo determinado y minimizar los problemas de entropía.

Figura. 4. Sistema de búsqueda federada de CoroArchivo

4.2. Test

Se decidió realizar la evaluación de acuerdo con los procedimientos de Popitsh y Haslhofer (Popitsh y Haslhofer, 2011), que verifica el funcionamiento de una plataforma con información que implica control de autoridades. En nuestro caso queríamos analizar en qué medida la precisión y la exhaustividad (F-measure) influían en el rendimiento de la herramienta. Para ello se utilizó: un vocabulario capaz de almacenar la información del proceso (Leiva-Mederos y otros, 2013); un módulo de monitorización de eventos; un simulador encargado de representar los datasets en otros entornos de la Universidad; y un analizador que mostraba y examinaba los parámetros seleccionados para emitir los resultados de la evaluación.

Con un algoritmo denominado jump (Domínguez-Velasco y Leiva-Mederos, 2013) se registraron los valores de entropía de las propiedades de los namespaces que se alojan en la plataforma con una cobertura menor del 20%. Estos valores se originaban en caso de que los recursos involucra-

dos en el proceso tuvieran cubiertos sus datasets con al menos un 20% de instancias. Los recursos evaluados fueron Europeana, VIAF, DLBP y el archivo de la UCLV, seleccionando los elementos que caracterizan un documento que tienen buena representación en los datasets:

$$H(p) = - \sum_{i=1}^n p_i \ln(p_i)$$

La evaluación de la entropía sirve para seleccionar los namespaces con mejores valores de C, H y Hnorm, de acuerdo con las características de un documento (tabla III). Los namespaces evaluados fueron:

<http://xmlns.com/foaf/0.1/>
<http://xmlns.com/foaf/0.1/>
<http://www.europeana.eu/schemas/ese/>
<http://purl.org/dc/elements/1.1/>
<http://viaf.org/viaf/data/>

Tabla III. Cobertura, entropía y entropía normalizada para los datasets de este experimento con un 20% de cobertura

Nombre	Cobertura	H	Hnom
skos:Concept	0,999	0,8345	0,999
skos:prefLabel	0,876	0,8523	0,876
skos:altLabel	<u>0,981</u>	<u>0,8698</u>	<u>0,981</u>
skos:broader	0,976	0,8765	0,976
skos:narrower	0,956	0,8773	0,956
skos:related	0,888	0,3458	0,888
skos:broadMatch	0,947	0,3456	0,947
skos:narrowMatch	0,876	0,8765	0,876
skos:relatedMatch	0,999	0,7772	0,999
skos:exactMatch	0,976	0,7834	0,976
skos:closeMatch	0,978	0,6783	0,978
skos:note	0,999	0,9865	0,999
skos:notation	0,913	0,4531	0,913
skos:inScheme	0,999	0,7803	0,999
rdf:ResourceOrLiteralType	0,991	0,2456	0,991
rdf:LiteralType	0,998	0,7735	0,998
edm:currentLocation	0,965	0,8456	0,965
edm:hasMet	0,953	0,3903	0,953
edm:hasType	0,975	0,7931	0,975
edm:incorporates	0,876	0,563	0,876
edm:isDerivativeOf	0,981	0,455	0,981
edm:isNextInSequence	0,976	0,7653	0,976
edm:isRelatedTo	0,68	0,3457	0,68
edm:isRepresentationOf	0,45	0,2345	0,45
edm:isSimilarTo	0,67	0,3455	0,67
edm:isSuccessorOf	0,98	0,3454	0,98
edm:realices	0,76	0,3467	0,76
edm:EdmType	0,34	0,45678	0,34
edm:ProvidedCHOType	0,67	0,6789	0,67
edm:WebResourceType	0,7	0,5678	0,7
edm:AgentType	0,85	0,345	0,85
edm:PlaceType	0,45	0,567	0,45
edm:TimeSpanType	0,67	0,23	0,67
edm:aggregatedCHO	0,45	0,345	0,45
edm:collectionName	0,78	0,64	0,78
edm:country	0,89	0,71	0,89
edm:hasView	0,678	0,57	0,678
edm:isShownBy	0,89	0,84	0,89
edm:preview	0,78	0,63	0,78
edm:landingPage	0,56	0,4123	0,56

Nombre	Cobertura	H	Hnom
edm:language	0,45	0,33	0,45
edm:rights	0,99	0,93	0,99
owl:same as	0,99	0,93	0,99
dc:contributor	0,99	0,93	0,99
dc:description	0,76	0,7	0,76
dc:title	0,99	0,95	0,99
dc:creator	0,99	0,92	0,99
foaf:document	0,99	0,89	0,99
foaf:mane	0,96	0,91	0,96
owl:differentFrom	0,99	0,9123	0,99
uclvf:title	0,99	0,85	0,99
uclvf:topic	0,96	0,89	0,99
uclvf:type	0,86	0,84	0,86

El resultado de la tabla III muestra que los mejores resultados en el proceso de evaluación de entropía, H y Hn son superiores en los datasets de Europeana y en los de VIAF, debido a la calidad de las descripciones de los datos. Los datasets locales, etiquetados con la sigla uclvf, mantienen buenos índices de evaluación, todos por encima de 0,5. Se decidió usar en el test las características *uclvf:type*, *skos:Concept*, *edm:collectionName*, *foaf:document*, *foaf:mane*, *uclvf:title*, *edm:EdmType*, que tienen mayor representatividad en cada uno de los sistemas de información seleccionados, y que representan lo esencial de un documento.

A continuación se llevó a cabo una simulación, poniendo en funcionamiento el software durante 60 segundos para acceder al 30% de las instancias, como en trabajos que pretendían evaluaciones similares (Sing-Borrajo, 2013), para así monitorizar los cambios en los datasets con una tasa media de 7,03 eventos, resultado del cociente 422/60 seg. Se trataba de analizar la influencia de los eventos de la plataforma en su exhaustividad y precisión. Para ello se repitió la prueba con diferentes intervalos (4 s., 8 s., 15 s., 25 s., 45 s.), que registraron un promedio de 12, 21, 34, 43 y 379 por cada segundo de ejecución del sistema.

Evaluación de la exhaustividad y la precisión

La exhaustividad registra valores muy favorables para el dataset de CoroArchivo, denominado *uclvf:title* y *uclvf:type*. Los porcentajes de estos valores se mantienen por encima del 80%. A medida que aumentan los eventos tienden a mejorar los valores de exhaustividad. Los niveles inferiores de exhaustividad solo aparecen en algunos eventos cuando la plataforma comienza a trabajar. Los

elementos de Europeana y VIAF muestran resultados favorables de exhaustividad a medida que se van produciendo eventos. Los tipos de datos que mayor exhaustividad alcanzan a medida que aumentan los eventos son los de DLBP (Figura 5).

Precisión

La precisión también crece a medida que aumentan los eventos. Las etiquetas de Europeana y uclvf registran el mayor nivel de precisión. Pero, si bien es cierto que tiende a acrecentarse a medida que aumentan los eventos, en la mayoría de los namespaces analizados se mantiene por debajo del 40%. Aparecen mayores índices de precisión en los *foaf:name* y *foaf:document* pertenecientes a DLBP, a *uclvf title* de la base CoroArchivo y a *edm:Edmtype*: de Europeana (Figura 6).

F-measure

La figura 7 muestra cómo disminuyen de manera general los valores de F-measure a medida que aumenta el tiempo de trabajo de la plataforma CoroArchivo. Esto no obliga a usar un mayor número de vectores de rasgos para lograr mayores niveles de exhaustividad y precisión, pues ha de tenerse en cuenta que la herramienta procesó datasets de gran tamaño y con altos niveles de información, más de 50 gigas de datos. Por lo tanto los resultados de F-measure parecen aceptables. Es importante destacar que, entre el octavo y el décimo evento, los valores de exhaustividad y precisión (F-measure) alcanzaron resultados muy cercanos a 100%, ya que el servidor tiende a recargarse en los primeros minutos. Con el paso del tiempo mejora su capacidad.

Figura 5. Influencia del número de eventos del sistema CoroArchivo en la exhaustividad

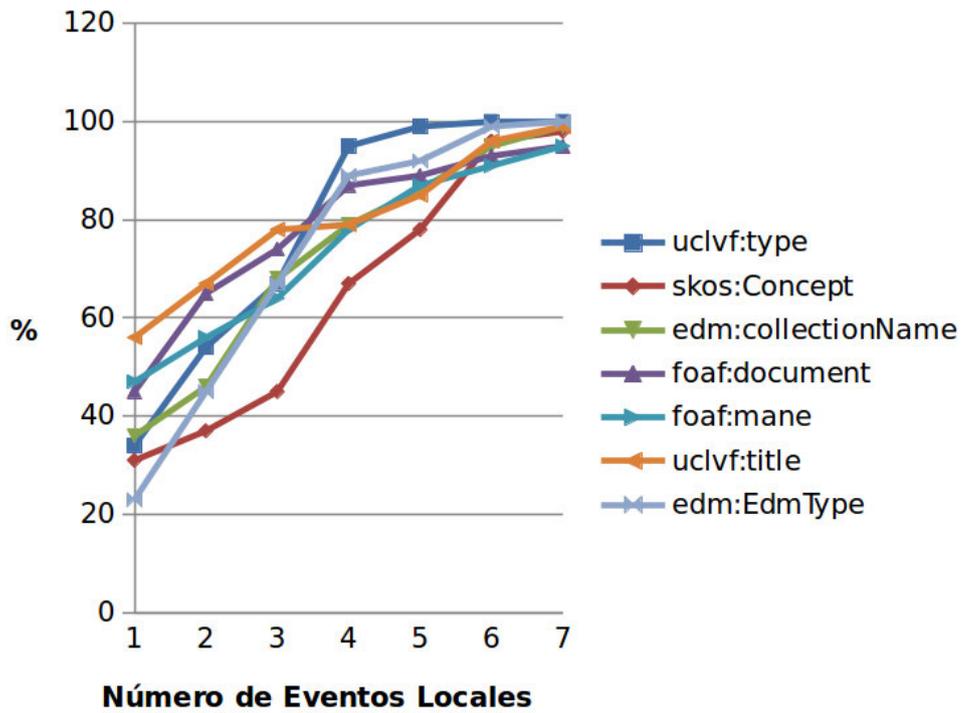


Figura 6. Influencia de los eventos locales de la plataforma en la precisión del sistema

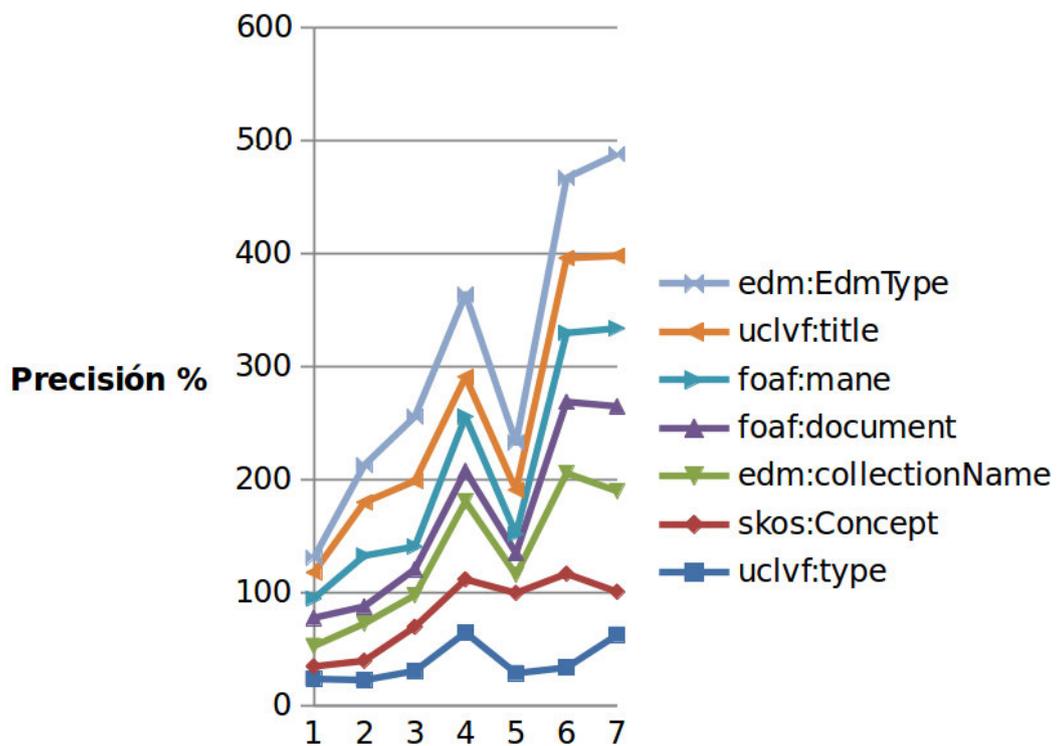
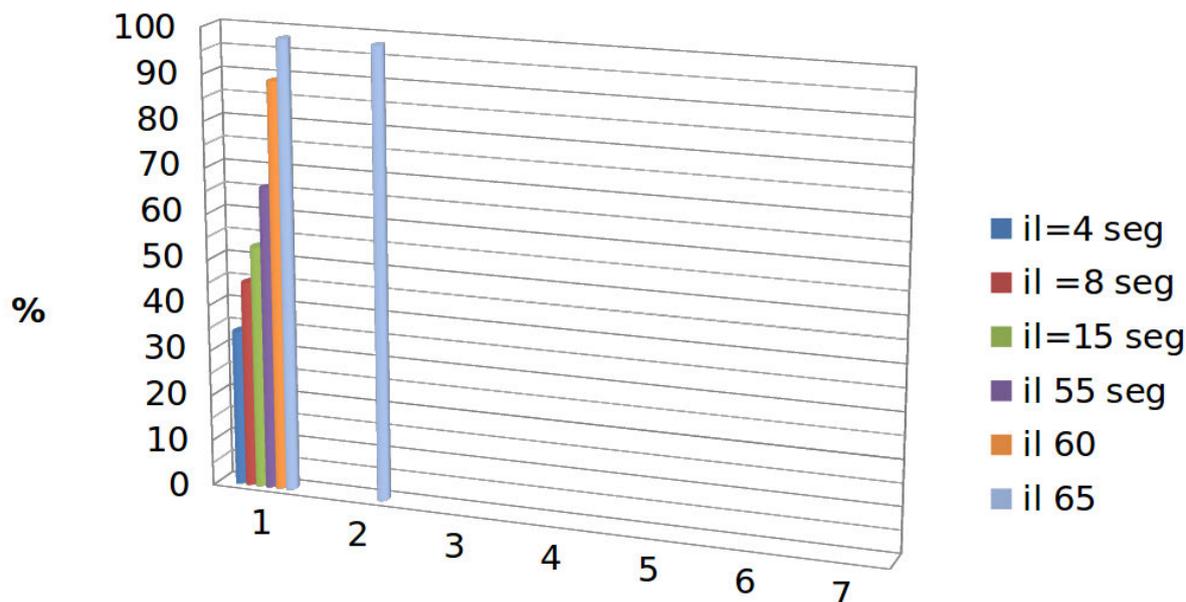


Figura 7. Influencia de los eventos locales de la plataforma en el parámetro F-measure con variables de tiempo

Rendimiento en la búsqueda federada

Para evaluar la búsqueda federada usamos el SPARQL Endpoint interno de los fondos de archivos de la UCLV, dataset con 30.000 registros de archivo publicados simultáneamente en un SPARQL Endpoint local y en otro remoto. En el local solo se almacenan las URI de los recursos, mientras que en el remoto se incluyen recursos de información con propiedades ontológicas específicas para igualar y excluir elementos. La búsqueda en el SPARQL Endpoint local y en el remoto facilita la consulta de información federada y la evaluación de OWL2tips. Se analizó la consulta distribuida de excelencia, realizada a través de la plataforma CoroArchivo y su SPARQL Endpoint, ponderando la eficiencia de la herramienta en función del tiempo y el número de consultas generadas con el algoritmo. Otra parte del proceso de validación de la herramienta realizó el mismo procedimiento en otros datasets (Europeana, DLBP y VIAF) descargados en el servidor de la UCLV, con 68.000 registros, a los que se le adjudicaron constructores y modelos de etiquetado similares a los de CoroArchivo en OWL (tabla IV).

La evaluación de la efectividad del sistema de búsqueda federada a través de las capacidades de su algoritmo demuestra que la herramienta logra valores muy positivos cuando se realizan las búsquedas sobre los recursos internos de la UCLV. Los primeros resultados de las consultas con el algoritmo global aportan más del 30% de los datos alojados en los recursos semánticos, tanto durante

los primeros momentos de la recuperación de la información como en el final del proceso, momento en el que aumenta el nivel de exhaustividad hasta el 97% de los recursos con instancias. La valoración de la búsqueda distribuida perfecta también mostró resultados favorables durante este proceso. Cuando se realizó el mismo experimento con los datos de otras fuentes como Europeana, DLBP y VIAF, descargados en los servidores de la UCLV, los resultados fueron similares, y presentan mayores niveles de recuperación de información. Sin embargo aún existen deficiencias en la descripción, que ha dejado muchos recursos de archivo de valor bibliográfico mal descritos y, en consecuencia, no recuperados (tabla V).

5. CONCLUSIONES

Tras analizar las principales tecnologías aportadas por la Web Semántica, algunos de los proyectos que las están implementando para la gestión de fondos de archivos, así como la plataforma que se ha desarrollado específicamente para el presente trabajo, queda manifiesto, a nuestro juicio, que resulta beneficioso utilizar dichas tecnologías para gestionar y recuperar informaciones de archivo con datos enlazados.

La puesta en funcionamiento de este tipo de sistemas puede hacer uso de diversas herramientas ya existentes, como Silk, Virtuoso, Pubby, Authoris y Metharto, que facilitan la generación de modelos de datos enlazados de alto nivel semántico.

Tabla IV. Efectividad de la búsqueda federada en el dataset local

Algoritmo de búsqueda perfecta	Primeros resultados de tiempo	Últimos resultados de tiempo
Búsqueda 1	1023 ms	13456 ms
Búsqueda 8	4213 ms	14233 ms
Búsqueda 10	6783 ms	16234 ms
Uso global del algoritmo	Primeros resultados de Tiempo	Últimos resultados de tiempo
Búsqueda 1	12034 ms (+41%)	22116 ms (+73%)
Búsqueda 8	19076 ms (+63%)	29123 ms (+97%)
Búsqueda 10	10023 ms (+33,41%)	29113 ms (+97%)

Tabla V. Efectividad de la búsqueda federada en los datasets de Europeana, DLBP y Library of Congress

Algoritmo de búsqueda perfecta distribuida	Primeros resultados de tiempo	Últimos resultados de tiempo
Búsqueda 1	150023 ms	205677 ms
Búsqueda 8	345783 ms	345680 ms
Búsqueda 10	387643 ms	416234 ms
Uso global del algoritmo	Primeros resultados de tiempo	Últimos resultados de tiempo
Búsqueda 1	386650 ms (56%)	560770 ms (82%)
Búsqueda 8	493050 ms (72%)	602304 ms (88%)
Búsqueda 10	582340 ms (85%)	675143 ms (99%)

El modelo propuesto en este artículo se basa en las prestaciones ofrecidas por las propiedades de los objetos para formular búsquedas federadas, haciendo posibles nuevas opciones para la consulta de conjuntos de datos.

Consideramos, por otra parte, que la experiencia de evaluación incluida en estas páginas demuestra suficientemente la eficiencia del modelo, pues la plataforma logra altos niveles de F-measure, en la mayoría de los casos por encima del 60%.

6. REFERENCIAS

- Addis, M.; Allasia, W.; Bailer, W. (2010). 100 Million Hours of Audiovisual Content: Digital Preservation and Access in the PrestoPRIME Project. *First International Digital Preservation Interoperability Framework (DPIF) Symposium*. Dresden, Germany: ACM.
- Alonso-Sierra, L. E., Ortiz-Muñoz, E.; Hidalgo-Delgado, Y. (2012) Los sistemas de gestión de contenidos en el ámbito de la web semántica: una breve revisión. *Serie científica de la Universidad de las Ciencias*, 5, 1-9.
- Baker, T. (2012). Libraries, languages of description, and linked data: a Dublin Core perspective. *Library Hi Tech*, 30, pp. 116 - 133. <http://dx.doi.org/10.1108/07378831211213256>
- Berners-Lee, T.; Hendler, J. y Ora Lassila (2001). The Semantic Web. *Scientific American*, vol. 284 (5), 28-37. <http://dx.doi.org/10.1038/scientificamerican0501-34>
- Bizer, C.; Seaborne, A. (2004). D2RQ - Treating Non-RDF Databases as Virtual RDF Graphs. *3rd International Semantic Web Conference*. Hiroshima, Japan: Springer.
- Brickley, D.; Miller, L. (2010). *FOAF Vocabulary Specification*.
- Byron, A., Berry, A.; De-Bondt, B. (2012). *Using Drupal: choosing and configuring modules to build dynamic websites*. Sebastopol (California), O'Reilly Media.
- Coppens, S.; Mannes, E.; Walle, R. V. D. (2009). Disseminating heritage records as linked open data. *International Journal of Virtual Reality*, 8, 39-44.
- Coppens, S.; Verborgh, R.; Mannes, E.; Walle, R. V. D. (2013). Querying the Linked Data Graph using owl:sameAs Provenance. *Proceedings of the 16th International Conference on Model Driven Engineering Languages and Systems*.
- Cygniak, R.; Bizer, C. (2008). Pubby - A Linked Data Frontend for SPARQL Endpoints. <http://wifo5-03.informatik.uni-mannheim.de/pubby>
- Doerr, M.; Gradmann, S.; Henniecke, S.; Isaac, A.; Meghini, C.; Van De Sompel, H. (2011). The Europeana data model (EDM). *IFLA 2011: World library and information congress: 76th IFLA general conference and assembly*. Gothenburg, Suecia: IFLA.
- Domínguez-Velasco, S. (2013a). *OWL2tips: herramienta para generar y visualizar ontologías desde Drupal*. Beta 2 ed. Santa Clara, Cuba: Universidad Central de las Villas.

- Domínguez-Velasco, S. (2013b). *Sampras*. Alfa 2 ed. Santa Clara, Cuba: Universidad Central de las Villas.
- Domínguez-Velasco, S.; Leiva-Mederos, A. (2013). *Jump*. Beta ed. Santa Clara, Cuba: Universidad Central de las Villas.
- Ferro, N.; Silvello, G. (2013). NESTOR: A formal model for digital archives. *Information Processing and Management*, vol. 49 (6), 1206-1240. <http://dx.doi.org/10.1016/j.ipm.2013.05.001>
- Giasson, F.; D'arcus, B. (2009). *Bibliographic Ontology Specification*. Structured Dynamics LLC. <http://bibliontology.com/specification>
- Görlitz, O.; Staab, S. (2011). SPLENDID: SPARQL Endpoint Federation Exploiting VOID Descriptions. *The 10th International Semantic Web Conference*. http://iswc2011.semanticweb.org/fileadmin/iswc/Papers/Workshops/COLD/GoerlitzAndStaab_COLD2011.pdf
- Gracy, Karen F. (2014). *Archival description and linked data: a preliminary study of opportunities and implementation challenges*. *Arch Sci*. Disponible en <http://doi10.1007/s10502-014-9216-2>
- Grimoüard, Claire Sibille-de (2014). The Thesaurus for French Local Archives and the Semantic Web. *Procedia - Social and Behavioral Sciences*, (147), 206-212. <http://dx.doi.org/10.1016/j.sbspro.2014.07.153>
- Hartig, O.; Bizer, C.; Freytag, J. D. (2009). Executing SPARQL Queries over the Web of Linked Data. *ISWC '09 Proceedings of the 8th International Semantic Web Conference*. Springer-Verlag. http://dx.doi.org/10.1007/978-3-642-04930-9_19
- Haslhofer, B.; Roochi, E. M.; Schandl, B.; Zander, S. (2011). *Europeana RDF Store Report*. Viena, Austria: Universidad de Viena.
- Haslhofer, B.; Schnadl, B. (2008). The OAI2LOD Server: Exposing OAI-PMH Metadata as Linked Data. In: *International Workshop on Linked Data on the Web (LDOW2008)*. Beijing, China.
- Haslhofer, B.; Schnadl, B. (2010). Interweaving OAI-PMH data sources with the linked data cloud. *International Journal of Metadata, Semantics and Ontologies*, vol. 5 (1), 17-31. <http://dx.doi.org/10.1504/IJMSO.2010.032648>
- Hernández, A. (2007). *Organización y representación del conocimiento: paradigmas, hipertextos y fundamentación metamodélica*. Universidad de La Habana. Cuba.
- Hidalgo-Delgado, Y.; Rodríguez-Puente, R.; Ortiz-Muñoz, E.; Alonso-Sierra, L. E. (2013). Herramienta para la recolección de metadatos bibliográficos mediante el protocolo OAI-PMH. *II Conferencia Internacional de Ciencias Computacionales e Informáticas*. La Habana, Cuba.
- Jentzsch, A.; Isele, R.; Bizer, C. (2010). Silk - Generating RDF Links while publishing or consuming Linked Data. *International Semantic Web Conference Posters&Demos*, Shanghai, China.
- Koutsomitropoulos, D. A.; Solomou, G. D.; Theodore S. Papatheodorou (2008). Semantic Interoperability of Dublin Core Metadata in Digital Repositories. *Innovations in Information Technology, 2008. IIT 2008. International Conference on IEEE*. <http://dx.doi.org/10.1109/INNOVATIONS.2008.4781709>
- Ledesma, G. A. (2013). *Coroimagen. Ontología para el tratamiento de archivo*. Santa Clara, Cuba: Universidad Central de las Villas.
- Leiva-Mederos, A.; Senso, J. A.; Domínguez-Velasco, S.; Hípola, P. (2013). Authoris: a tool for authority control in the Semantic Web. *Library Hi Tech*, vol. 31 (3), 536-553. <http://dx.doi.org/10.1108/LHT-12-2011-0135>
- Lynch, Tom J. (2014). Social Networks and Archival Context Project: A Case Study of Emerging Cyberinfrastructure. *Digital Humanities Quarterly*, 8(3). Disponible en: <http://www.digitalhumanities.org/dhq/vol/8/3/000184/000184.html> [Consulta: 1 de septiembre de 2015].
- Mena, M. (2006). *Retos de la actividad archivística: reporte de conferencias*. La Habana, Cuba: Universidad de la Habana.
- Merlino-Santesteban, C. (2012). Repositorios institucionales y buscadores web: una interrelación no tan exitosa. *10^a Jornada sobre la biblioteca digital universitaria*. Buenos Aires. Argentina.
- Moyano Collado, Julián. (2013). La Descripción Archivística. de los Instrumentos de Descripción Hacia la Web Semántica. *Anales de Documentación*, 16(2), 3-13.
- Nam-Park, Ok. (2015). Development of Linked Data for Archives in Korea. *DLib Magazine*, 21(3-4), 1-13.
- Nolle, A.; Nemirovski, G. (2013). ELITE: An Entailment-based Federated Query Engine for Complete and Transparent Semantic Data Integration. In Eiter, T.; Glimm, B.; Kazakov, Y.; Krötzsch, M. (eds.) (2013). *26th International Workshop on Description Logics*. Ul, Alemania.
- Peroni, S.; Shotton, D. (2012). FaBiO and CiTO: ontologies for describing bibliographic resources and citations. *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 17, 33-43. <http://dx.doi.org/10.1016/j.websem.2012.08.001>
- Popitsch, N.; Haslhofer, B. (2011). DSNotify - A solution for event detection and link maintenance in dynamic datasets. *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 9 (3), 266-283. <http://dx.doi.org/10.1016/j.websem.2011.05.002>
- Quillitz, B.; Leser, U. (2008). Querying Distributed RDF Data Sources with SPARQ. *ESWC'08 Proceedings of*

- the 5th European semantic web conference on The semantic web: research and applications*. Springer-Verlag.
- Rademaker, A.; Borges-Oliveira, D. A.; de Paiva, V.; Higuchi, S.; Medeiros e Sá, A.; Alvim, M. (2015). A linked open data architecture for the historical archives of the Getulio Vargas Foundation. *Int J Digit Libr*, (15), 153-167. <http://dx.doi.org/10.1007/s00799-015-0147-1>
- Sánchez Alonso, S.; Sicilia Urbán, M. Á.; Rato Leguina, G. D. (2008). Sobre la interoperabilidad semántica en las descripciones archivísticas digitales. *Revista Española de Documentación Científica*, vol. 31 (1), 11-34.
- Schöpfel, J.; Bescond, I.; Prost, H. (2012). Open is not enough: a case study on grey literature in an OAI environment. *The Grey Journal*, vol. 8 (2), pp. 112-124.
- Schwarte, A.; Haase, P.; Hose, K.; Schenkel, R.; Schmidt, M. (2011). FedX: A Federation Layer for Distributed Query Processing on Linked Open Data. *The Semantic Web: Research and Applications*, 481-486. http://dx.doi.org/10.1007/978-3-642-21064-8_39
- Sing-Borrajó, P. (2013). *Reporte de cargas en los datasets de la biblioteca universitaria de la Universidad Central de las Villas*. Santa Clara, Cuba: Universidad Central "Marta Abreu" de las Villas, Facultad de Ingeniería Eléctrica, Departamento de Telecomunicaciones.
- Solomou, G.; Papatheodorou, T. (2010). The Use of SKOS Vocabularies in Digital Repositories: the DSpace case. *Fourth International Conference on Semantic Computing*. IEEE. <http://dx.doi.org/10.1109/icsc.2010.83>
- Tummarello, G.; Delbru, R.; Oren, E. (2007). *Sindice.com: Weaving the open linked data*. 6th international The semantic web and 2nd Asian conference on Asian semantic web conference.
- Vasallo, S. (2010). Descrizioni Archivistiche e web semántico: un connubio possibile. *Italian Journal of Library and Information Science*, 1(1), 169 - 163.

ESTUDIOS / RESEARCH STUDIES

Análisis bibliométrico de las tesis doctorales españolas indexadas con el descriptor "Sector de la educación" (1976/2014)

Olga Moreno-Fernández*, Pilar Moreno-Crespo**

*Facultad de Ciencias de la Educación. Universidad de Sevilla

**Facultad de Formación del Profesorado. Universidad de Extremadura

Correos-e: omoreno@us.es, pilarmoreno@unex.es

Recibido: 03-09-2015; 2ª versión: 08-10-2015; Aceptado: 08-12-2015.

Cómo citar este artículo/Citation: Moreno-Fernández, O.; Moreno-Crespo, P. (2016). Análisis bibliométrico de las tesis doctorales españolas indexadas con el descriptor "Sector de la educación" (1976/2014). *Revista Española de Documentación Científica*, 39(3): e146. doi: <http://dx.doi.org/10.3989/redc.2016.3.1331>

Resumen: El presente trabajo es un análisis de la estructura productiva de la investigación española catalogada con el descriptor "sector de la educación" a través de los datos extraídos de las tesis doctorales en España en el período 1976/2014. Para ello se ha aplicado el análisis bibliométrico, utilizando la base de datos bibliográfica TESEO para localizar las tesis doctorales. Los resultados describen la producción según quinquenios, universidades, directores, evaluadores, así como el idioma de los trabajos, el sexo de los doctorandos y la mención europea/internacional del título. El número total de tesis producidas en el período estudiado es de 485 títulos. La tendencia general indica un incremento constante, con un crecimiento exponencial en los últimos 15 años, con una alta presencia de trabajos relacionados con la Economía de la Educación en la primera mitad del período estudiado que pasa a inclinarse posteriormente por los trabajos en Ciencias de la Educación.

Palabras clave: Indexación; bibliometría; tesis doctorales; sector de la educación; educación superior.

Bibliometric analysis of Spanish doctoral thesis indexed with the descriptor "Education sector" (1976/2014)

Abstract: This study is an analysis of Spanish research output cataloged with the descriptor «Education Sector», through data extracted from doctoral dissertations in Spain during the period 1976/2014. For this we have applied a bibliometric analysis using the TESEO bibliographic database to locate doctoral theses. The results describe the output by quinquennium, university, director, evaluator and the language of work, sex of doctoral candidate and either a European or international mention in the title. The total number of theses produced in the period under review is 485. The overall trend shows a steady increase, with an exponential growth in the last 15 years, with a high proportion of works related to the Economics of Education in the first half of the period studied, subsequently leaning more toward work on Education Sciences.

Keywords: Indexing; bibliometrics; dissertations; education sector; higher education.

Copyright: © 2016 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Attribution (CC BY) España 3.0.

1. INTRODUCCIÓN

Los cambios que en los últimos años está experimentando el sector de la educación presentan a la Universidad la necesidad de aumentar la calidad de las investigaciones que se realizan, así como la visibilidad y transferencia de los conocimientos obtenidos en estas investigaciones (Bermúdez y otros, 2009; Carbonell y Calvó, 2009; Conferencia de Rectores de Universidades Españolas, 2013). Hecho que ha repercutido en que en los últimos años se hayan publicado un gran número de estudios que analizan esta productividad en las universidades desde distintos enfoques (Agudelo y otros, 2003; Bermúdez y otros, 2011; Jiménez-Conteras y otros, 2014; Olivas-Ávila y otros, 2012; Oscalluch y otros, 2013). Estudios que abordan diversidad de criterios como el número de tesis doctorales dirigidas (Musi-Lechuga y otros, 2011; Olivas-Ávila y Musi-Lechuga, 2010); la relación que existe entre la dirección de tesis doctorales y la publicación de artículos o el tipo de citaciones que reciben los profesores e investigadores (Ariza y otros, 2012; Buena-Casal, 2005, 2010; Buena-Casal y Zych, 2010; Olivas-Ávila y otros, 2012; Robinson y otros, 2011; Schultz, 2004; Villar, 2011).

Este interés por la producción científica ha dado lugar al fortalecimiento de los estudios bibliométricos, los cuales aportan datos valiosos para verificar la situación actual de diversas áreas de conocimiento, siendo el análisis bibliométrico de las tesis doctorales todo un campo de estudio (Casaneva y otros, 2007; Castillo y Xifra, 2006; Delgado y otros, 2006; Fuentes y Argimbau, 2010; Haba y otros, 2000; Marcos y otros, 2012; Ortega, 2010; Repiso y otros, 2011a, 2011b; Torres y Torres, 2005, 2007). Y es que, aunque son diversas las vías por las que el conocimiento científico se abre paso y deja su huella (tesis doctorales, artículos científicos, libros, o actas de congresos entre otros), la tesis doctoral sigue siendo el trabajo de investigación que inicia al estudiante de doctorado al campo de la investigación, con lo que ello conlleva. Un trabajo inédito que aporta nuevos conocimientos al campo científico y que debe ser avalado y evaluado por autoridades competentes en la materia de estudio, directores y tribunales cualificados para esta labor. Por consiguiente, estamos de acuerdo con Delgado y otros (2006) con que "son un excelente referente para conocer la estructura social de la investigación en la universidad", razón por la que se han constituido en sí mismas en un objeto de investigación, estableciéndose como uno de los mejores ejemplos donde se reflejan las líneas, tendencias y potencialidades de la investigación de las universidades.

En España existen numerosos estudios aplicados al análisis de estos trabajos de investigación. Casi todos los elaborados hasta 2009 se recogen en el estudio de Fuentes y Argimbau (2010), en el que se distinguen hasta dieciocho campos científicos distintos que han sido estudiados y caracterizados a través de las tesis doctorales defendidas en los mismos. Por similitud temática o metodológica cabría señalar también los estudios sobre tesis doctorales defendidas en "Bibliometría" (Delgado y otros, 2006), "Estudios de la Mujer" (Torres y Torres, 2005, 2007), "Relaciones Públicas" (Castillo y Xifra, 2006), "Televisión" (Repiso y otros, 2011a), "Radio" (Repiso y otros, 2011b), "Fotografía" (Sánchez y otros, 2014), o "Internet" (Díaz del Campo, 2014), entre otros.

En el campo educativo también está aumentando paulatinamente este tipo de estudios, entre los que podemos destacar los centrados en "Educación" en general (Fernández y Bueno, 1998), "Educación matemática" (Max y otros, 2012), "el Dibujo en los libros de texto" (Collados, 2010), sobre "Historia de la Educación" (Galante y Sanchidrián, 1996), o acerca de revistas educativas (Ariza y Quevedo-Blasco, 2013; Carpallo y Burgos, 2008; Gómez-García y otros, 2012; Granados y otros, 2011; López-Meneses y otros, 2015), siendo estos últimos los más generalizados.

En el ámbito de la investigación dedicada a la educación en España aún son escasos los trabajos descriptivos que atienden al análisis en base a los descriptores que se utilizan para la identificación de los trabajos (Gutiérrez-Saldivia, 2014; Pérez-Sánchez y otros, 2009). Por lo cual, podemos afirmar que la investigación en el campo de la educación es relativamente reciente y todo lo que le hace referencia implica factores que llegan a ser aparentemente ajenos al propio sector educativo. Sin embargo, y a pesar de estas cuestiones, consideramos que este estudio, junto a los aspectos metodológicos que ya hemos comentado, aporta una visión sobre las investigaciones educativas novedosa y nos brinda un hilo conductor de lo que ha sido la contribución científica realizada a través de tesis doctorales en los últimos lustros y cómo ésta se ha ido fortaleciendo.

Este trabajo pretende describir la producción científica catalogada bajo el descriptor "sector de la educación" recogida en las tesis doctorales inscritas entre 1976 y 2014 en la base de datos TESEO, mediante un estudio pormenorizado, con el objetivo de cuantificar y trazar cronológicamente la producción de tesis doctorales por sexo, idioma, universidades, directores y participación en tribunales, e identificar a los principales actores que han protagonizado la investigación científica en lo

que se ha denominado "sector de la educación" y las estructuras que generan, ofreciendo una panorámica de los cambios, así como del desarrollo que se está produciendo en los trabajos catalogados bajo este descriptor, ayudándonos a entender la evolución que se ha producido en los últimos años.

Consideramos, por tanto, que el interés de este trabajo se centra en la información que la evaluación de la investigación sobre el sector de la educación en España nos va a ofrecer, pues aporta información relevante para diferentes colectivos (profesionales de distintos ámbitos, académicos, investigadores, futuros doctores e instituciones interesadas en el área), ya que se presenta información de carácter científico actualizada sobre la producción científica de este campo disciplinar.

2. MATERIAL Y MÉTODOS

Esta investigación pretende caracterizar la producción española de tesis doctorales que se han adscrito al descriptor "Sector de la educación" del Tesauro, con código de normalización UNESCO [531204], en el período que va desde septiembre de 1976 hasta diciembre de 2014. Se ha seleccionado este sistema de código por tratarse de un criterio normalizado, consistente en una nomenclatura internacional para los campos de ciencia y tecnología, utilizada en España desde 1973.

Por tanto, este trabajo se plantea el objetivo de cuantificar y trazar cronológicamente la producción de tesis doctorales por sexo del doctorado, idioma, universidades, directores y participación en tribunales e identificar a los principales actores que han protagonizado la investigación científica en lo que se ha denominado "sector de la educación" y las estructuras que generan. En cuanto a la elección del descriptor el Tesauro de la UNESCO clasifica para educación los siguientes descriptores: "sector de la educación", "historia de la educación", "organización y planificación de la educación", "educación de adultos", "planificación y financiación de la educación", "educación básica", "educación especial", "educación superior", y "sociología de la educación". Nos interesan las tesis doctorales que se han considerado deben etiquetarse con dicho código, ya que implica un posicionamiento por parte del investigador, descartando aquellas en las que no ha sido considerado. Debemos indicar que las tesis doctorales analizadas, aun ubicándose en diversos campos científicos, todas se encuentran relacionadas con el sector educativo. No obstante, nuestro objeto de estudio dista de evaluar la pertinencia del código de normalización UNESCO. Nos decidimos, por lo tanto, por "Sector de la educación" al ser más general, considerando que en

él cabían los restantes descriptores. El número de trabajos registrados fue uno de los factores determinantes para hacer el análisis de este único descriptor, dejando a futuras investigaciones el cruce con otros descriptores del mismo campo.

Para ello se realiza un estudio descriptivo longitudinal retrospectivo de las tesis doctorales que utilizan el descriptor "Sector de la educación" en España en el período que va desde 1976 a 2014. Por lo tanto, esta elección del descriptor se ha realizado para dar cabida a la producción científica realizada desde la elaboración de tesis doctorales que se han catalogado como relevantes, referentes y/o relativas al campo educativo que, tal y como se ha indicado en párrafos anteriores, se encuentra en crecimiento y diversificación. Nuestro interés se centra, por tanto, en elaborar una hoja de ruta de estas investigaciones.

La metodología empleada puede ser resumida en 4 fases de investigación: documentación, obtención y organización de la información, análisis de la información, y elaboración y difusión del informe final.

Para la recuperación de la información se ha empleado la base de datos TESEO, dependiente del Ministerio de Educación y cuya finalidad es recopilar los registros de las tesis leídas en las universidades españolas. Entre los aspectos que nos han llevado a seleccionar esta base de datos sobre otras es su amplio recorrido cronológico (a partir de 1976) y la exhaustiva descripción que realiza de sus registros bibliográficos, incluyendo datos de gran interés, como la composición de los tribunales, siendo fuente recurrente en los estudios españoles sobre tesis doctorales (Fernández-Cano y otros, 2008).

Con respecto a la base de datos TESEO, y como ya se ha apuntado en otros trabajos, presenta algunas limitaciones relacionadas con la normalización de sus registros (Delgado y otros, 2006; Fuentes y Arguimbau, 2010; Repiso y otros, 2011a, 2011b; Martínez Pestaña, 2004, 2011a, 2011b). En relación con su cobertura hay que señalar que no recoge el total de tesis existentes, ya que aquellas que no han sido remitidas por la universidad o el doctorando no figuran en el catálogo. A pesar de esta cuestión, es la única herramienta que recoge, en su conjunto, las tesis de las universidades españolas y prácticamente la única empleada en la realización de estudios cuantitativos de corte bibliométrico (Sorli y Merlo, 2002), lo que nos ha llevado a decidimos por esta base de datos en concreto. No obstante existen otras bases de datos, que sin embargo, presentan un recorrido cronológico menor, menos exhaustivo y con un número de registros inferior. Por ejemplo,

podemos citar la base de datos de Tesis en Red (TDR), que se pone en funcionamiento en el año 2001 por el Consorcio de Servicios Universitarios de Cataluña (CSUC), facilitando el acceso abierto a los textos íntegros de los trabajos producidos en distintas universidades españolas (principalmente universidades catalanas) y permitiendo su recuperación a través de datos como autor, director, título, universidad, departamento, materia de la Clasificación Decimal Universal (CDU), palabras clave, texto libre y año. En este caso los trabajos catalogados bajo el descriptor "Sector de la educación" dieron como resultado la recuperación de 3 trabajos, 2 de los cuáles ya estaban recogidos en la base de datos de TESEO, lo que nos llevó a no considerar esta última base de datos como fuente de recogida de datos de la investigación.

La información recuperada de TESEO recoge 485 trabajos de investigación. Por tanto, esta es la muestra con la que se ha trabajado. Los datos se exportaron a una base de datos ad-hoc de estructura relacional. Como señalan Fuentes y Arguimbau (2010), las dimensiones más comunes en este tipo de análisis son los indicadores temáticos (disciplinas y áreas de estudio) y los de productividad (directores, tribunales, años, universidades y centros de lectura). Otras variables que se estudian en los trabajos examinados son el idioma de lectura o el sexo de los doctorandos. Los registros realizados para la presente investigación recogieron: nombre y apellidos del autor/a, sexo, título de la tesis doctoral, año académico de lectura, universidad en la que se leyó, departamento y/o facultad en la que se en-

contraba inscrito el trabajo, nombre y apellidos del director/a o directores de la tesis y de los miembros que conformaron el tribunal, sexo del director/a o directores y de los miembros del tribunal, idioma de presentación del trabajo y posesión o no de la mención europea/internacional del título.

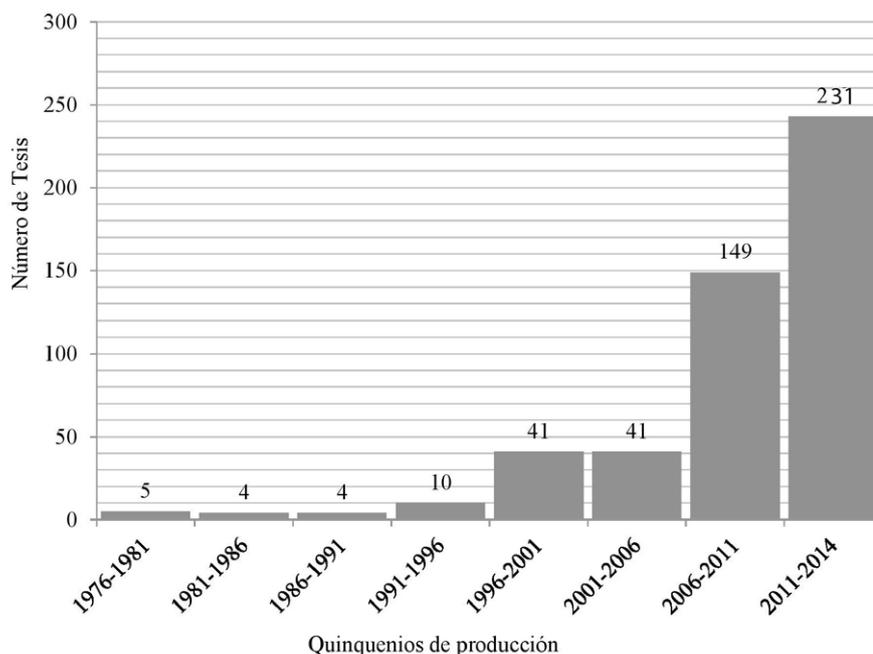
3. RESULTADOS

De los datos analizados, consideramos de gran relevancia poner de manifiesto la evolución que se ha producido en las tesis doctorales defendidas en las universidades españolas indexadas bajo el descriptor "Sector de la educación" y recogidas en la base de datos de TESEO a lo largo del periodo de estudio: entre 1976 y 2014.

Evolución de la producción y Universidades

El número total de tesis doctorales defendidas en España que tienen asignado como descriptor "Sector de la educación" entre 1976/2014 ha sido de 485 trabajos inéditos de investigación. En la figura 1, se muestra la evolución por quinquenios desde el año 1976, constatándose un crecimiento continuo. La producción de tesis que se clasifican dentro de este descriptor comenzó a principios de las década de los 80, aunque no es hasta la primera década del siglo XXI cuando se consolida con una producción significativa de 149 tesis clasificadas dentro del descriptor a estudio. El mayor nivel de producción se alcanza en el último período (2011-2014) con 231 tesis defendidas a fecha de 31 de diciembre de 2014, fecha hasta la que se presentan los resultados.

Figura 1. Evolución por quinquenios académicos del nº de tesis que han utilizado para su indexación el descriptor "Sector de la educación" en España (1976/2014)



En cuanto al campo de estudio al que se encuentran inscritas y teniendo presente el enfoque multidisciplinar que se comentaba con anterioridad, podemos hacer referencia a 23 campos de estudios distintos y variados, abarcando desde las Ciencias Económicas y Empresariales hasta las Ciencias Ambientales, pasando por ramas de conocimiento como la Historia, la Geografía, la Arquitectura, la Sociología, la Informática, la Educación y la Pedagogía, entre otras.

De estos campos de conocimiento, son las Ciencias Económicas y Empresariales, las Ciencias de la Educación y los Estudios Pedagógicos los que recogen el mayor volumen de produc-

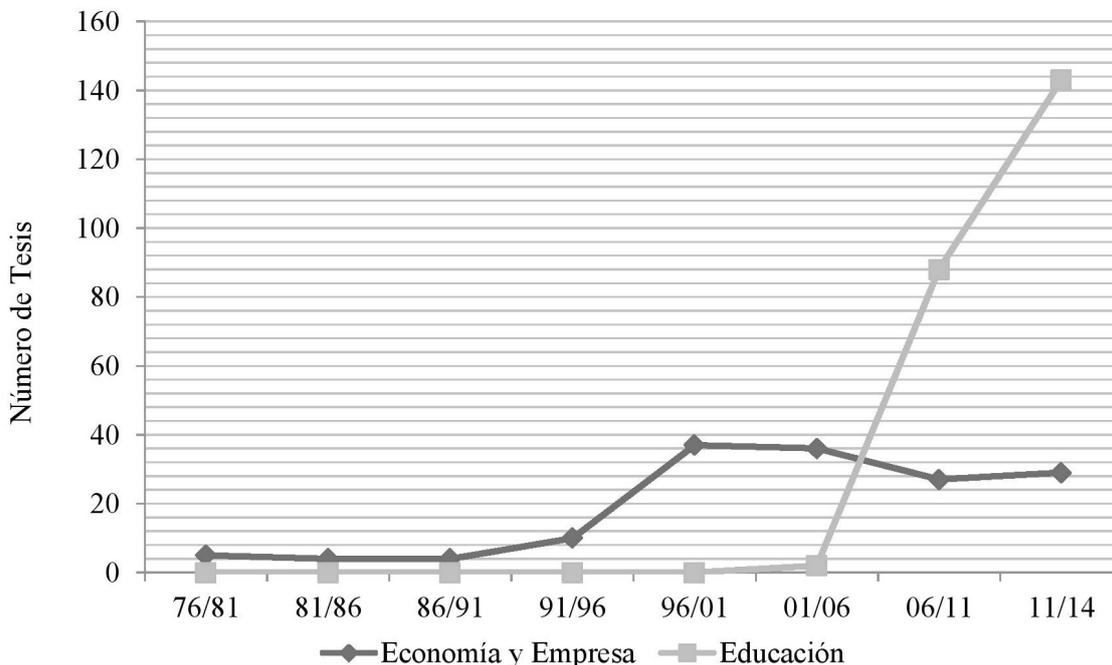
ción, siendo los campos de la Arquitectura, la Informática, las Ciencias Ambientales, el campo de la Gestión, los Sistemas Oceánicos y Navales, la Medicina, la Antropología y la Geografía, los menos productivos bajo el descriptor "Sector de la educación", con tan solo un trabajo cada uno de ellos (véase tabla I).

Si nos detenemos exhaustivamente en los campos más productivos, podemos observar que los trabajos de Ciencias Económicas y Empresariales arrancan con fuerza desde los años 70. Por otro lado, las tesis doctorales de Ciencias de la Educación, inician su auge en el descriptor a estudio a partir del inicio del siglo XXI (véase figura 2).

Tabla I. Nº de tesis doctorales indexadas bajo el descriptor "Sector de la educación" en España (1976/2014) en relación con el año de lectura y el campo de conocimiento

	Economía y Empresa	Historia	Geografía	Arquitectura	Pedagogía	C. de la Educación	Farmacia	Sist. oceánicos y navales	Ingeniería	Gestión	Comunicación y Publicidad	Sociología	Lengua y Literatura	Humanidades	Filología	Psicología	Desarrollo y Cooperación	Filosofía	Informática	Antropología	Enfermería y Nutrición	Medicina	C. Ambientales
76/77	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
77/78	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
78/79	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
79/80	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
80/81	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
81/82	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
82/83	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
83/84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
84/85	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
85/86	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
86/87	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
87/88	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
88/89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
89/90	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
90/91	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
91/92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
92/93	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
93/94	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
94/95	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
95/96	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
96/97	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
97/98	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
98/99	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
99/00	11	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
00/01	10	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
01/02	16	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
02/03	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
03/04	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
04/05	6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
05/06	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
06/07	8	0	0	0	1	2	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
07/08	4	0	0	0	1	4	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
08/09	3	0	0	0	4	12	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0
09/10	3	0	0	0	5	30	0	0	0	0	0	0	1	1	2	0	0	0	0	0	0	0	0
10/11	9	0	0	0	4	40	0	0	0	0	0	1	2	1	3	1	0	0	0	0	0	0	0
11/12	10	0	0	0	4	38	0	0	0	0	1	0	0	0	2	0	1	1	0	0	0	0	0
12/13	9	1	0	0	6	59	0	0	1	0	2	0	0	3	0	9	1	1	0	1	2	0	0
13/14	10	1	1	0	6	47	1	0	3	0	1	1	2	0	4	6	1	0	0	0	1	1	1

Figura 2. Evolución del nº de trabajos indexados bajo el descriptor "Sector de la educación" en Ciencias Económicas y Ciencias de la Educación en España (1976/2014)



Quinquenios productivos. Campos "Economía y Empresa" y "Educación"

Fuente: Elaboración propia.

En cuanto a los centros donde se realizaron las lecturas de las tesis doctorales, se trata de 52 universidades diferentes, de las cuales, las 12 más productivas (entendemos por universidad productiva cuando cuenta con al menos 10 trabajos defendidos), conforman el 56,90% del total (véase la tabla II). Entre éstas, la primera en producción es la Universidad de Barcelona con 66 trabajos leídos (13,61%), seguida por la Universidad de Granada con 44 (9,07%) y la Universidad Autónoma de Barcelona con 28 (5,77%). Respecto a la evolución por quinquenios académicos de las cinco universidades con mayor producción, destaca la Universidad de Barcelona cuyo crecimiento se ha disparado desde el curso 2006/2007 y sigue prosperando. La tabla II recoge aquellas Universidades productivas en las que desde 1976 hasta el 2014 se han leído un número igual o superior a 10 tesis doctorales.

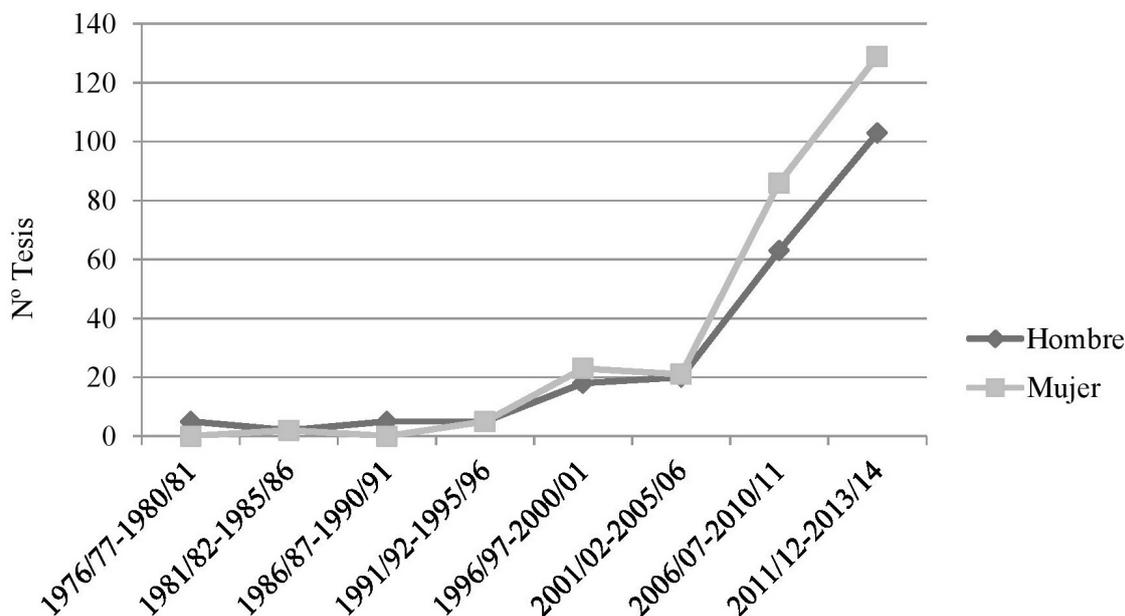
A título meramente informativo, se ha añadido una nube de palabras (Figura 3) generada con el programa Wordle® a partir de los términos que aparecen en los títulos de las 485 tesis analizadas. El tamaño menor o mayor en la imagen de cada una de las palabras está directamente relacionado con su frecuencia de aparición. Como se observa en la figura, el término educación es el protagonista, seguido por análisis y

España, así como también pueden destacarse los conceptos de universidad, aplicación y formación con una fuerte presencia entre los términos más utilizados.

Idioma, sexo de los doctorandos y mención europea/internacional del título

El castellano es el idioma empleado por la mayoría de los autores en las tesis analizadas, ya que 396 de ellas (81,64%) han sido escritas en esa lengua. El resto de lenguas que tienen carácter oficial en España suman otro 8,45% que corresponden a 41 tesis redactadas en catalán, 1 en euskera y 1 en valenciano. Por último, en lengua extranjera encontramos 29 tesis escritas en inglés (5,97%), 16 en portugués (3,29%) y 1 en francés (0,20%).

Por lo que respecta al sexo de los doctorandos, las mujeres representan el 55,05% de las tesis defendidas, con un total de 267 trabajos, mientras que los hombres representan el 44,94% (218 trabajos doctorales). Examinando la distribución anual de autorías por sexo, los hombres lideran la producción hasta el curso académico 2004/05 cuando comienza a invertirse esta situación, siendo las mujeres las que toman el relevo en producción científica, hasta llegar a superar a los hombres en 2005/06, tendencia que se mantiene hasta la actualidad (figura 4).

Figura 4. Evolución de la producción de tesis doctorales por sexo

En cuanto al doctorado europeo o internacional (la base de datos TESEO no diferencia entre uno u otro), el primero que se detecta categorizado con el descriptor "Sector de la educación" es en el año académico 2009/2010, concretamente 2 tesis del Departamento de Didáctica y Organización Escolar de la Universidad de Granada y se presentan en portugués. De 2009 al 31/12/2014 se registran 17 tesis con mención internacional en TESEO, siendo el periodo más prolífico el que va desde septiembre de 2013 a diciembre de 2014 con 9 menciones, lo que también coincide con uno de los periodos académicos en los que más tesis se leyeron.

Dirección de tesis doctorales y participación en tribunales

El número de directores por tesis doctorales en el periodo del 1976/2014, indexadas bajo el descriptor "Sector de la educación", ha sido de 292 (60,20%) trabajos con 1 sola dirección, 172 (35,46%) trabajos con 2 direcciones y 21 (4,32%) trabajos con 3 direcciones. No se han encontrado tesis doctorales con más de tres directores. Sin embargo, consideramos interesante destacar que hay 4 trabajos registrados donde se contempla la figura administrativa del tutor/ponente, es decir, que han sido dirigidas fuera de la universidad en las que han estado inscritas y han necesitado de una figura nexo entre la universidad de inscripción y la universidad de origen de la dirección. En todos los casos la dirección de estas tesis se ha llevado

a cabo por un solo profesor. Esta modalidad de dirección de tesis doctoral representa el 0,82% de la muestra, por lo que podemos afirmar que no es habitual que la universidad de inscripción del trabajo no cuente con al menos un director en el proyecto.

Centrándonos en las direcciones de las tesis doctorales defendidas, se ha considerado reseñable quienes han dirigido al menos 4 trabajos de investigación. Este criterio se ha seguido considerando que permite establecer una trayectoria relevante como director o directora de trabajos de investigación.

En este sentido, consideramos relevante resaltar el caso de la Universidad Antonio de Nebrija, que cuenta con el director más productivo, el Doctor José Manuel Saiz Álvarez, con ocho tesis dirigidas de las nueve que se han defendido en dicha Universidad bajo el descriptor "Sector de la educación" entre los años estudiados. Tal y como señala la propia universidad en su página web oficial, entre las razones para elegirla como centro de estudios está que es "deliberadamente pequeña y personalizada", con 3.000 alumnos matriculados, factor por el que podemos inferir que la producción de tesis totales analizadas en el presente estudio no sea significativa, pero se encuentren concentradas en pocos directores.

La segunda universidad con mayor número de directores, en base al número de investigaciones dirigidas, es la Universidad de Granada. En dicha

universidad, ocupa la segunda posición el Doctor Juan Torres Guerrero, del Departamento de Didáctica de la Expresión Musical, Plástica y Corporal, con seis trabajos dirigidos. Le siguen los Doctores Mar Cepero González (Didáctica de la Expresión Musical, Plástica y Corporal) y Manuel Lorenzo Delgado (Didáctica y Organización Escolar), con cuatro trabajos dirigidos respectivamente. De los datos extraídos y analizados, podemos afirmar que es la Facultad de Ciencias de la Educación de la Universidad de Granada la que concentra el mayor número de directores con mayor número de tesis dirigidas dentro de la muestra analizada.

Las Universidades de Barcelona junto con las Universidades de Sevilla, Vic, y Autónoma de Madrid cuentan cada una de ellas con un director que haya dirigido hasta cuatro trabajos de investigación (véase tabla III).

En relación con el sexo de los docentes que han dirigido tesis doctorales, llama la atención que el 63,55% de la muestra analizada han sido dirigidas por hombres (347 hombres) y el 36,44% han sido dirigidas por mujeres (199 mujeres). En relación al número de hombres (1.108) que forman parte como miembros de los tribunales de las tesis es superior al de las mujeres (539) con porcentajes del 67,27% y 32,72% respectivamente. También el número de hombres que han participado en más ocasiones como miembros de los tribunales de tesis (30), es superior al de las mujeres (9). Sin embargo, en el caso de los autores (doctorandos), el número de mujeres (267) es superior al de los hombres (218), con porcentajes del 55,05 % y 44,94% respectivamente, como ya se mencionaba con anterioridad (véase figura 5).

La tabla IV muestra quiénes son los académicos que han sido seleccionados en un mayor número

de ocasiones para evaluar tesis doctorales participando para ello en los correspondientes tribunales, así como el cargo con el que han actuado en ellos.

De esta forma encontramos un académico que ha participado en tribunales más de 10 veces, cinco que han sido convocados al menos siete, dos que han participado seis veces en tribunales, once académicos que han estado cinco veces en ellos, y veinte que han formado parte de tribunales de evaluaciones de tesis doctorales al menos cuatro veces. Un total de 39 docentes que durante el periodo estudiado han participado como mínimo cuatro veces como evaluadores de trabajos doctorales.

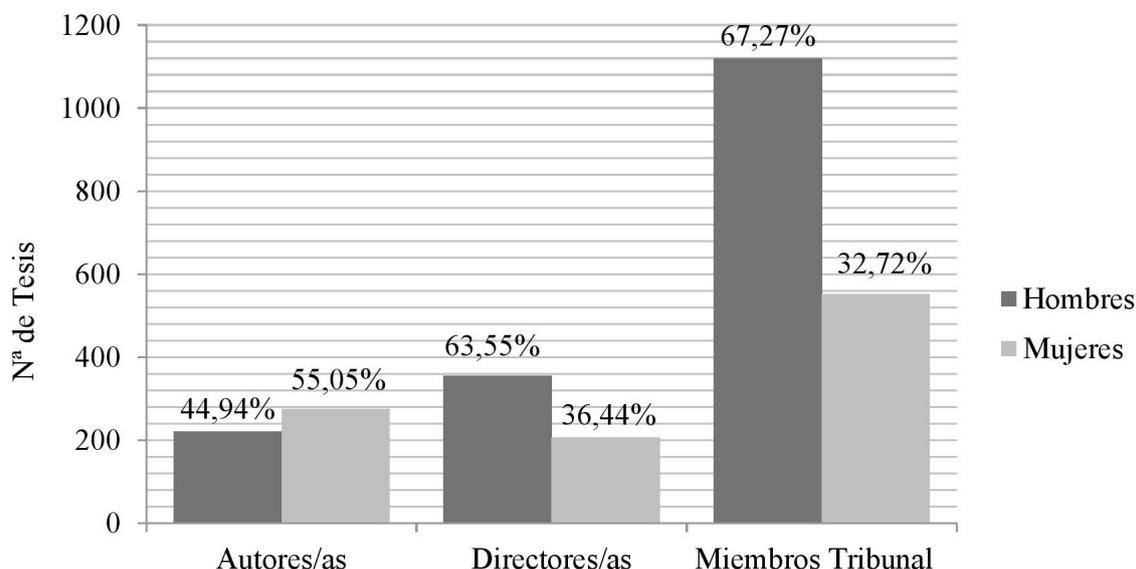
De estos casos, destacar al Doctor José Ginés Mora Ruiz de la Politécnica de Madrid, del ámbito de las Ciencias Económicas y Empresariales, con más de 10 participaciones en tribunales con diferentes cargos aunque, como se puede observar, sólo una de las veces ha actuado como presidente.

Como presidentes de tribunales, son el Doctor Francisco Imbernón Muñoz (Universidad de Barcelona) y la Doctora Leonor Buendía Eisman (Universidad de Granada), los investigadores que más veces han ocupado este cargo, cinco veces respectivamente.

Con al menos cuatro participaciones en tribunales de evaluación de tesis doctorales encontramos 20 profesores, de los que consideramos relevante destacar a los Doctores Manuel Lorenzo Delgado (Universidad de Granada) y Jorge Calero Martínez (Universidad de Barcelona) se encuentran tanto entre los directores de tesis más productivos, tal y como se señalaba anteriormente en la tabla III, como entre los que más han participado en tribunales, con cuatro tesis dirigidas y cuatro participaciones en tribunales respectivamente.

Tabla III. Profesorado con más de tres direcciones de tesis indexadas en "Sector de la educación" durante el período 1976/2014

Director/a	Tesis Dirigidas	Departamento	Universidad
Saiz Álvarez, José Manuel	8	Economía y administración de empresas	Antonio de Nebrija
Torres Guerrero, Juan	6	Didáctica de la expresión musical, plástica y corporal	Granada
Cepero González, Mar	4	Didáctica de la expresión musical, plástica y corporal	Granada
Lorenzo Delgado, Manuel	4	Didáctica y organización escolar	Granada
Calero Martínez, Jorge	4	Economía política, hacienda pública, derecho financiero y tributario	Barcelona
Murillo Torrecilla, Francisco J.	4	Didáctica y teoría de la educación	Autónoma de Madrid
Romero Granados, Santiago	4	Educación física y deporte	Sevilla
Simó Gil, Núria	4	Departament de ciències de l'activitat física	Vic

Figura 5. Número de autores, directores y miembros de los tribunales de las tesis según su sexo

En cuanto a qué Universidad tiene más profesorado participante en tribunales destaca la Universidad de Barcelona, con un total de cuatro docentes, que han participado en tribunales entre seis y cuatro veces, ostentando diferentes cargos.

4. DISCUSIÓN Y CONCLUSIONES

Destacamos que desde 1976 y hasta el curso 1999/2000 las producciones registradas son del ámbito del sector de las Ciencias Económicas, cuyo campo de estudio se centra, en general, en las enseñanzas superiores o universitarias. Por este motivo, las temáticas más estudiadas son las relativas al mercado de trabajo de los titulados superiores, el gasto público de la enseñanza a distintos niveles, los costes de la educación, la incidencia económica de la universidad en el ámbito local, la eficacia del sistema educativo, o el salario percibido por los titulados superiores.

Este interés de las Ciencias Económicas en el campo de la educación siempre ha estado presente, la economía de la educación como tal es un área aún joven, con no más cincuenta años. Son las aportaciones de Schultz (1961), y posteriormente las de Becker (1964), las que marcan un antes y un después en el interés de la economía en el sector educativo. Este auge de trabajos de investigación de economía centrados en el sector de la educación está estrechamente relacionado con el fuerte crecimiento económico que se ocasionó tras la finalización de la II Guerra Mundial en la mayoría de los países occiden-

tales. La democratización de la enseñanza y la demanda educativa que se produjo a partir de los años 60 cambió por completo el nivel de producción de los sistemas educativos, lo que conllevó el interés de los economistas por este sector. La crisis económica de la primera mitad de los años 70 modificó las perspectivas de progreso de los sistemas educativos, incidiendo en que los presupuestos públicos no pudieran afrontar un aumento sostenido de la demanda de la enseñanza. Esta situación derivó en que las Ciencias Económicas se interesaran en la investigación sobre cómo financiar los gastos de los sistemas educativos de una manera eficiente. Es destacable que estos estudios también se interesaron por el personal docente, sus cualificaciones y sus condiciones de trabajo, así como por sus motivaciones a la hora de elegir la carrera docente, tal y como hemos podido confirmar a través de la presente investigación. Estudios sobre la economía de la educación que, como señalan Lassibille y Navarro (2012), han estado y seguirán estando muy presentes en este nuevo contexto de cambios en el que nos encontramos.

En España, la primera tesis registrada bajo el descriptor "Sector de la educación" es el trabajo del Doctor Tomás Moltó García, presentada en español bajo el título *El mercado de trabajo de los titulados superiores en Cataluña 1964-1974*, leída el 01/01/1978 en la Universidad Autónoma de Barcelona y dirigida por el ya fallecido Doctor Santiago Roldán López, que fue catedrático de Estructura Económica.

Tabla IV. Número y tipología de las participaciones de los profesores en tribunales de tesis indexadas en "Sector de la educación" durante el período 1976/2014

Tribunal	Universidad	Nº Trib.	Nº de veces/ Cargo Ocupado		
			Pres.	Sec.	Voc.
Mora Ruiz, José Ginés	Polit. Valencia	12	1	0	11
Calero Martínez, Jorge	Barcelona	7	0	1	6
Díaz Gómez, Blanca Maravillas	País Vasco	7	4	0	3
Toharia Cortés, Luis	Alcalá	7	4	0	3
San Segundo Gómez Cadiñanos, M ^a Jesús	País Vasco	7	0	2	5
Oroval Planas, Esteban	Barcelona	7	2	0	5
Imbernon Muñoz, Francisco	Barcelona	6	5	0	1
Giné Giné, Climent	Ramón Llul	6	3	1	2
Buendía Eisman, Leonor	Granada	5	5	0	0
Coruego de las Cuevas, Indalecio	Compl.Madrid	5	2	1	2
González Gallego, Isidoro	Valladolid	5	4	0	1
Jiménez Eguizábal, Alfredo	Burgos	5	1	1	3
Jimeno Serrano, Juan F.	Alcalá	5	0	0	5
Llorent Bedmar, Vicente	Sevilla	5	1	3	1
Zagalaz Sánchez, María Luisa	Jaén	5	3	0	2
Salinas Jiménez, Francisco Javier	Extremadura	5	0	2	3
Sáenz-López Buñuel, Pedro	Huelva	5	3	0	2
Puig Rovira, Josep Maria	Barcelona	5	4	1	0
Pedraja Chaparro, Francisco	Extremadura	5	2	0	3
Arruza Gabilondo, José Antonio	País Vasco	4	2	0	2
Caballero Fernández, Rafael	Málaga	4	2	1	1
Cabero Almenara, Julio	Sevilla	4	2	0	2
Calaf Masachs, Roser	Oviedo	4	2	0	2
Caride Gómez, José Antonio	S. de Compostela	4	1	0	3
Domingo Segovia, Jesús	Granada	4	1	3	0
Gilar Corbi, Raquel	Alicante	4	0	3	1
González González de Mesa, Carmen	Oviedo	4	0	1	3
Hernández Pina, Fuensanta	Murcia	4	1	1	2
Joyanes Aguilar, Luis	Pontificia	4	3	0	1
Torrego Egido, Luis Mariano	Valladolid	4	1	2	1
Sola Martínez, Tomás	Granada	4	4	0	0
Sanchís Ramírez, José Pascual	Sevilla	4	0	1	3
Salinas Ramos, Francisco	Pontificia	4	0	0	4
Pèlach Busom, Joaquim	Girona	4	1	0	3
Martínez Patiño, María José	Vigo	4	0	3	1
Martínez Martín, Miquel	Barcelona	4	4	0	0
Mallart i Navarra, Joan	Barcelona	4	4	0	0
Lorenzo Delgado, Manuel	Granada	4	2	1	1
Prats Cuevas, Joaquín	Barcelona	4	1	0	3

En el curso 2000/2001 se registran nueve trabajos indexados bajo el descriptor "Sector de la educación", aunque si bien siguen perteneciendo al ámbito de la empresa y/o la economía se detecta que se centran en la vivienda y la construcción, por lo que el descriptor utilizado no parece ser el más adecuado para referirse a ellas.

No es hasta el curso 2001/2002 cuando se detectan las primeras tesis relativas al ámbito de la Educación o la Pedagogía bajo este descriptor, aunque aún con carácter anecdótico. De las veinte tesis registradas en ese curso académico, dieciséis son relativas a la Empresa o a la Economía, dos a la Pedagogía y dos a la Educación, aunque de estas últimas, una de ellas se centra en el valor económico de la educación.

Es a partir del curso 2002/2003 que la tendencia empieza a cambiar y, ya en el curso 2013/2014 se contabilizan cincuenta y nueve trabajos los que abordan cuestiones relativas a las didácticas específicas o a temáticas relativas del campo de la Pedagogía. El resto de la muestra de este curso la conforman trabajos de los campos de la Economía, la Empresa, la Psicología, la Ingeniería, las Humanidades o la rama sanitaria, vinculándose todas ellas, de alguna forma, a temáticas educativas. Es en esta cuestión donde ponemos el acento, no podemos perder de vista que son varias las disciplinas científicas que intervienen en el estudio del campo educativo, cada una de ellas centrándose en aquellas dimensiones que le corresponde abordar.

Una de las cuestiones que pueden explicarlo es que en España, hasta finales del siglo XX, no hubo una Facultad propia de Educación en la que se pudiera obtener el grado académico de Doctor. No fue hasta el curso académico 1975/1976, que la Facultad de Filosofía y Letras, en donde ya se impartían los estudios de Pedagogía, se reestructura y pase a denominarse Facultad de Filosofía y Ciencias de la Educación. Aun así, hasta 1991 no comenzarán a crearse en España las primeras Facultades de Educación.

Existe un cambio en el nivel de producción que traslada el protagonismo de las investigaciones realizadas desde el campo de las Ciencias Económicas y Empresariales en los años 70, lo que consideramos está relacionado con el auge de estos estudios como respuesta a las crisis económica que sufrió el país y que repercutió en la educación fuertemente. Por otro lado, las tesis doctorales de Ciencias de la Educación, inician su auge en el descriptor a estudio a partir del inicio del siglo XXI. Este hecho no creemos que sea mera casualidad, sino que este aumento de

investigaciones se ha visto favorecido con la consolidación de las Facultades de Educación y el afianzamiento en el ámbito de la investigación de las didácticas específicas.

Durante la década de 2005/2014, el número de tesis catalogadas con el descriptor "Sector de la educación" en universidades españolas se ha incrementado de forma extraordinaria, habiéndose multiplicado por tres la producción con respecto a la década anterior. Los datos disponibles muestran que la evolución sigue siendo creciente, superando ya los 60 registros anuales. Este progreso ha transformado la distribución temática: se ha reducido el peso porcentual de las disciplinas relacionadas con la empresa y la economía, que son las que abordan este campo hasta que en 2002 la tendencia empieza a cambiar. Es a partir de 2013 cuando las Ciencias de la Educación comienzan a posicionarse, abordándose cuestiones relativas a las didácticas específicas o a temáticas relativas del campo de la pedagogía.

Ambos hechos, incremento de tesis y variación de la orientación temática, reflejan un cambio notable en la presencia de los temas educativos dentro de los estudios de doctorado en España, que posiblemente no se haya dado en décadas anteriores debido a la escasa tradición de las carreras de magisterio a finalizar con tesis doctorales.

La producción por universidades está liderada por las Universidades de Barcelona y Granada. También entre los primeros puestos de directores más productivos aparecen estas dos universidades junto con la Universidad Antonio de Nebrija. Los datos relativos al idioma de lectura indican que sigue siendo el español el idioma predominante, aunque aumentan las tesis con el resumen y las conclusiones en otros idiomas para la obtención de la mención europea/internacional.

Sería interesante realizar un estudio sobre la representación correspondiente a la figura 4, ya que consideramos debe existir una relación entre las proporciones de sexo tanto en autores, como en directores y miembros de tribunal, que explique la evolución de la representatividad de la mujer en los estudios superiores. Una de las cuestiones que nos hacen plantearnos esta nueva pregunta es que, como hemos podido comprobar, a través del tiempo el aumento del número de mujeres universitarias ha derivado en una creciente producción de tesis doctorales, superando el número total defendidas por hombres en la última década (figura 2). Esto debe conllevar un aumento escalonado en la representación de mujeres como miembros de tribunal y, por lo tanto, como directoras.

5. REFERENCIAS

- Agudelo, D.; Ortiz-Recio, G.; Vico Fuillerat, C.; Valor-Segura, I.; Bretón-López, J.; Poveda-Vera, J.; Teva Álvarez, I. (2003). Análisis de la productividad científica de la psicología española a través de las tesis doctorales. *Psicothema*, vol. 15 (4), 595-609.
- Ariza, T.; Quevedo-Blasco, R. (2013). Bibliometric analysis of the Revista de Investigación Educativa (2000-2012). *Revista de Investigación Educativa*, vol. 31(1), 31-52. <http://dx.doi.org/10.6018/rie.31.1.160321>
- Ariza, T.; Quevedo-Blasco, R.; Bermúdez, M.P.; Buela-Casal, G. (2012). Los estudios de doctorado en España: de la mención de calidad a la mención hacia la excelencia. *Aula Abierta*, vol. 40 (2), 39-52.
- Becker, G. S. (1964). *Human Capital: A theoretical and Empirical Analysis, with Special References to Education* (1ª ed.). New York, National Bureau of Economic Research, p. 264.
- Bermúdez, M. P.; Castro, Á.; Sierra, J. C.; Buela-Casal, G. (2009). Análisis descriptivo transnacional de los estudios de doctorado en el EEES. *Revista de Psicodidáctica*, vol. 14 (2), 193-210.
- Bermúdez, M. P.; Guillén-Riquelme, A.; Gómez García, A.; Quevedo-Blasco, R.; Sierra, J. C. (2011). Análisis del rendimiento en el doctorado en función del sexo. *Educación XX1: Revista de la Facultad de Educación*, vol. 14(1), 17-33. <http://dx.doi.org/10.5944/educxx1.14.1.261>
- Buela-Casal, G. (2005). An overview of scientific productivity of Spanish Universities. *International Journal of Clinical and Health Psychology*, vol. 5 (1), 175-190.
- Buela-Casal, G. (2010). Índices de impacto de las revistas científicas e indicadores para medir el rendimiento de los investigadores. *Revista de Psicodidáctica*, vol. 15 (1), 3-19.
- Buela-Casal, G.; Zych, I. (2010). Analysis of the relationship between the number of citations and the quality evaluated by Experts in psychology journals. *Psicothema*, vol. 22 (2), 270-276.
- Carbonell, X.; Calvó, N. (2009). Las revistas españolas de Psicología: cómo elegir la revista donde publicar. *Anales de Psicología*, vol. 25 (2), 209-216.
- Carpallo, A.; Burgos, E. (2008). Estudio bibliométrico y de calidad de la Revista Complutense de Educación (1990-2007). *Revista Complutense de Educación*, vol. 19 (1), 13-30.
- Casanueva, C.; Escobar, B.; Larrinaga, C. (2007). Red social de contabilidad en España a partir de los tribunales de tesis. *Revista Española de Financiación y Contabilidad*, núm. 136, 707-726.
- Castillo, A.; Xifra, J. (2006). Investigación bibliométrica de las tesis doctorales españolas sobre relaciones públicas. *Anàlisi: quaderns de comunicació i cultura*, núm. 34, 141-161.
- Collados, E. (2010). La enseñanza del Dibujo a través de los libros de texto de educación obligatoria publicados en España (1915-1990): estudio bibliométrico de contenidos. *Revista de Educación*, núm. 352, 517-544.
- Conferencia de Rectores de Universidades Españolas [CRUE] (2013). *La universidad española en cifras 2012*, Madrid, Cátedra UNESCO de Gestión y Política Universitaria de la Universidad Politécnica de Madrid, p. 244.
- Delgado, E.; Torres, D.; Jiménez, E.; Ruiz-Pérez, R. (2006). Análisis bibliométrico y de redes sociales aplicado a las tesis bibliométricas defendidas en España (1976-2002): temas, escuelas científicas y redes académicas". *Revista Española de Documentación Científica*, vol. 29 (4), 493-524.
- Díaz del Campo, J. (2014). La investigación sobre Internet en las facultades de comunicación española. Análisis bibliométrico de tesis doctorales (1997-2012). *Documentación de las Ciencias de la Información*, núm. 37, 305-320.
- Fernández, A.; Bueno, Á. (1998). "Síntesis de estudios bibliométricos españoles en Educación. Una dimensión evaluativa". *Revista Española de Documentación Científica*, vol. 21, núm. 3, pp. 269-285.
- Fernández-Cano, A.; Torralbo, M.; Vallejo, M. (2008). Revisión y prospectiva de la producción española en tesis doctorales de Pedagogía (1976-2006). *Revista de Investigación Educativa*, vol. 26 (1), 191-208.
- Fuentes, E.; Argimbau, L. (2010). Las tesis doctorales en España (1997-2008): análisis, estadísticas y repositorios cooperativos. *Revista Española de Documentación Científica*, vol. 33 (1), 63- 89. <http://dx.doi.org/10.3989/redc.2010.1.711>
- Galante, R.; Sanchidrián, M. C. (1996). Evaluación de los trabajos de grado en Historia de la Educación desde una perspectiva bibliométrica (1980-1990). *Revista de Investigación Educativa, RIE*, vol. 14 (1), 37-62.
- Gómez-García, A.; Ramiro, T.; Ariza, T.; Granados, R. (2012). Estudio bibliométrico de educación XX1. *Educación XX1*, vol. 15 (1), 17-41.
- Granados, R.; Ariza, T.; Gómez-García, A., Ramiro, T. (2011). Estudio bibliométrico de Aula Abierta. *Aula Abierta*, vol. 39 (3), 97-110.
- Gutiérrez-Saldivia, X. (2014). *Evaluación de la investigación sobre la educación especial en España mediante el análisis cuantitativo de sus tesis doctorales (1978-2013)*. Trabajo Fin de Máster. Dirigido por Antonio Fernández Cano. Granada, Universidad de Granada.
- Haba, J.; Osca-Lluch, J.; Muñoz, J. (2000). *La investigación en cardiología de la Comunidad Valenciana vista a través de las tesis doctorales*. Valencia, Instituto Cardiovascular Clínica Quirón, p. 94.

- Jiménez-Contreras, E.; Ruiz-Pérez, R.; Delgado, E. (2014). El análisis de las tesis doctorales como indicador evaluativo: reflexiones y propuestas. *Revista de Investigación Educativa*, vol. 32 (2), 295-308.
- Lassibille, G.; Navarro, M. L. (2012). Un compendio de investigaciones en economía de la educación. *Presupuesto y Gasto Público*, núm. 67, 9-28.
- López-Meneses, E.; Vázquez, E.; Sarasola, J. L. (2015). Estudio bibliométrico de Pixel-bit, revista de medios y educación (2000-2013). *Pixel-Bit: Revista de medios y educación*, núm. 46, 65-85.
- Marcos Recio, J. C.; Martínez Pestaña, M. J.; Blasco López, M. F. (2012). Producción y dirección de tesis doctorales sobre publicidad en la universidad española (1971-2010). *Revista Española de Documentación Científica*, vol. 35 (3), 433-452. <http://dx.doi.org/10.3989/redc.2012.3.890>
- Martínez Pestaña, M. J. (2004). La producción de tesis doctorales sobre temas publicitarios (1971-2001). *Documentación de las Ciencias de la Información*, núm. 27, 237-267.
- Martínez Pestaña, M. J. (2011a). *Evaluación de la producción científica sobre publicidad (1971-2001)*. Tesis doctoral. Facultad de Ciencias de la Información. Madrid, Universidad Complutense de Madrid.
- Martínez Pestaña, M. J. (2011b). La investigación universitaria en Publicidad: producción y temática de las tesis doctorales (1971-2001). *Documentación de las Ciencias de la Información*, núm. 34, 119-156.
- Max, A.; Bracho, R.; Torralbo, M.; Gutiérrez, M. P.; Jiménez, N.; Adamuz, N. (2012). Redes académicas generadas por las tesis doctorales de educación matemática en España. *Revista de Investigación Educativa, RIE*, vol. 30 (2), 253-270.
- Musi-Lechuga, B.; Olivas-Ávila, J. A.; Castro, Á. (2011). Productividad en tesis de los programas de doctorado en Psicología con mención de calidad. *Revista Mexicana de Psicología*, núm. 28, 93-100.
- Olivas-Ávila, J. A.; Musi-Lechuga, B. (2010). Producción en tesis doctorales de los profesores funcionarios de Psicología en España más productivos en la Web of Science. *Psicothema*, vol. 22 (4), 917-923.
- Olivas-Ávila, J. A.; Musi-Lechuga, B.; Guillén-Riquelme, A.; Castro, Á. (2012). Diferencias en la producción investigadora en tesis y artículos de los profesores funcionarios de Psicología en España en función del sexo. *Anales de Psicología*, vol. 28 (2), 597-603. <http://dx.doi.org/10.6018/analesps.28.2.132691>
- Ortega, E. (2010). La investigación en marketing a través de las tesis doctorales españolas. *Investigaciones Europeas de Dirección y Economía de la Empresa*, vol. 16 (1), 127-147. [http://dx.doi.org/10.1016/S1135-2523\(12\)60006-9](http://dx.doi.org/10.1016/S1135-2523(12)60006-9)
- Oscala-Lluch, J.; Haba, J.; Fonseca, S.; Civera, C.; Tortosa, F. (2013). Tesis doctorales españolas sobre análisis bibliométrico en psicología. *Aula Abierta*, vol. 41 (2), 99-110.
- Pérez-Sánchez, A. M.; López-Alacid, M. P.; Poveda-Serra, P. (2009). Aprendizaje cooperativo y formación del profesor. Un estudio bibliométrico (1997-2008). *Anales de Documentación*, núm. 12, 209-220.
- Repiso, R.; Torres, D.; Delgado, E. (2011a). Análisis bibliométrico y de redes sociales en tesis doctorales españolas sobre televisión (1976/2007). *Comunicar*, núm. 37, 151-159.
- Repiso, R.; Torres, D.; Delgado, E. (2011b). Análisis de la investigación sobre Radio en España: una aproximación a través del Análisis Bibliométrico y de Redes Sociales de las tesis doctorales defendidas en España entre 1976-2008. *Estudios sobre el mensaje periodístico*, vol. 17 (2), 417-429.
- Robinson, N.; Delgado, E.; Torres, D. (2011). Cómo comunicar y diseminar información científica en Internet para obtener mayor visibilidad e impacto. *Aula Abierta*, vol. 39 (3), 41-50.
- Sánchez, J. M.; Marcos, J. C.; Olivera, M. (2014). Tesis doctorales sobre fotografía en la universidad española. Análisis de la producción y dirección (1976-2012). *Revista Española de Documentación Científica*, vol. 37 (1), 1-14.
- Schultz, T. W. (1961). Investment in Human Capital. *American Economic Review*, núm. 51, 1-17.
- Schultz, T. W. (2004). School subsidies for the poor: evaluating the Mexican PROGRESA poverty program. *Journal of Development Economics*, núm.74, 199-250. <http://dx.doi.org/10.1016/j.jdeveco.2003.12.009>
- Sorli, Á.; Merlo, J. A. (2002). Bases de datos y recursos en Internet de tesis doctorales. *Revista Española de Documentación Científica*, vol. 25 (1), 95-106.
- Torres, I.; Torres, D. (2005). Tesis doctorales sobre estudios de las mujeres en España (1976- 2002): A propósito de un indicador definitivo en investigación. *Revista Española de Documentación Científica*, vol. 28 (4), 479- 499.
- Torres, I.; Torres, D. (2007). *Tesis doctorales sobre estudios de las mujeres en España (1976- 2002): Análisis bibliométrico y repertorio bibliográfico*. Sevilla, Instituto Andaluz de la Mujer, p. 411.
- Villar, A. (2011). El "eigenfactor": un nuevo y potente instrumento bibliométrico para evaluar la investigación. *Aula Abierta*, vol. 39 (3), 85-96.

CRÍTICA DE LIBROS / BOOK REVIEWS

Métricas de la Web Social para Bibliotecas

Nieves González-Fernández-Villavicencio

Barcelona: Editorial UOC, 2014 (El profesional de la información: 4). 282 p. ISBN: 978-84-9116-181-3

Los medios sociales online han ido incorporándose, con mayor o menor acierto, a la gestión bibliotecaria en los últimos años. Ha habido rechazo por parte de algunos profesionales que no consideraban que fuera parte de sus atribuciones y también ha habido administraciones públicas que han retardado la presencia institucional en las redes sociales. Por otra parte, tenemos profesionales de la biblioteconomía, inspirados por la cultura digital, que llevan ya unos años actuado como 'punta de lanza' en la utilización de estas plataformas, y gracias a su empuje podemos señalar que, en la actualidad, el uso de la web social en las bibliotecas se encuentra en parte consolidado.

El trabajo constante de todos ellos demuestra que los medios sociales son útiles para ampliar el alcance de la labor bibliotecaria, conforme a las necesidades que la sociedad demanda en la era de la información. Aunque la principal finalidad de las bibliotecas no es el lucro, puede resultar legítimo asimilarlas a empresas en lo referente a que reciben un dinero que necesitan gastar para cumplir sus objetivos, buscando la satisfacción de los usuarios, que serían los clientes. Por ello, si los medios sociales se utilizan como herramienta de marketing, su uso requiere un planteamiento sistemático, incluyendo la consideración de métricas e indicadores (KPIs) cuantitativos y cualitativos, que permitan comprobar que la cantidad de tiempo y de dinero invertido produce un retorno adecuado de la inversión.

Nieves González-Fernández se ocupa, en el nº 4 de la colección EPI Scholar, de la evaluación del rendimiento de la estrategia de marketing, me-

diante la revisión de informes, de sitios web comerciales y de datos referentes a bibliotecas. Realiza un recorrido por algunas de las iniciativas y buenas prácticas en relación con el uso de métricas e indicadores de la web social, que sirve para elaborar un cuadro de las más utilizadas, tanto las que permiten evaluar un momento puntual como aquellos que facilitan explorar la evolución y permiten la comparación de las bibliotecas con técnicas de benchmarking. Describe las principales herramientas necesarias para la toma de datos y esboza un mapa mental de las mismas. Mediante técnicas de correlación y regresión, identifica las variables predictivas y las reglas de dependencia que existen entre ellas y sus respectivos objetivos de negocio. Aunque no se ha encontrado un modelo matemático determinante para la predicción de resultados, las aproximaciones evidenciadas han permitido deducir con cierta seguridad que las actividades realizadas en la web social producirán a medio plazo un impacto positivo.

Este trabajo emplea un enfoque mixto, con una primera parte en la que realiza un desarrollo esencialmente teórico, y los resultados de la aplicación de la definición y la selección de indicadores y métricas en los capítulos finales. El lector podrá inicialmente actualizar su conocimiento sobre la evolución de las relaciones del usuario con la biblioteca, como consecuencia de la incorporación de los medios sociales online. El segundo capítulo, dedicado a los costes, resulta una matización fundamental en cuanto a su aplicación a las bibliotecas, ya que al no ser el lucro su principal objetivo, no resulta muy apropiado aplicar el concepto de ROI.

Para solventar la especial situación de la biblioteca respecto a la consideración de costes-beneficios, se han considerado el IOR (impact on relationship), el SROI (ROI social) y el ROC o retorno en colaboración, que miden la rentabilidad como consecuencia del grado de fidelización y de compromiso con la marca que adquiere el usuario. La estimación del coste que supone la dedicación a los medios sociales se evalúa en base al coste de la tecnología empleada, del personal dedicado y la inversión en anuncios patrocinados. La autora ha ejemplificado su análisis con un estudio de dieciocho bibliotecas realizado en 2013.

La obtención del ROI necesita métricas que lo evidencien de forma concreta. El principal problema es la falta de normativa estandarizada para elaborarlas. La primera parte del capítulo reflexiona sobre la <<Necesidad de métricas>> y los problemas asociados a la medición y cuantificación de los beneficios. Es destacable el esfuerzo de redefinición conceptual de las métricas de comparación y se expone la distinción entre métricas, indicadores y KPIs y entre medios propios, pagados y ganados. En una segunda parte, sintetiza las métricas que proponen expertos y organizaciones como WOMMA, IAB Spain, AMEC, IPR, Katie Paine, Dolores Vela, Isra García, la profesora Castelló-Martínez, John Lovett, Social Bakers, Juan Carlos Martínez Gallardo, y, específicamente, las propuestas para bibliotecas por el Grupo de Trabajo Estratégico para el Estudio del Impacto Socioeconómico de las Bibliotecas en la Sociedad, del Consejo de Cooperación Bibliotecaria; por Nuria Lloret-Romero; por Helene Bowers; por Julián Márquina; por María del Carmen Martín Marichal, por REBIUN y por aquellas bibliotecas que han creado sus propios sistemas de métricas como la Red de Bibliotecas Municipales de La Coruña o la Biblioteca Nacional española.

Derivado del anterior, el capítulo 4 está dedicado a describir los objetivos de negocio y los indicadores para medirlos, a partir de los cuales las bibliotecas extraerán los indicadores clave de rendimiento. Los objetivos estratégicos considerados han sido la visibilidad, la fidelización, la influencia y el retorno de la inversión, a los que se corresponden respectivamente los objetivos tácticos de alcance y frecuencia de actividad; tráfico web; percepción de la marca; y 'engagement' y conversiones. La reputación se desmarca de los anteriores al centrarse en la búsqueda de una rentabilidad no financiera (IOR), más acorde con la filosofía y modo de funcionamiento de los medios sociales, cuya medición se realiza considerando los objetivos tácticos de influencia de la marca, alcance, engagement y fidelización. En la segunda parte realiza una propuesta de indicadores para cada uno de

los objetivos tácticos descritos, con una estructura formada por la definición del objetivo y el proceso de generación de indicadores, que se repetirá en capítulos posteriores. Junto al cuadro resumen de indicadores (epígrafe 3.7), se ofrecen elementos prácticos para llevar a cabo la obtención de indicadores y su medición, como un modelo de formulario de recogida de datos; un plantilla de medición de indicadores social media desde un punto de vista sincrónico y diacrónico; una metodología de tratamiento de datos para su posterior análisis y, finalmente, el análisis de los datos mediante su estudio descriptivo, el estudio de correlaciones y el estudio de regresión.

Los capítulos 5, 6 y 7 son la aplicación práctica de las bases fijadas principalmente en el capítulo 4 del libro. El primero de ellos está dedicado a mostrar los indicadores con mayor capacidad predictiva en bibliotecas, apoyándose en un estudio estadístico de 2013. Continúa con un estudio comparativo del uso de las variables del capítulo anterior aplicadas a los principales tipos de bibliotecas, de forma que ha sido posible compararlas entre sí en base a los valores medios obtenidos. En cuanto al alcance, se consideran variables como los seguidores en Twitter y los fans de Facebook; como representativas de la frecuencia de actividad, los posts publicados en el blog y los comentarios que reciben; en relación con la fidelización, los porcentajes de rebote, de visitas y de sesiones social media; la influencia viene determinada por el índice Klout y las menciones; la participación, tomando en cuenta el número de personas hablando de esto en Facebook, los me gusta recibidos y los archivos descargados; y la conversión, con indicadores relativos al préstamo de materiales, los cursos descargados, el número de cursos impartidos o las solicitudes de información y referencia, etc.

Finalmente, el capítulo 7 constituye la comunicación de los resultados obtenidos a partir del estudio descriptivo elaborado por la autora en 2012, tomando como referencia cuatro bibliotecas pertenecientes a las tipologías más comunes: la Biblioteca de la Universidad de Cádiz, la Biblioteca Pública de La Coruña, las Bibliotecas especializadas de la Red del Instituto Cervantes y la Biblioteca Juan Leiva del IES Vega del Guadalete (Cádiz). Se presenta un análisis de correlaciones y dependencias, rectas de regresión y variables predictoras de los indicadores de uso de la web social, junto a las métricas tradicionales, con el objetivo de reflejar la actividad realizada durante un año y ponerla en relación con los objetivos generales del plan estratégico de la biblioteca, lo cual ha permitido valorar el impacto de los mismos en dicha actividad y cuantificar la rentabilidad.

Este libro es absolutamente recomendable para ponerse al día de las tendencias de uso de los medios sociales online. Ciertamente, va de lo general a lo particular, puntualizando el uso teórico de la terminología, confirmando la interdependencia de las variables de medición y demostrando que la presencia de la biblioteca en medios sociales permite, efectivamente, obtener beneficios ROI y no ROI. Estas páginas constituyen una oportunidad para que las bibliotecas entiendan la importancia de que la presencia en la web social, debe hacerse partiendo de un planteamiento sistemático que permita la medición del cumplimiento de los objetivos fijados, como parte del plan de marketing general de la institución.

Tenemos ante nosotros un libro que, probablemente, en lo referente al dato, se quede obsoleto

con el tiempo (con muestras de 2012 y 2013), pero, en cambio, realiza valoraciones de conjunto significativas que marcaran la pauta, en los próximos, para evaluar la evolución de las repercusiones de los medios sociales en la actividad bibliotecaria. Sus aportaciones se pueden completar con una segunda lectura de las referencias bibliográficas, que junto a las recomendaciones metodológicas, las figuras y las tablas, resultan de gran valor para la comprensión del conjunto del libro y de su objeto de investigación.

Adrián Macías Alegre

Director de DoKumentalistas.com
Consultor en Gestión de la Información
amaciasalegre@dokumentalistas.com

CRÍTICA DE LIBROS / BOOK REVIEWS

SEO: Introducción a la disciplina del posicionamiento en buscadores

Jorge Serrano-Cobos

Editorial UOC, 2015. Colección EPI Scholar: n. 3, 200 p. ISBN: 978-84-9064-956-5

El camino hacia la optimización de la experiencia de búsqueda o la entrada del posicionamiento web en la academia

El SEO y la academia

El SEO, por sus siglas en inglés Search Engine Optimization, tuvo su necesario nacimiento en el seno de la industria. En algún momento de los años 90 las empresas comprendieron que si querían aumentar el tráfico de sus sitios debían asegurarse de que sus páginas tuvieran visibilidad en los buscadores. Y esto último, solo podía significar que los contenidos de su sitio debían aparecer entre las primeras posiciones de la página de resultados de los buscadores.

De aquí siguió con naturalidad el perfil profesional del responsable en la optimización de sitios web. Estos profesionales debían escrutar con la mayor eficacia posible la forma de analizar e indizar la información que seguía cada buscador (sí, hubo una época en la cual tenía sentido usar el término en plural). En paralelo, los académicos de nuestra disciplina estudiaban y criticaban a partes iguales a los buscadores como sistemas de información.

La cuestión es que durante años teníamos una importante brecha que se manifestaba en varios frentes. Por un lado, la Biblioteconomía-Documentación como disciplina estudiaba los buscadores, pero solo para evaluar hasta qué punto eran eficaces para encontrar información, no sin un fuerte criticismo al que muchos se sentían obligados para defender, alternativamente, el papel de las bibliotecas o de las bases de datos académicas frente a los primeros.

Por el otro, aunque se supone que los profesionales de nuestro sector están inmejorablemente preparados para todo aquello que tiene relación con la gestión de la información (entendida como conocimiento, y no como meros datos), el SEO no formaba parte de su horizonte, y por tanto, tampoco estaba entre sus intereses y, como consecuencia, tampoco como parte de su perfil competencial.

Como señalábamos antes, esta brecha, por suerte para todos, incluidas tanto la sociedad como la academia, se ha ido cerrando. Poco a poco, la Biblioteconomía-Documentación como disciplina ha ido incorporando el estudio de la búsqueda en la web sin complejos y sin sentir la necesidad de desacreditarla. Los planes de estudio de nuestros grados o postgrados incluyen con naturalidad aspectos de posicionamiento, poniendo así en el mercado profesionales del SEO con un perfil competencial ideal para ahora que cada vez más el SEO se basa en una combinación de estrategia de contenidos y de aspectos de SEO técnicos como el código fuente y los metadatos.

Todo esto viene a cuento de la reciente obra de Serrano-Cobos que aborda el SEO, precisamente, como disciplina. Una obra como esta actúa, en paralelo con sus cualidades intrínsecas, como símbolo tangible de la desaparición, al menos parcial, de esta brecha.

La cuestión es que, siguiendo el modelo propio de la colección, la obra que nos ocupa combina aspectos que serán útiles para profesionales, pero

también para estudiosos (*academia*). De todos modos, seguramente sea éste último el perfil que más se beneficie de este trabajo debido a su capacidad para presentar los temas fundamentales de una forma muy clarificadora, y por su elección de aquellos aspectos que tienen un interés más estructural y permanente en el tiempo, como corresponde a una monografía, ya que se supone que debe ser vigente en un periodo que se cuente en años (y no en meses).

Índice de la obra y factores SEO

El contenido del libro tiene como núcleo sendos capítulos dedicados a la investigación de palabras clave y a los denominados factores de posicionamiento, tanto los internos como los externos. En concreto, los títulos de los capítulos son los siguientes:

- Introducción al SEO
- Keywords & Context Research: análisis de palabras clave y del contexto SEO
- SEO On Site: técnicas dentro del sitio web
- SEO Off Site: técnicas fuera del sitio web
- Analítica web para SEO

De forma acertada incluye también aspectos de SEM (Search Engine Marketing) en diversos apartados, dado que cada vez es más difícil separar uno de otro, al menos cuando se trata de exponer los fundamentos del posicionamiento en buscadores. Otra cosa es desde el punto de vista económico, donde la diferencia puede parecer evidente: el SEO no requiere inversión directa en publicidad, mientras que se supone que el SEM sí.

Pero tampoco debemos confundirnos, y establecer las ecuaciones fáciles (y engañosas) según las cuales $SEO = \text{coste cero}$ y $SEM = \text{inversión}$, porque el buen posicionamiento requiere también inversión en contenidos, en mantenimiento del sitio, en monitorización, etc.; y todo esto implica tiempo y dinero.

Como hemos señalado antes, dos de los capítulos centrales se dedican a los factores de posicionamiento que, acertadamente, el autor divide en OnSite y OffSite, o internos y externos.

Como explica Serrano-Cobos, los criterios internos se refieren principalmente al contenido y, por lo tanto, por un lado, a aspectos estadísticos o de frecuencia de ocurrencia de la palabra (o palabras) clave de la pregunta. En concreto, las páginas web que tengan mayor densidad de la palabra clave, serán más relevantes; siempre que esa frecuencia no sea anti natural, es decir, que no se deba a un

caso de sobre optimización SEO, ya que es algo que los buscadores actuales detectan. Cuando en una página los términos del lenguaje ocurren con frecuencias superiores a las estadísticas del lenguaje natural, los buscadores consideran que se trata de una página fraudulenta o de baja calidad.

Por este motivo, esta frecuencia estadística debe estar relacionada con su aparición en contenidos de calidad. Por tanto, los aspectos intrínsecos se refieren en definitiva a la calidad de los contenidos del sitio, tanto en contenidos de texto como multimedia.

Otros aspectos, como que la palabra clave aparezca rodeada de etiquetas como `<h1>` o `<h2>`, es decir, formando parte de los títulos o subtítulos de la página, también otorga mayor importancia relativa a la página correspondiente, así como el hecho de que la palabra clave forme parte de la URL de la página.

También son decisivos aspectos como la calidad del código fuente (uso de estándares), la utilización de metadatos como los que corresponden a las etiquetas `<title>` y `<meta description>` y, por último, pero no menos importante, diseños que sean compatibles con la Web móvil (*responsive design*).

Por su parte, los criterios externos se refieren al resultado que pueda arrojar el análisis de los enlaces de entrada de la página considerada. En teoría, cada buscador tiene sus propios conjuntos de criterios y sus propias reglas para asignar pesos a cada criterio, pero, en general, las características de los enlaces que recibe una página suele ser uno de los más importantes, al menos a igualdad de los otros factores. Entre tales características podemos señalar: el número de enlaces de entrada procedentes de sitios web de calidad, el texto utilizado en el enlace de entrada o la relación temática de los enlaces de entrada con el tema (o temas) de la Web, entre otros.

Como un criterio fácil de fijar, podemos decir que un sitio web mejora su posicionamiento de cara a los motores de búsqueda si recibe muchos enlaces de entrada de sitios que, a su vez, son muy enlazados y que están temáticamente relacionados con el tema del sitio considerado.

En los últimos años, la cada vez mayor importancia de la web social ha provocado que también sean importantes las llamadas señales de usuario, como el número de twitts de una página, sus menciones en la web social o clics en iconos de "me gusta" en Facebook, entre otros.

El apartado de Analítica web sirve, entre otras, para señalar que existen barreras técnicas que impiden que un motor de búsqueda "entienda" un

sitio. La primera de las que hay que ser conscientes es que los motores de búsqueda no entienden los gráficos, ni los vídeos, ni las animaciones multimedia, hechas con flash o con otra tecnología alternativa. En general, no entiende nada que no sea textual, lo cual no es tan fácil como decir que nuestro sitio solamente debe tener texto, porque tampoco funciona bien. Al contrario, sin contenido multimedia es difícil que un sitio sea considerado una autoridad.

Otras barreras pueden proceder de la tecnología utilizada para codificar el sitio que impide que los motores de búsqueda puedan interpretarlo. Por ejemplo, el sitio puede utilizar muchos gráficos y poco texto, o bien utiliza técnicas obsoletas como marcos (frames), o utilizar de forma incorrecta tecnologías como JavaScript o Flash, que impiden que los motores de búsqueda puedan explorar de forma adecuada el sitio ya que carece de contenido textual.

Por otro lado, aunque no tiene un enorme impacto en el posicionamiento, los metadatos <title> y

<meta description> son esenciales porque forman parte del resumen que aparece en la página de resultados del buscador. Por tanto, su importancia es grande en cuanto a otros aspectos de SEO como la ratio de clics.

Para concluir esta reseña, y si me permiten un punto de vista algo más subjetivo, si tuviera que destacar, aparte de lo señalado, algunos apartados, en mi opinión, serían los dedicados al denominado Link Building y Link Baiting, por la extensión y el detalle que dedica a las posibilidades que puede tener un sitio web para su aplicación. También me han parecido especialmente acertados los apartados dedicados a cuestiones como el Knowledge Graph, así como los dedicados a microdatos, también porque aporta informaciones precisas y actualizadas en unos temas que, incluso los profesionales, suelen tener dudas.

Lluís Codina

UPF

lluis.codina@upf.edu