



Revista española de documentación científica

vol. 40, n. 3 (2017)

Estudios

Primera generación de autores con difusión internacional en la investigación turística española

Jesús Manuel López-Bonilla, Concepción Granados-Perea, Luis Miguel López-Bonilla

El uso de las licencias libres en los datos públicos abiertos

Luis Fernando Ramos-Simón

Estudio métrico de ALFIN en Iberoamérica: de la bibliometría a las altmetrics

Alejandro Uribe-Tirado, Joel Alhuay-Quispe

Determinación de grupos de usuarios de bibliotecas digitales mediante el análisis de ficheros log

Juan Antonio Martínez-Comeche

Características generales de las revistas científicas peruanas

Julio Santillán-Aldana, Mónica Arakaki, Aurora de la Vega, Mónica Calderón-Carranza, Josmel Pacheco-Mendoza

Influencia de la fecha de publicación online de los artículos científicos en los indicadores bibliométricos

Mercedes Echeverría, David Stuart, José Antonio Cordón-García

Modelos evaluativos de Metaliteracy y alfabetización en información como factores de excelencia académica

Miguel Ángel Marzal, Jussara Borges

ESTUDIOS / RESEARCH STUDIES

Primera generación de autores con difusión internacional en la investigación turística española

Jesús Manuel López-Bonilla*, Concepción Granados-Perea*, Luis Miguel López-Bonilla*

*Universidad de Sevilla

Correo-e: lopezbon@us.es | ORCID iD: <http://orcid.org/0000-0003-0623-1048>

Correo-e: conchagrana2@gmail.com | ORCID iD: <http://orcid.org/0000-0003-2466-2669>

Correo-e: luismi@us.es | ORCID iD: <http://orcid.org/0000-0002-5324-7715>

Recibido: 12-05-2016; 2ª versión: 19-08-2016; Aceptado: 12-11-2016

Cómo citar este artículo/Citation: López-Bonilla, J. M.; Granados-Perea, C.; López-Bonilla, L. M. (2017). Primera generación de autores con difusión internacional en la investigación turística española. *Revista Española de Documentación Científica*, 40(3): e178. doi: <http://dx.doi.org/10.3989/redc.2017.3.1399>

Resumen: Los investigadores españoles han mostrado un interés creciente en los análisis bibliométricos sobre la investigación turística. Sin embargo, no se ha realizado todavía un análisis de autoría suficientemente exhaustivo en la investigación turística española. Este estudio analiza la producción científica con difusión internacional de los autores afiliados a instituciones españolas. Para ello, la búsqueda se ha realizado a través de la base de datos de Scopus durante el período 2002-2013. Se ha establecido un ranking de 79 autores españoles que han publicado seis o más trabajos en revistas científicas con difusión internacional. La media de co-autores españoles por trabajo publicado es elevada respecto a los autores internacionales. Asimismo, se observa una clara desigualdad de género, con predominio de autores masculinos. En este ranking destacan por número de trabajos publicados las áreas de Economía Aplicada y de Comercialización e Investigación de Mercados, así como las universidades de Islas Baleares y de Alicante.

Palabras clave: Turismo; análisis de autoría; Scopus; producción científica; ranking.

First generation of Spanish authors to disseminate hospitality and tourism research internationally

Abstract: Spanish researchers have shown an increasing interest in the bibliometric analysis of tourism research. However, as of yet a sufficiently comprehensive authorship analysis on Spanish tourism research has not been performed. This study analyses the international scientific output of scholars in Spain. The search was performed using the Scopus database from 2002 to 2013. It established a ranking of 79 Spanish authors who have published six or more papers in international scientific journals. The average number of Spanish co-authors per paper is high with respect to international authors. Also, a clear gender inequality is observed, with male authors dominating. The areas of Economy and Marketing stand out for the total number of papers produced in them, as do the universities of Islas Baleares and Alicante for their output.

Keywords: Tourism; authorship analysis; Scopus; scientific output; ranking.

Copyright: © 2017 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

El sector turístico es una de las industrias más importantes del mundo en términos de Producto Interior Bruto y de número de empleados (Chang y McAleer, 2012). España lleva muchos años situándose entre los países principales del sector. Actualmente, España es el tercer país con mayor número de turistas internacionales que lo visitan y el segundo país en el que se obtienen mayores ingresos por este concepto. Sin embargo, la formación y la investigación española en turismo han estado más rezagadas de lo que deberían. En concreto, la formación universitaria en turismo se inicia con la aprobación del Real Decreto 259/1996, a partir del cual se crea la Diplomatura de Turismo. Se trataba de una titulación de tres años académicos que no tenía continuación con una titulación de licenciatura ni, por supuesto, con la posibilidad de realizar un doctorado en turismo. Como indican Ceballos y otros (2010), la Unión Europea creó el Espacio Europeo de Educación Superior con idea de armonizar los sistemas universitarios de los países miembros. La adaptación de España al Espacio Europeo de Educación Superior supone la aprobación de la Ley Orgánica 6/2001 de Universidades, que es modificada por la Ley Orgánica 4/2007, y que define la estructura de las enseñanzas universitarias en tres ciclos: Grado, Máster y Doctorado.

Como ya avanzaban López-Bonilla y López-Bonilla (2004) hace algo más de una década, la gran importancia que el sector turístico ha demostrado alcanzar en España no se ha correspondido con la atención prestada por la comunidad científica. Aunque ha habido un escaso desarrollo de la formación universitaria en turismo hasta hace relativamente poco tiempo, la investigación turística española empieza a consolidarse a partir del siglo XXI. Todos los investigadores españoles pertenecen a distintas áreas de conocimiento científico, pero ninguna es específica de turismo porque no existe esta categoría en España. Actualmente, muy pocos investigadores españoles postdoctorales han realizado los estudios universitarios de turismo. Afortunadamente, los estudiantes de turismo ya pueden cursar el Doctorado de Turismo desde 2008. Recientemente, a finales de 2013, se ha creado el primer Doctorado Interuniversitario en Turismo, que ha sido promovido por las Universidades de Alicante, Málaga, Rey Juan Carlos y Sevilla.

En general, la producción científica española ha crecido exponencialmente en las últimas décadas (Jiménez-Contreras y otros, 2003; Granadino y otros, 2006; Ibáñez y otros, 2013). Siguiendo a Moreno-Pulido y otros (2013), se puede constatar que el primer impulso para fomentar la producción científica española lo han dado la creación

de la Agencia Nacional de Evaluación y Prospectiva (ANEP) en 1986 y de la Comisión Nacional de Evaluación de la Actividad Investigadora (CNEAI) en 1989. En ambos casos se ha basado en la evaluación de publicaciones de prestigio y, especialmente, en el ámbito internacional. En cambio, un segundo impulso investigador lo ha provocado la creación de la Agencia Nacional de Evaluación de la Calidad y la Acreditación (ANECA) en 2002, que ha supuesto un verdadero fomento de la producción científica a través de la evaluación de las actividades docentes e investigadoras de los solicitantes para el acceso a las distintas categorías de profesorado universitario.

En los últimos años, los investigadores españoles han mostrado interés en los análisis bibliométricos sobre la investigación en el ámbito del turismo. Se han realizado diversos estudios que se han enfocado desde distintas disciplinas, tales como la geografía (v.gr., Antón y otros, 1996; Corral y San Eugenio, 2013; Corral y Cànoves, 2014), el marketing (v.gr., Bigné y otros, 2008; López-Bonilla y López-Bonilla, 2008), la economía (v.gr., Bote, 1996; Figuerola, 2000; Aguiló, 2010) o la gestión de empresas (v.gr., Camisón y Monfort, 1996; Camisón y otros, 1997; Sánchez y Marín, 2003; Ruiz y otros, 2006; Hernández y otros, 2011; Albacete y otros, 2013). Pero también algunos estudios se han planteado con un carácter general o multidisciplinar (v.gr., Albacete-Sáez y Fuentes-Fuentes, 2010; Vargas, 2011a, 2011b; Moreno y Picazo, 2012; Corral y Cànoves, 2013; Corral-Marfil y otros, 2015). La mayoría de estos trabajos se han basado en las publicaciones en revistas científicas, aunque hay otros que lo han hecho a través de las tesis doctorales (v.gr., Esteban, 2000; Ortega y Rodríguez, 2004; Agencia Valenciana de Turismo, 2006; López-Bonilla y López-Bonilla, 2008; Vargas, 2011b) o mediante las actas de congresos especializados en turismo (v.gr., Miralbell y otros, 2012; Corral-Marfil, 2013; Corral y San Eugenio, 2013).

Sin embargo, ninguno de estos trabajos se ha dedicado en profundidad al análisis de autorías sino que se ha tratado parcialmente, en algunos casos como un aspecto más del análisis de la investigación turística en general en España. Además, todos estos estudios se han publicado en revistas españolas que no están indexadas en bases de datos internacionales, salvo los trabajos de Albacete y otros (2013) y de Corral y San Eugenio (2013), aunque en ambos casos se centran en áreas específicas de la investigación turística como la economía de la empresa y la geografía, respectivamente.

Asimismo, en términos generales, cabe señalar que los trabajos que se ocupan parcialmente del análisis de autorías se limitan a un número redu-

cido de autores analizados, así como a un número y tipo de revistas estudiadas, que se circunscribe concretamente a revistas especializadas en turismo y ocio, o bien a revistas especializadas en la gestión de empresas. En la tabla I se recoge un resumen de los estudios bibliométricos publicados en revistas científicas sobre la investigación turística

en España, que incluyen de manera destacada un análisis de autorías. Tan solo uno de estos estudios citados se ha publicado en una revista indexada en *Web of Science (Journal Citation Reports)*, aunque dicho trabajo se ha ocupado especialmente de una disciplina concreta, como es la gestión de empresas (Albacete y otros, 2013).

Tabla I. Análisis de autorías sobre investigación turística española a través de publicaciones en revistas científicas

ESTUDIOS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Albacete-Sáez y Fuentes-Fuentes (2010)	1998-2009	31	31	Turismo y ocio	Internacional	45	Multidisciplinar
Hernández, Campón y Folgado (2011)	2001-2010	19	2	Gestión de Empresas	Nacional	14	Gestión de empresas
Moreno y Picazo (2012)	2006-2011	31	15	Turismo	Nacional e internacional	30	Multidisciplinar
Albacete, Fuentes y Haro-Domínguez (2013)	1997-2011	26	26	Turismo y ocio	Internacional	18	Gestión de empresas

(1) Período de estudio; (2) Nº de revistas totales analizadas; (3) Nº de revistas en Scopus; (4) Tipos de revistas; (5) Origen de publicaciones; (6) Nº de autores analizados; (7) Áreas de estudio.

En la investigación turística internacional parece que los estudios bibliométricos centrados en las autorías tampoco han tenido un gran desarrollo hasta bien entrado este siglo XXI. Así, por ejemplo, como sugieren Zhao y Ritchie (2007), el análisis de autoría ha sido una práctica común en otras disciplinas maduras, mientras que es un fenómeno mucho más reciente en la investigación turística.

Por lo tanto, el objetivo principal del presente estudio se basa en la identificación de los principales investigadores españoles en el ámbito turístico y que hayan estado afiliados a una institución española. Para ello, se va a analizar la producción científica de estos investigadores en prestigiosas revistas internacionales durante los últimos doce años. En este contexto, se pretende detectar el perfil de los investigadores más prolíficos a través de las revistas de prestigio donde han publicado, la afiliación de estos autores, las zonas geográficas, el número de autores por publicación y el género. En definitiva, con el presente estudio se trata de cubrir este gap en la literatura académica española y, de este modo, fortalecer la difusión internacional de los investigadores españoles en el ámbito turístico. Además, puede representar una mayor contribución a la investigación académica tanto en el ámbito del estudio turístico como en el ámbito de la información científica. La posición de España en el sector turístico mundial, así como la importancia de la investigación turística española, suponen un motivo extraordinario para desarrollarlo en el presente trabajo. De este modo, se pretende

aportar un mayor conocimiento sobre este tópico a la comunidad científica nacional e internacional en el ámbito turístico. Los hallazgos pueden ser útiles para instituciones, investigadores y estudiantes de doctorado y master.

2. METODOLOGÍA

El presente estudio se realiza desde un enfoque cuantitativo bajo las perspectivas multidisciplinar, longitudinal e internacional. La metodología de este estudio hay que plantearla desde distintos aspectos, como son los siguientes: (1) la identificación del concepto estudiado; (2) la selección de revistas y de trabajos publicados, (3) la selección de la búsqueda de palabras claves; (4) el período de tiempo analizado; y (5) la medida de las autorías. En primer lugar, es necesario definir el concepto que va a ser investigado en el análisis bibliométrico. Este concepto de búsqueda se corresponde con los trabajos realizados sobre la investigación turística por autores afiliados a alguna institución española, que han sido publicados en prestigiosas revistas internacionales y durante un período de 12 años que va desde 2002 a 2013. De este modo, se han seleccionado autores que estaban adscritos a instituciones españolas, por lo que se han excluido aquellas organizaciones que están situadas en España pero que no son de nacionalidad española, como ocurre, por ejemplo, con la Organización Mundial del Turismo, con sede en Madrid. Hay que precisar también que la investigación turística internacional abarca las publicaciones realizadas en

torno a los términos anglosajones de *hospitality* y *tourism*, que en este contexto se entienden separados pero que en España tienen un mismo significado, es decir, se aglutinan ambos términos dentro del concepto de turismo.

La selección de revistas ha sido un aspecto controvertido en la literatura académica. Las revistas de prestigio se pueden definir como aquellas que están indexadas en grandes bases de datos internacionales y que disponen de un índice de impacto reconocido por la comunidad científica internacional. Asimismo, solo se consideran las publicaciones en revistas científicas porque gozan de un reconocimiento académico más común y extendido en todas las disciplinas o campos de estudio. Actualmente, las dos bases de datos más relevantes y más utilizadas en los estudios bibliométricos son *ISI Web of Knowledge (WoS)*, de la empresa *Thomson Reuters*, y *Scopus*, del grupo *Elsevier*. Como indica Ardanuy (2014), estas bases de datos bibliográficas internacionales son frecuentemente utilizadas para analizar la producción científica y establecer indicadores bibliométricos.

Además, se ha creído conveniente analizar en este trabajo todas las publicaciones incluidas en Scopus y no solo las revistas especializadas en turismo. Se ha considerado que utilizar solo unas pocas revistas de prestigio especializadas en turismo puede resultar una limitación importante en este tipo de estudios bibliométricos. Esto es debido a dos razones principales. Por una parte, los investigadores españoles en ciencias sociales y, especialmente en el ámbito turístico, llevan relativamente poco tiempo publicando en revistas internacionales. Y, por otra parte, las revistas centradas en investigación turística han crecido en importancia en los últimos años. Muchas de estas revistas se han incorporado recientemente a las bases de datos WoS y Scopus.

Como se ha indicado, el presente estudio se ha basado en una doble revisión de revistas especializadas y no especializadas en investigación turística. En primer lugar, se ha realizado una revisión exhaustiva de todas las revistas especializadas netamente en investigación turística que están incluidas en la categoría de "Tourism, Leisure and Hospitality Management" de Scopus. También se han incluido otras revistas especializadas que se consideran importantes, pero que aparecen en otras categorías distintas, tales como *Journal of Travel Research* y *Anatolia: An International Journal of Tourism and Hospitality Research*. En segundo término, se han revisado también de manera exhaustiva todas las revistas científicas más específicas sobre la investigación del ocio, depurando aquellos trabajos publicados que no se co-

rrespondían con el ámbito turístico. Y, en tercer lugar, se ha realizado una revisión del resto de trabajos publicados en revistas científicas incluidas en Scopus, pero que no eran consideradas netamente especializadas en el ámbito turístico y del ocio. En total han sido 704 revistas científicas de las que se han obtenido, al menos, una publicación por autores de instituciones españolas.

Para seleccionar los trabajos publicados en revistas no especializadas en turismo que están incluidas en Scopus, se ha optado por la búsqueda de cuatro palabras clave que se identifiquen claramente con la investigación turística y que eviten confusiones con otros ámbitos de estudio al estar bien diferenciadas. Así, se han elegido dos palabras genéricas, como son "tourism" y "tourist", que abarcan la gran mayoría de trabajos publicados, y otras dos palabras clave específicas relacionadas con las principales empresas turísticas del sector, como son "hotel" y "travel agency". Esta búsqueda se ha realizado a través de los títulos de artículos, los resúmenes y las palabras claves (*title, abstracts and keywords*) de Scopus. Además, se ha realizado una última búsqueda complementaria a la anterior que se ha basado en buscar un amplio conjunto de palabras clave para encontrar nuevos trabajos de los autores prolíficos ya detectados, o bien, para hallar nuevos autores prolíficos, en su caso. Para ello, en concreto, se ha utilizado la totalidad de palabras clave que se recoge en el trabajo de Corral y Cànoves (2013), que suman la cincuentena de conceptos de búsqueda, y que se ha basado a su vez en el trabajo de González-Albo y otros (2009). La lista de terminología de búsqueda es la siguiente: accommodation*, agritour*, aparthotel*, boarding hous*, bed and breakfast*, booking*, camping*, caravan*, computer reserve system*, cruises*, cycletour*, destination, image, destination marketing, destination branding, destination positioning, ecotourism*, excursio*, geotourism*, holiday*, hospitalit*, hostel*, hosteler*, *hotel*, inns, journey*, laketourism*, lodge, lodging hous*, lodging*, motel*, overbook*, overnight stay*, parador*, pilgrim*, resort*, restaurant*, roadhouse*, rooming hous*, salute per aqua, sightsee*, souvenir*, spa, timeshare*, tour*, tourism*, tours, travel*, trip*, vacation*, weekend stay*.

Por otra parte, en el proceso de búsqueda se han incluido exclusivamente aquellos trabajos que aparecen como publicaciones en revistas (identificadas en Scopus con el término de "journals") y que hayan sido objeto de revisión científica, es decir, se han eliminado aquellos documentos que hayan sido considerados como editoriales, libros, capítulos de libros y reseñas de libros. Asimismo, se ha realizado una exhaustiva labor de depuración de

los trabajos publicados. Sobre todo, se ha realizado una revisión muy detallada de las firmas de los autores, ya que muchos de ellos aparecen con distintas firmas y ha sido necesario reagruparlas para cada autor.

En cuanto al ámbito temporal, el presente estudio se ha realizado durante un período de doce años, abarcando desde 2002 hasta 2013. Se ha elegido esta duración temporal por dos cuestiones fundamentales: (1) debido a que 2002 es el primer año completo tras la aprobación de la nueva Ley de universidades (Ley 6/2001, de 21 de diciembre de 2001); y (2) dado que se ha asimilado el período de estudio a dos sexenios de investigación, que es el complemento de productividad más importante de las universidades españolas, que acredita la capacidad y experiencia investigadora del personal docente e investigador, además de ser una condición necesaria para poder ser director de tesis doctorales conforme a la normativa vigente.

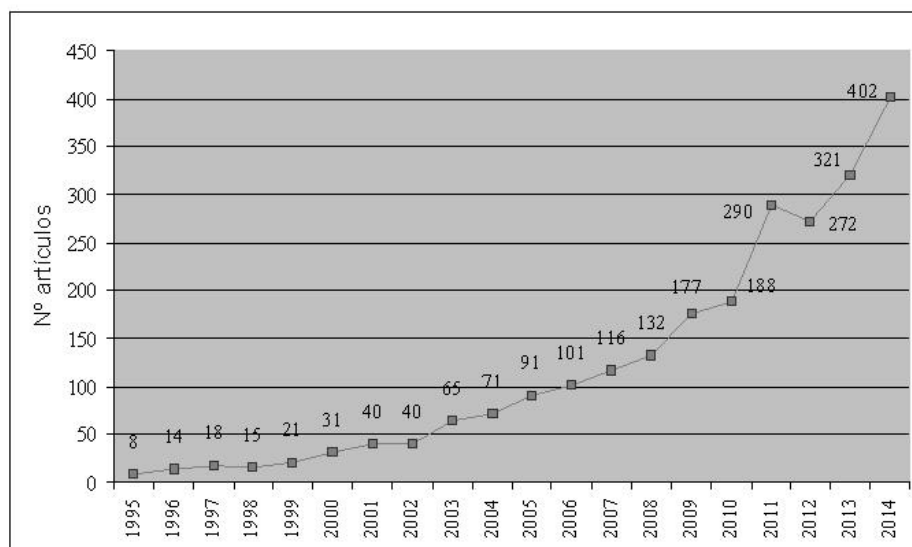
Por último, para la selección de autores más prolíficos, se había pensado en elegir inicialmente a todos aquellos autores que hubieran publicado cinco trabajos en prestigiosas revistas internacionales por cada período de seis años. Se consideraba esta cifra porque significaría el número de trabajos a completar hipotéticamente en un sexenio de investigación como mérito a alcanzar por los investigadores españoles. Sin embargo, esta cifra de diez trabajos durante los doce años analizados aportó un número de autores muy reducido. Una serie de trabajos relevantes sobre análisis de autorías en investigación turística a nivel internacional se ha basado en la selección de un número de autores que supera ligeramente el medio centenar (v.gr.,

Zhao y Ritchie, 2007; McKercher, 2008; Park y otros, 2011). Por ello, se ha establecido el punto de corte en seis trabajos publicados, coincidiendo con Albacete y otros (2013). Así, los autores prolíficos identificados en el presente estudio deben haber publicado un mínimo de seis trabajos en prestigiosas revistas internacionales. Esto significa publicar un trabajo por cada dos años, lo cual parece una cifra razonable para que un autor, al menos, pueda empezar a ser considerado prolífico en un área de investigación específica que está relativamente menos desarrollada que otras en el ámbito internacional. Además, en el análisis se ha utilizado un doble sistema de recuento de trabajos mediante el recuento total de artículos de cada autor y el recuento fraccionado, según el número de colaboraciones por artículo.

3. RESULTADOS

Los primeros resultados generales se presentan en la figura 1, que recoge el número de trabajos de autores afiliados a instituciones españolas y que han publicado en revistas indexadas en Scopus. Los datos aparecen desde 1995 debido al escaso número de publicaciones de los años anteriores. La primera publicación española en el ámbito de la investigación turística que aparece en Scopus data de 1977 y está firmada por Miguel Ángel García Dory, aunque no aparece su afiliación. Su trabajo se ocupa de los problemas de gestión y conservación del Parque Nacional de Covadonga (Asturias) y está publicado en *Conservation Biological*. Desde esa fecha, el número de publicaciones españolas sobre turismo ha sido muy escaso anualmente hasta 1996, en el que se supera la decena de artículos.

Figura 1. Evolución de la producción científica española en el ámbito turístico con difusión internacional



Como se observa en la figura 1, la producción científica española sobre turismo con difusión internacional comienza su verdadero despegue con el inicio del siglo XXI. Así, el mayor aumento relativo interanual de trabajos publicados se ha dado entre los años 2002 y 2003, pasando de 40 a 65 trabajos, que supone un aumento del 62,5%. Precisamente, estas fechas son coincidentes con la entrada en vigor de la Ley de Universidades a finales de 2001. El crecimiento anual de publicaciones ha mantenido una tendencia ascendente, acentuándose en los últimos años, en el que se destaca el fuerte incremento de 2011 al pasar de 188 a 290 trabajos publicados, que supone un 54,3% de aumento interanual. La tasa media de aumento interanual desde 1995 hasta 2014 ha sido del 25%.

En el presente estudio se han recogido un total de 1864 publicaciones en revistas científicas indexadas en Scopus que cumplen con los requisitos previstos durante el período 2002-2013. Pero hay que precisar que 1380 artículos se han publicado en el segundo sexenio (2008-2013), lo que representa un 74% de todo el período de 12 años analizado. Respecto a la cantidad total, se han publicado 576 trabajos en revistas especializadas en el ámbito turístico, lo cual supone alrededor del 31% del total de trabajos publicados en materia turística. Sin embargo, las publicaciones en revistas especializadas sobre turismo han descendido en el segundo período de estudio, alcanzando el 35,5% entre 2002-2007 y pasando al 29,5% entre 2008-2013. En la tabla II se enumeran las revistas científicas en las que los autores españoles han publicado más artículos. Casi todas las revistas aumentan su número de publicaciones de autores españoles en el segundo sexenio, salvo algunas excepciones, como son: *Estudios Geográficos*, que publica nueve artículos; *Desalination*, con seis artículos; *Anatolia* y *Tourism Geographics*, que publican solo cuatro artículos cada una de ellas; y, especialmente, *Annals of Tourism Research*, con 20 artículos publicados en este período. Otra revista relevante en turismo, *Journal of Sustainable Tourism*, publica la mitad de trabajos en cada período de seis años.

En la tabla II aparecen las 30 revistas científicas principales que difunden la investigación turística de los autores españoles. En este listado se incluyen 16 revistas especializadas en turismo (señaladas con asterisco) y otras 14 revistas no especializadas en este ámbito de investigación. Las primeras posiciones del ranking están ocupadas por tres revistas especializadas en turismo, siendo dos de ellas las que poseen actualmente un mayor prestigio internacional, contando con los más altos factores de impacto en su ámbito, como son *Tou-*

rism Management (con 128 artículos publicados) y *Annals of Tourism Research* (con 50 artículos). Asimismo, hay 21 revistas en esta lista que están indexadas en JCR, perteneciendo las ocho primeras a esta base de datos. Además, se hallan siete revistas españolas, aunque tan solo dos de ellas forman parte del JCR, con unos factores de impacto reducidos: *Boletín de la Asociación de Geógrafos Españoles* y *Scripta Nova*. Curiosamente, se puede observar también que la mayoría de revistas españolas están impulsadas por áreas de conocimiento de Geografía.

Se podría hacer un análisis previo de algunos datos relevantes de todos de los artículos publicados por autores españoles para obtener una primera visión en su conjunto de la difusión internacional que ha conseguido la investigación turística en los últimos años. Para ello, a continuación, se van a considerar aspectos como la procedencia, temática y áreas de estudio, co-autoría y colaboraciones nacionales e internacionales, así como se tendrá en cuenta tanto el desglose de revistas especializadas en turismo frente a otras revistas como los efectos producidos por la evolución en el tiempo que merezcan destacarse.

La base de datos Scopus recoge las grandes áreas de investigación en las que más publican los autores españoles sobre turismo entre 2002 y 2013, destacando las Ciencias Sociales, con un total de 859 artículos; Negocios, Gerencia y Contabilidad, con 746 artículos; y Ciencias Medioambientales, con 390 artículos. Las tres cuartas partes de estas cifras se corresponden con el período 2008-2013. Cada artículo suele clasificarse en más de una de estas grandes áreas en Scopus. Este mismo predominio de estas tres áreas se mantiene de manera muy similar en los dos sexenios analizados, aunque en el segundo período de 2008-2013 se ha producido un incremento importante de dos grandes áreas, como son, por un lado, Economía y Finanzas, que asciende a la cuarta posición en este ranking de áreas con 119 artículos y, por otro, Artes y Humanidades, que ha tenido el mayor aumento porcentual de todas las áreas, al pasar de solo 6 artículos en el primer período a 80 artículos en el segundo.

Las instituciones de las que proceden los autores españoles son mayoritariamente universitarias y públicas. Esto se confirma por las diez primeras instituciones que se citan seguidamente: Universidad Islas Baleares (139 artículos), Universidad de Alicante (128), Universidad de Valencia (92), Universidad de Las Palmas de Gran Canaria (78), Universidad de Málaga (75), Universidad Autónoma de Barcelona (75), Universidad de Barcelona (67), Universidad Autónoma de Madrid (65), Universidad

Tabla II. Revistas científicas donde publican los autores españoles en la investigación turística con difusión internacional

Orden	TÍTULO DE REVISTA	PAÍS DE ORIGEN	FACTOR IMPACTO 2014	Nº ART.
1	Tourism Management*	Reino Unido	JCR=2,554	128
2	Tourism Economics*	Reino Unido	JCR=0,515	110
3	Annals of Tourism Research*	Reino Unido	JCR=2,685	50
4	Boletín de la Asociación de Geógrafos Españoles	España	JCR=0,103	44
5	Service Industries Journal	Reino Unido	JCR=0,832	34
6	International Journal of Hospitality Management*	Reino Unido	JCR=1,939	29
7	Journal of Coastal Research	USA	JCR=0,980	25
8	Journal of Travel Research*	Reino Unido	JCR=2,442	24
9	Estudios Geográficos	España	SJR=0,112	21
9	WIT Transactions on Ecology and the Environment	Reino Unido	SJR=0,154	21
11	International Journal of Tourism Research*	Reino Unido	JCR=1,314	19
12	Document D'Anàlisi Geogràfica	España	SJR=0,363	18
13	Journal of Travel and Tourism Marketing*	Reino Unido	JCR=0,736	17
14	Journal of Sustainable Tourism*	Reino Unido	JCR=1,959	16
15	International Journal of Contemporary Hospitality Management*	Reino Unido	JCR=1,409	15
15	Revista Galega de Economía	España	SJR=0,157	15
17	Cornell Hospitality Quarterly*	USA	JCR=1,746	14
18	Desalination	Holanda	JCR=3,756	13
19	Current Issues in Tourism*	Reino Unido	JCR=0,918	12
19	European Journal of Tourism Research*	Bulgaria	SJR=0,245	12
19	Investigaciones Europeas de Dirección y Economía de la Empresa	España	SJR=0,106	12
19	Ocean & Coastal Management	Reino Unido	JCR=1,748	12
19	Tourismos*	Grecia	SJR=0,240	12
24	Anatolia: An International Journal of Tourism and Hospitality Research*	Turquía	SJR=0,388	11
24	Cuadernos de Turismo*	España	SJR=0,130	11
24	Environmental Engineering and Management Journal	Rumania	JCR=1,065	11
24	Expert Systems with Applications	Reino Unido	JCR=2,240	11
24	Scripta Nova	España	JCR=0,195	11
29	Environmental Management	USA	JCR=1,724	10
29	Tourism Geographics*	Reino Unido	JCR=1,695	10

Se señalan con * las revistas especializadas en turismo.

de Sevilla (59) y Universidad Complutense (58). La primera institución distinta a las universidades públicas es el Museo Nacional de Ciencias Naturales, perteneciente al Consejo Superior de Investigaciones Científicas (CSIC), con 18 artículos. En el primer sexenio (2002-2007) aparecerían entre las diez primeras instituciones las universidades de Murcia, La Laguna y Politécnica de Valencia, que ocupaban las posiciones de las tres últimas universidades citadas anteriormente.

En general, la gran mayoría de artículos se ha publicado por un número comprendido entre uno

y cuatro autores, alcanzando un porcentaje del 83,1% entre 2002-2007 y del 85,9% entre 2008-2013. Esto supone que en torno al 15% de trabajos publicados superan las cuatro colaboraciones. El mayor porcentaje de co-autoría está representado por dos autores en los primeros seis años (con un 26,4% sobre el total de este período), mientras que las colaboraciones de tres autores son más destacadas en el segundo período estudiado (con un 30,2%). A pesar de ello, el valor de la mediana es de tres autores en ambos sexenios. En concreto, la co-autoría por cada publicación se mantiene en

una media similar en torno a tres autores durante los doce años analizados. De este modo, el número medio de autores por artículo fue de 3,04 en el primer período, mientras que aumentó ligeramente hasta 3,26 en el segundo período estudiado. Sin embargo, estas cifras de co-autorías son inferiores en las revistas especializadas en turismo, alcanzando 2,39 autores por artículo publicado durante 2002-2007 y 2,54 autores durante 2008-2013.

Las colaboraciones entre autores de distintas áreas de conocimiento, distintas instituciones o distintas nacionalidades, son relativamente escasas, aunque han aumentado en gran medida en el último sexenio. Para hacerse una idea de este progreso, el país con el que los autores españoles tuvieron más colaboraciones entre 2002 y 2007 fue Italia, con 15 participaciones de autores italianos. En su conjunto, los autores españoles han colaborado con autores extranjeros de 78 países, siendo los más destacados los autores británicos (con 92 colaboraciones de estos autores), estadounidenses (con 70 colaboraciones), italianos (con 66 colaboraciones), alemanes y mexicanos (con 55 colaboraciones por cada nacionalidad), portugueses (con 29 colaboraciones) y franceses (con 27 colaboraciones). Entre las instituciones extranjeras de procedencia con las que han participado los autores españoles, se pueden destacar la Universidad de Bolonia y la Organización Mundial del Turismo, en ambos casos con ocho colaboraciones cada una de ellas; la Universidad Nacional Autónoma de México, con siete colaboraciones; la Universidad Michoacana de San Nicolás de Hidalgo (México), con seis colaboraciones; y la Universidad del Estudio de Florencia y la Liverpool John Moores University, con cinco colaboraciones en ambos casos.

La tabla III detalla una lista de 79 autores adscritos a universidades españolas que tienen, al menos, seis trabajos publicados en la base de Scopus durante el doble sexenio entre 2002 y 2013 y que están relacionados con el ámbito de la investigación turística. Todos estos autores suman un total de 285 publicaciones en revistas indexadas en Scopus. Estos autores representan a 28 instituciones, estando integradas por un centro de investigación del CSIC, una universidad privada y 27 universidades públicas. Uno de los autores aparece adscrito a dos instituciones.

En la tabla III se recoge, además de los nombres de los investigadores, su afiliación académica e institucional, el número de trabajos totales que han publicado en el período de estudio, los años de antigüedad de sus publicaciones y el número de trabajos fraccionados en función del número de autores de cada publicación realizada. Es decir, si un autor ha publicado un trabajo conjuntamente con otro investigador supone que ambos han realizado 0,5 trabajos en este caso. Por su parte, la antigüedad se mide con la media ponderada de los artículos publicados en los doce últimos años, teniendo en cuenta que si se publica en 2013 significa un año de antigüedad, mientras que si se publica en 2002 serían 12 años. Asimismo, se han incluido dos clasificaciones en las columnas extremas, que representan un orden general en relación al número de trabajos totales publicados por cada autor y un orden fraccionado en función del número de co-autorías o trabajos fraccionados de cada autor. En el orden total se ha considerado el número de trabajos fraccionados para desempatar, mientras que se ha hecho a la inversa en el orden fraccionado.

Tabla III. Producción científica en investigación turística de los autores más productivos afiliados a universidades españolas con difusión internacional (2002-13)

Orden total	Autores	Afiliación de autores		Nº trabajos totales	Años Antigüedad	Nº trabajos fraccionados	Orden fracc.
		Institución	Área de conocimiento				
1	J.L. Nicolau	Universidad de Alicante	Comercialización e Invest.Mdos.	27	3,96	17,76	1
2	E. Claver-Cortés	Universidad de Alicante	Organización de Empresas	26	5,04	8,05	3
3	J. Rosselló	Universidad Islas Baleares	Economía Aplicada	24	4,67	10,30	2
4	I. Gil Saura	Universidad de Valencia	Comercialización e Invest.Mdos.	21	3,57	7,75	4
5	D.M. Frías Jamilena	Universidad de Granada	Comercialización e Invest.Mdos.	21	2,67	6,77	7
6	J. Pereira-Moliner	Universidad de Alicante	Organización de Empresas	20	4,55	6,15	9
7	J.F. Molina-Azorín	Universidad de Alicante	Organización de Empresas	20	4,60	5,80	10

Orden total	Autores	Afilación de autores		Nº trabajos totales	Años Antigüedad	Nº trabajos fraccionados	Orden fracc.
		Institución	Área de conocimiento				
8	J. Alegre	Universidad Islas Baleares	Economía Aplicada	19	5,11	7,69	5
9	M.M. Alonso-Almeida	Univ. Autónoma de Madrid	Organización de Empresas	19	1,89	7,36	6
10	M.J. Such Devesa	Universidad de Alcalá de Henares	Economía Aplicada	16	4,13	6,40	8
11	M.A. Rodríguez-Molina	Universidad de Granada	Comercialización e Invest.Mdos.	15	2,87	4,59	14
12	J. Rey-Maqueira	Universidad Islas Baleares	Economía Aplicada	15	5,07	4,02	22
13	L.A. Gil-Alaña	Universidad de Navarra	Métodos Cuantit. Economía y Empr.	12	6,25	4,96	11
14	J.A. Fraiz Brea	Universidad de Vigo	Comercialización e Invest.Mdos.	12	2,83	4,36	16
15	S. Moreno-Gil	Univ. Las Palmas de Gran Canaria	Comercialización e Invest.Mdos.	11	5,18	4,31	18
16	M.E. Ruiz-Molina	Universidad de Valencia	Comercialización e Invest.Mdos.	11	3,55	3,90	24
16	A. Riera-Font	Universidad Islas Baleares	Economía Aplicada	11	7	3,90	24
18	A.I. Polo-Peña	Universidad de Granada	Comercialización e Invest.Mdos.	11	2	3,78	27
18	T. López-Guzmán	Universidad de Córdoba	Economía Aplicada	11	2,82	3,78	27
20	G. Cànoves Valiente	Univ. Autónoma de Barcelona	Análisis Geogr. Regional	10	4,30	4,35	17
21	I. Rodríguez del Bosque	Universidad de Cantabria	Comercialización e Invest.Mdos.	10	4,30	4,17	21
22	A. Molina Collado	Univ. de Castilla-La Mancha	Comercialización e Invest.Mdos.	10	2,80	3,57	30
23	J.J. Tari Guillo	Universidad de Alicante	Organización de Empresas	10	3,90	2,86	42
24	L. Parte Esteban	UNED	Economía Fra. y Contabilidad	10	3,90	2,25	58
25	B. Plaza	Universidad del País Vasco	Economía Aplicada	9	3,89	4,76	12
25	T.F. Espino-Rodríguez	Univ. Las Palmas de Gran Canaria	Organización de Empresas	9	7	4,76	12
27	A. Mantecón	Universidad de Alicante	Sociología	9	3,44	4,50	15
28	J.M. López-Bonilla	Universidad de Sevilla	Comercialización e Invest.Mdos.	9	3,33	4,26	19
28	L.M. López-Bonilla	Universidad de Sevilla	Comercialización e Invest.Mdos.	9	3,33	4,26	19
30	J.I. Pulido-Fernández	Universidad de Jaén	Economía Aplicada	9	2,78	3,86	26
31	C.J. León González	Univ. Las Palmas de Gran Canaria	Economía Aplicada	9	6	3,52	34
32	E. Aguiló	Universidad Islas Baleares	Economía Aplicada	9	7,33	3,13	40
33	J.M. Peiró	Universidad de Valencia	Psicología Social	9	7,33	2,79	43
34	E. Bigné-Alcañiz	Universidad de Valencia	Comercialización e Invest.Mdos.	9	6,55	2,77	44
35	J.A. Castañeda-García	Universidad de Granada	Comercialización e Invest.Mdos.	9	3,67	2,70	46
36	J. Sánchez-García	Universidad Jaume I	Comercialización e Invest.Mdos.	9	4,33	2,45	49

Orden total	Autores	Afilación de autores		Nº trabajos totales	Años Antigüedad	Nº trabajos fraccionados	Orden fracc.
		Institución	Área de conocimiento				
37	V. Ramos	Universidad Islas Baleares	Economía Aplicada	9	5,33	2,25	59
38	J.A. Jiménez	Univ. Politécnica de Cataluña	Ingeniería Hidráulica	9	7,78	1,84	71
39	J.I. Castillo-Manzano	Universidad de Sevilla	Economía Aplicada	8	2,50	4,00	23
40	H. San Martín	Universidad de Cantabria	Comercialización e Invest.Mdos.	8	3,87	3,56	31
40	J.B. Garau-Vadell	Universidad Islas Baleares	Comercialización e Invest.Mdos.	8	3,88	3,56	31
40	B. Ortega Aguaza	Universidad de Málaga	Economía Aplicada	8	4,56	3,56	31
43	F. Orfila-Sintes	Universidad Islas Baleares	Organización de Empresas	8	5,50	3,42	35
44	D.R. Medina-Muñoz	Univ. Las Palmas de Gran Canaria	Organización de Empresas	8	4,88	3,20	38
44	R.D. Medina-Muñoz	Univ. Las Palmas de Gran Canaria	Organización de Empresas	8	4,88	3,20	38
46	L. Andreu	Universidad de Valencia	Comercialización e Invest.Mdos.	8	7,38	2,29	57
47	R. Rigall-I-Torrent	Universidad de Girona	Fund. Análisis Económico	8	3,50	2,06	64
48	M.F. Schmitz	Univ. Complutense de Madrid	Ecología	8	5,25	1,73	73
49	R. Huete	Universidad de Alicante	Sociología	7	4,14	3,76	29
50	E. Ruiz-Ballesteros	Universidad Pablo de Olavide	Antropología	7	3,71	3,27	36
50	M. Sánchez-Rivero	Universidad de Extremadura	Métodos Cuantitat. Economía y Empr.	7	2,57	3,27	36
52	F. Vera-Rebollo	Universidad de Alicante	Análisis Geográfico Regional	7	4,71	2,88	41
53	L. Pou Garcías	Universidad Islas Baleares	Economía Aplicada	7	5,29	2,72	45
54	M. Santana-Gallego	Universidad Islas Baleares	Economía Aplicada	7	2,86	2,58	47
55	J.A. Campos-Soria	Universidad de Málaga	Economía Aplicada	7	4,86	2,45	50
55	A. Esteban-Talaya	Univ. de Castilla-La Mancha	Comercialización e Invest.Mdos.	7	4,71	2,45	50
57	J. Uche-Marcuello	Universidad de Zaragoza	Máquinas y Motores Térmicos	7	4	2,33	56
58	V. Martínez-Tur	Universidad de Valencia	Psicología Social	7	6,83	1,69	74
59	G. Anfuso	Universidad de Cádiz	Geología	7	3,57	1,58	77
60	F. Díaz Pineda	Univ. Complutense de Madrid	Ecología	7	5,43	1,48	78
61	M. Seric	Universidad de Valencia	Comercialización e Invest.Mdos.	6	2	2,58	48
62	M. Cladera	Universidad Islas Baleares	Economía Aplicada	6	3,83	2,40	52

Orden total	Autores	Afiliación de autores		Nº trabajos totales	Años Antigüedad	Nº trabajos fraccionados	Orden fracc.
		Institución	Área de conocimiento				
62	J.M. Hernández	Univ. Las Palmas de Gran Canaria	Métodos Cuantit. Economía y Empr.	6	3,67	2,40	52
62	F. Martínez-Roget	Univ. de Santiago de Compostela	Economía Aplicada	6	3,50	2,40	52
62	X.A. Rodríguez González	Univ. de Santiago de Compostela	Economía Aplicada	6	3,67	2,40	52
66	S. Sánchez-Cañizares	Universidad de Córdoba	Organización de Empresas	6	3,50	2,25	60
67	J.E. Araña Padilla	Univ. Las Palmas de Gran Canaria	Economía Aplicada	6	5,17	2,12	61
67	M. Gómez Rodríguez	Univ. de Castilla-La Mancha	Comercialización e Invest.Mdos.	6	2,67	2,12	61
67	J.M. Rodríguez Antón	Univ. Autónoma de Madrid	Organización de Empresas	6	2,33	2,12	61
70	J. Cuñado Eizaguirre	Universidad de Navarra	Economía Aplicada	6	7,83	2,00	65
70	J. Lozano	Universidad Islas Baleares	Economía Aplicada	6	4,67	2,00	65
70	B. Moliner-Velázquez	Universidad de Valencia	Comercialización e Invest.Mdos.	6	3,50	2,00	65
70	F. Pérez de Gracia	Universidad de Navarra	Economía Aplicada	6	7,83	2,00	65
74	A. García-Pozo	Universidad de Málaga	Economía Aplicada	6	2,17	1,89	69
74	J.L. Sánchez-Ollero	Universidad de Málaga	Economía Aplicada	6	2,17	1,89	69
76	M.D. López-Gamero	Universidad de Alicante	Organización de Empresas	6	4	1,80	72
77	J.M. Calaforra	Universidad de Almería	Geodinámica	6	8,83	1,63	75
77	A. Fernández-Cortés	Univ. Almería y CSIC (Museo C. Naturales)	Geología	6	8,83	1,63	75
79	J. Ramos	Universidad de Valencia	Psicología Social	6	5,83	1,44	79

Los dos ranking de la tabla III están encabezados por Juan Luis Nicolau, que tiene 27 trabajos publicados en total, pero que reafirma sus diferencias con el resto de autores en cuanto al número de trabajos fraccionados, al disponer de una cifra de 17,8 artículos. Otros autores prolíficos son E. Claver-Cortés, con 26 trabajos publicados totales y 8,1 trabajos fraccionados, y J. Rosselló, con 24 trabajos totales y 10,3 trabajos fraccionados. Hay que subrayar que los diez primeros autores del ranking presentan una media de cuatro años de antigüedad respecto a sus publicaciones, cuya cifra está por debajo de la media de antigüedad general. Preci-

samente, en estas posiciones más avanzadas se encuentra M.M. Alonso-Almeida, que es quien tiene el menor ratio de antigüedad de todos los autores prolíficos analizados, y le siguen a gran distancia D.M. Frías JAMILENA y J.L. Nicolau.

Por otro lado, se puede observar que la mayoría de investigadores que ocupan las primeras posiciones por el número de trabajos totales publicados mantienen también posiciones similares de acuerdo al número de trabajos fraccionados. En este sentido, cabe mencionar algunos autores que han mejorado sus posiciones para situarse en el pri-

mer cuartil de este ranking general, considerando el número de trabajos fraccionados, como son B. Plaza y T.F. Espino-Rodríguez, que alcanzan la posición 12; A. Mantecón, que se sitúa en el puesto 15; J.M. López-Bonilla y L.M. López-Bonilla, que ascienden a la posición 19; y J.I. Castillo-Manzano, que alcanza la posición 23. Otros investigadores que están más atrás en la clasificación, pero que han conseguido avances de más de diez posiciones según el número de trabajos fraccionados, son los siguientes: R. Huete, M. Sánchez-Rivero, E. Ruiz-Ballesteros, F. Vera-Rebollo y M. Seric. Entre todos estos autores, merece resaltar la mejora considerable de R. Huete, avanzando 20 posiciones en el ranking y ocupando finalmente el puesto 29 en la clasificación fraccionada.

A continuación, se van a considerar en el presente análisis otras tres cuestiones relevantes, como son el género de las autorías, el número de co-autorías y la participación en doctorados de turismo. En lo que respecta al género de los autores prolíficos, se encuentran 22 mujeres y 57 hombres, con lo cual hay una representación femenina del 27,8% del total. Sin embargo, hay cuatro investigadoras entre los diez primeros autores prolíficos, como son I. Gil Saura, M.D. Frías Jamilena, M.M. Alonso-Almeida y M.J. Such Devesa. Además, las mujeres tienen un ratio de antigüedad de sus publicaciones (4,2) inferior al de los hombres (4,6). En cuanto a la cifra de co-autorías de los trabajos publicados, se ha obtenido una media de 2,86 autores por artículo, siendo J.L. Nicolau quien ostenta el promedio más reducido con 1,52 autores por trabajo, seguido de R. Huete, con una media de 1,86 autores y B. Plaza (1,89 autores). A su vez, la media de las autoras (2,72) es inferior a la de los hombres (2,97). Por último, se hallan ocho autores prolíficos que pertenecen al Doctorado Interuniversitario en Turismo, de reciente creación en 2013, que se ha constituido inicialmente por 57 investigadores pertenecientes a las universidades de Alicante, Barcelona, Málaga, Rey Juan Carlos y Sevilla. Este programa de doctorado es hoy en día el más relevante en España por su amplitud y especialización en distintas áreas de investigación turística. Sus tres grandes líneas de estudio son: (1) planificación de destinos turísticos; (2) economía y empresas turísticas; y (3) turismo y patrimonio. Estos ocho autores mencionados son los siguientes (citados por orden de clasificación en este ranking): J.L. Nicolau, J. Pereira-Moliner, J.J. Tari Guillo, J.M. López-Bonilla, B. Ortega-Aguaza, R. Huete, F. Vera-Rebollo y J.L. Sánchez-Ollero.

Las ramas científicas a las que pertenecen los autores prolíficos están dominadas por las Ciencias Económicas y Empresariales, apareciendo un total

de 64 autores que pertenecen a estas disciplinas, lo cual representa un 81% del total de autores del ranking. Otras Ciencias Sociales están representadas por la Psicología (con tres autores), la Geografía (con dos autores), la Sociología (con dos autores) y la Antropología (con un autor). Además, aparecen también de manera muy limitada otras ramas o áreas científicas, como la Ecología, la Ingeniería o la Geología.

Respecto a las áreas de conocimiento científico, las tres cuartas partes de estos autores se concentran en tres áreas de conocimiento, como son: (1) Economía Aplicada, con 25 autores; (2) Comercialización e Investigación de Mercados, con 22 autores; y (3) Organización de Empresas, con 12 autores. A gran distancia se encuentran las áreas de Métodos Cuantitativos para la Economía y la Empresa y de Psicología Social, ambas con tres autores. Las áreas de Análisis Geográfico Regional, Ecología, Geología y Sociología disponen de dos autores cada una de ellas. En la tabla IV se detallan los datos de las principales áreas de conocimiento, con al menos dos autores, ordenándose en función del número de trabajos totales. Se puede apreciar que existe una clara distinción de las áreas de Economía Aplicada y de Comercialización e Investigación de Mercados, quedando esta última ligeramente por delante en cuanto al número de trabajos fraccionados. También se observa que las áreas de Sociología y de Métodos Cuantitativos alcanzan un ratio de trabajos fraccionados respecto a los trabajos totales superior al 50%, lo cual indica que los autores pertenecientes a estas áreas realizan un mayor esfuerzo que en otras áreas al publicar con un menor número de co-autorías.

En la tabla V se recoge la producción científica de las principales instituciones españolas en relación con sus autores prolíficos. Se ordenan de acuerdo al número total de trabajos publicados. Se observa que destaca ampliamente la Universidad de Islas Baleares, con 12 autores y 86 trabajos publicados, seguida de la Universidad de Alicante, con 9 autores y 66 trabajos. La Universidad de Alicante iguala prácticamente a la Universidad de Islas Baleares si se contempla el número de trabajos fraccionados por número de autores de cada artículo publicado. A continuación, en un segundo grupo de instituciones en el ranking están situadas las universidades de Valencia (8 autores) y de Las Palmas de Gran Canaria (7 autores), con 41 trabajos publicados cada una de ellas, aunque ésta última ostenta un mayor número de trabajos fraccionados. También se pueden agrupar en un tercer bloque las universidades de Granada (4 autores), Málaga (4 autores), Autónoma de Madrid (2 autores), Sevilla (3 autores) y Alcalá de Henares (1 autor), que han

Tabla IV. Principales áreas de conocimiento de los autores más productivos en la investigación turística española con difusión internacional (2002-2013)

AREAS DE CONOCIMIENTO	Nº Autores	Nº Trabajos en total	Nº trabajos fraccionados	Ratio
Economía Aplicada	25	158	62,63	39,6%
Comercialización e Investigación de Mercados	22	154	62,88	40,8%
Organización de Empresas	12	73	27,69	37,9%
Métodos Cuantitativos para Economía y Empresa	3	18	9,08	50,4%
Análisis Geográfico Regional	2	17	7,23	42,5%
Sociología	2	10	5,50	55%
Psicología Social	3	9	2,36	26,2%
Ecología	2	8	1,73	21,6%
Geología	2	7	1,58	22,6%

Tabla V. Producción científica en turismo por universidades y número de publicaciones de los autores más productivos españoles (2002-13)

Orden total	UNIVERSIDADES	Nº Autores	Nº trabajos totales	Co-autorías	Nº trabajos fraccionados	Orden fraccionado
1	Universidad de Islas Baleares	12	86	2,58	33,33	1
2	Universidad de Alicante	9	66	2,05	32,21	2
3	Univ. de Las Palmas de Gran Canaria	7	41	2,35	17,42	3
4	Universidad de Valencia	9	41	3,01	13,63	4
5	Universidad de Granada	4	21	3,10	6,78	8
6	Universidad de Málaga	4	19	2,56	7,41	6
7	Universidad Autónoma de Madrid	2	19	2,58	7,36	7
8	Universidad de Sevilla	3	17	2,12	8,04	5
9	Universidad de Alcalá de Henares	1	16	2,50	6,40	9
10	Universidad de Cantabria	2	13	2,36	4,65	12
11	Universidad de Navarra	3	12	2,42	4,97	10
12	Universidad de Castilla-La Mancha	3	12	2,83	4,24	14
13	Universidad de Vigo	1	12	2,33	3,86	16
14	Universidad de Córdoba	2	11	2,91	3,78	18
15	Universidad Autónoma de Barcelona	1	10	2,30	4,35	13
16	UNED	1	10	2,40	4,17	15
17	Universidad del País Vasco	1	9	1,89	4,76	11
18	Universidad de Jaén	1	9	2,33	3,86	17
19	Universidad Jaume I	1	9	3,67	2,45	19
20	Universidad Politécnica de Cataluña	1	9	4,89	1,84	21
21	Universidad de Girona	1	8	3,88	2,06	20
22	Universidad Complutense	2	8	4,63	1,63	22

publicado en una horquilla entre 16 y 21 trabajos. Sin embargo, este grupo lo lidera la Universidad de Sevilla en cuanto al número de trabajos fraccionados, alcanzando una cifra de ocho publicaciones.

4. DISCUSIÓN Y CONCLUSIONES

El análisis bibliométrico ha demostrado que la difusión internacional de la investigación turística española tiene un desarrollo muy reciente. En ge-

neral, el número de publicaciones internacionales mantiene una tendencia de crecimiento continuo, especialmente marcada por la entrada en vigor de la Ley de Universidades a finales de 2001, que ha actuado como la verdadera impulsora de esta evolución al exigir publicaciones internacionales a profesores e investigadores para su reconocimiento y acreditación de la carrera docente e investigadora. La gran mayoría de autores han publicado sus trabajos en revistas de prestigio internacional a partir del segundo período analizado (2008-2013). De hecho, la media de años de antigüedad de las publicaciones internacionales de todos los autores citados es de 4,5 años. Por el contrario, los autores más prolíficos en el primer período de estudio (2002-2007) pueden considerarse como los verdaderos pioneros de la investigación turística con difusión internacional. Estos autores más destacables en el primer período son Jaume Roselló, con nueve trabajos, A. Riera-Font, con ocho trabajos, y Enrique Claver-Cortés, con siete trabajos.

En el presente trabajo se ha realizado un análisis de autoría a través de una búsqueda muy exhaustiva y laboriosa de trabajos publicados en el ámbito turístico internacional. Con ello, se ha conseguido revisar prácticamente la totalidad de las publicaciones de autores afiliados a instituciones españolas. Esto permite obtener unas conclusiones mucho más precisas y representativas de la realidad que las que se han presentado en otros trabajos previos, en los que se ha utilizado un número reducido de revistas científicas analizadas o han sido revistas exclusivamente españolas o se han centrado en un ámbito de investigación muy específico. Además, los estudios sobre autorías en la investigación turística son todavía muy escasos y se han tratado de una manera tangencial. Por todo ello, el presente trabajo ha intentado superar todas estas limitaciones de los estudios previos aportando un amplio y profundo análisis de autorías respecto a la investigación turística española.

Los estudios bibliométricos sobre la investigación turística española e internacional se han limitado normalmente a seleccionar tan solo una serie de revistas especializadas en el ámbito turístico. Esto se justifica por razones operativas a la hora de facilitar la búsqueda bibliográfica. Sin embargo, esta estrategia conlleva también la ausencia de trabajos publicados en otras revistas científicas de prestigio no especializadas en turismo, pero que pueden tener un impacto y una representatividad relevantes en la investigación turística. Así, por ejemplo, en un estudio bibliométrico similar y muy interesante (y publicado en una revista de economía de la empresa), Albacete y otros (2013) analizan la investigación turística española de impacto internacional

a través de 26 revistas especializadas en turismo, durante un período de 15 años (entre 1997 y 2011), y consiguen identificar tan solo a 18 autores prolíficos con seis o más trabajos publicados en el contexto de la economía de la empresa. Por ello, es imprescindible contar con este requisito de exhaustividad en los estudios bibliométricos de esta índole para procurar su mayor rigor. Hoy en día, las bases de datos bibliográficas permiten mayores facilidades para cumplir con estas condiciones.

Seguidamente, se detallan algunas cuestiones relevantes en relación con los resultados obtenidos. En primer término, un autor destacable de la investigación turística con difusión internacional es J.L. Nicolau, que resalta tanto por su número total de publicaciones como por su mayor esfuerzo investigador al publicar con un menor número de coautores en cada uno de sus trabajos. Esto último es muy importante destacarlo porque en las ciencias sociales no es tan necesario que haya muchos autores en la publicación de un artículo. No obstante, es razonable que un investigador en ciencias sociales pueda publicar con varios autores en distintos artículos, sobre todo, cuando se realizan proyectos de investigación y/o se trata de trabajos de ámbito internacional. Pero ya deja de ser significativo cuando la gran mayoría de sus trabajos publicados aparecen firmados con muchos autores. En muchos de estos casos puede ser el trasfondo de favoritismos o pactos para aumentar artificialmente la producción científica. En este sentido, la clasificación que se establece en función de los trabajos fraccionados debe tener un peso específico más elevado en estas valoraciones bibliométricas, aunque también se deberían considerar en las valoraciones más institucionales sobre la investigación académica. El número de trabajos fraccionados puede denotar una mayor productividad de los autores prolíficos al publicar trabajos con un menor número de colaboradores por artículo. De este modo, se ha comprobado que varios investigadores (v.gr., B. Plaza, T.F. Espino-Rodríguez, A. Mantecón, J.M. López-Bonilla, L.M. López-Bonilla, J.I. Castillo-Manzano y R. Hueite), mejoran claramente su posición en la clasificación fraccionada, demostrando con ello que han realizado un gran esfuerzo en la publicación de sus trabajos de investigación entre 2002 y 2013.

Por otra parte, se aprecia todavía una desigualdad de género destacable en la investigación turística. Esto es debido a que la mujer se ha incorporado más tarde a la realización de estas funciones, como se demuestra en los años de antigüedad de sus publicaciones. Sin embargo, estos datos no son malos en su conjunto, dado que, en primer lugar, si se compara con los resultados obtenidos por Zhao y Ritchie (2007), en un estudio similar

a nivel internacional, aunque sean datos menos recientes, las mujeres suponen tan solo el 10,5% del ranking de autores prolíficos. Y, en segundo término, como dato más evidente de una futura tendencia hacia la equidad investigadora, las mujeres ocupan cuatro de las diez primeras posiciones en el presente estudio.

La media de autores por trabajo publicado en el presente análisis ha sido de 2,86 y, comparado con el estudio de Zhao y Ritchie (2007), resulta ser algo elevada. La media de Zhao y Ritchie (2007) está en torno a 2,1 autores por artículo, pero en España se ha tenido costumbre de incluir a tres o más autores en cada trabajo publicado, debido posiblemente a la inexperiencia y a las mayores dificultades que pueden tener los autores españoles para publicar en un ámbito prácticamente anglosajón. Es lógico que la tendencia deba ser decreciente y más consecuente con la realidad de la investigación en ciencias sociales, en cuanto que los medios materiales para llevarla a cabo no suponen una limitación importante. En el presente estudio tan solo aparecen diez autores prolíficos que publican con un número similar de firmas que la media obtenida por Zhao y Ritchie (2007). Esto supone tan solo un 12,7% sobre el total de autores del ranking. Estos investigadores son los siguientes: J.L. Nicolau (1,52), R. Huete (1,86) B. Plaza (1,89), A. Mantecón (2), J.M. López-Bonilla (2,13), L.M. López-Bonilla (2,13), T.F. Rodríguez-Espino (2,13), T.F. Espino-Rodríguez (2,13), E. Ruiz-Ballesteros (2,14) y M. Sánchez-Rivero (2,14). Curiosamente, hay que indicar que el primer autor del ranking (J.L. Nicolau) es el que presenta una media de autoría más reducida entre todos los autores analizados. Por lo tanto, es necesario que el número de autores por cada trabajo publicado disminuya y se aproxime más a la media internacional, en términos generales. Los organismos españoles de evaluación de la investigación deberían considerar esta circunstancia, valorando con mayor equidad los trabajos publicados en ciencias sociales y dando un mayor peso a aquellos autores que realizan un esfuerzo investigador más acentuado y más acorde con el panorama internacional.

El presente trabajo cumple una función divulgadora de la investigación turística en España. Puede ser un punto de referencia para académicos y doctorandos en este ámbito de estudio para conocer y valorar, al menos, una parte importante de los autores principales en materia turística que publican sus trabajos a nivel internacional. Por ejemplo, la existencia de un nuevo programa de doctorado en turismo en España, de carácter interuniversitario, constituye una buena oportunidad de fomentar la investigación turística. Este programa de doctora-

do ha sido creado recientemente y está coordinado por el profesor Fernando Vera Rebollo, y se puede afirmar que goza de una plantilla de investigadores de una alta cualificación, siendo una prueba de ello el hecho de que el 10% de los autores prolíficos citados en el presente estudio pertenece a dicho programa, que ha firmado en torno al 30% de las 285 publicaciones. En concreto, se han detectado ocho autores prolíficos que forman parte del programa y que avalan su importancia. Estos autores están afiliados cinco de ellos a la Universidad de Alicante (J.L. Nicolau, J. Pereira-Moliner, J.J. Tari Guillo, R. Huete y F. Vera-Rebollo), dos son de la Universidad de Málaga (B. Ortega-Aguaza y J.L. Sánchez-Ollero) y uno es de la Universidad de Sevilla (J.M. López-Bonilla). Cabe resaltar que el primer autor de este ranking (J.L. Nicolau) figura en dicho programa de doctorado, así como se puede esperar que vayan incorporándose otros investigadores de prestigio, junto a otros investigadores más noveles, en un futuro próximo. Por el contrario, hay que precisar también que la participación femenina está representada escasamente por la profesora Huete entre estos ocho investigadores.

En cuanto a las instituciones a las que pertenecen los autores prolíficos, se puede observar el predominio de la Universidad de Islas Baleares, con 86 trabajos publicados, seguida de la Universidad de Alicante, con 66 trabajos. La mayor concentración de autores y publicaciones se encuentra en estas dos universidades y a gran distancia del resto de instituciones. En un segundo nivel se encuentran las universidades de Las Palmas de Gran Canaria y Valencia. Y en un tercer grupo se pueden incluir las universidades de Granada, Autónoma de Madrid, Málaga, Sevilla y Alcalá de Henares. En general, cabe destacar el mayor esfuerzo investigador a nivel individual que realiza la Universidad de Alicante, dado que el número de trabajos por autor se iguala al de la Universidad de Islas Baleares. Esto también ocurre en el caso de la Universidad de Sevilla, ya que posee un bajo ratio de co-autorías por artículo. Aunque es recomendable que no haya un exceso de autores por trabajo publicado, como se ha comentado anteriormente, también puede ser necesario que exista un mayor grado de colaboración entre autores pertenecientes a distintas áreas de conocimiento y, sobre todo, entre instituciones diferentes, así como de países distintos. En este caso, puede estar más justificado que estos trabajos estén realizados por un mayor número de autores. No cabe duda que esto facilitaría el desarrollo de la investigación turística y le otorgaría una mayor capacidad para afrontar retos a escalas más elevadas. Sería interesante estudiar en el futuro las redes de colaboración entre autores prolíficos y su conveniencia para mejorar la producción científica.

Por otra parte, hay que señalar que las universidades ubicadas en las regiones españolas con mayor desarrollo de su sector turístico son las que aparecen liderando este ranking, como son, por orden del número de trabajos publicados de cada una de ellas: Comunidad Valenciana (116 trabajos), Andalucía (97), Islas Baleares (86), Madrid (53), Canarias (41) y Cataluña (27). En el resto de las regiones españolas destaca Galicia, con 18 trabajos publicados, seguida de Cantabria (13), Castilla-La Mancha (12), Navarra (12 trabajos) y País Vasco (9).

De acuerdo con los resultados obtenidos en el presente estudio, si se observan las universidades más destacadas y con más autores prolíficos (sobre todo aquellas que tienen, al menos, dos autores), se puede distinguir una serie de corrientes de investigación en el ámbito turístico repartidas por todo el territorio español. De este modo, y en un ejercicio meramente ilustrativo, nos encontramos con varias corrientes que sobresalen por su número de investigadores y por la importancia de la zona geográfica en la que se ubican respecto al sector turístico en España. Así, cabe subrayar la existencia de dos corrientes isleñas, como son (1) la corriente balear (12 autores y 86 artículos) y (2) la corriente canaria (7 autores y 41 artículos); (3) una corriente levantina, fomentada especialmente por las Universidades de Alicante y Valencia, contando entre ellas con 18 autores prolíficos, que han publicado 107 artículos; y (4) una corriente investigadora del sur, apoyada mayormente por las Universidades de Granada, Málaga, Sevilla, Córdoba y Jaén, en cuyas instituciones suman 14 autores, que han publicado 77 artículos en su conjunto.

Además, se pueden citar otras dos tendencias investigadoras más heterogéneas, como la existencia de una corriente central, respaldada principalmente por las Universidades Autónoma de Madrid, Alcalá de Henares, Castilla-La Mancha, UNED y Complutense, con ocho autores, que han publicado 65 artículos en total; y una corriente investigadora del norte, formada sobre todo por las Universidades de Navarra, Cantabria y Vigo, con seis autores y 37 artículos publicados en total. Por último, es destacable que las universidades catalanas, con tres autores de tres instituciones distintas y 27 trabajos publicados, se sitúan algo rezagadas respecto a las tendencias de investigación en el ámbito turístico, teniendo en cuenta la posición privilegiada de su sector turístico en la economía española. En este sentido, hay que decir que las tres áreas de conocimiento científico principales, en cuanto al número de publicaciones realizadas, no parecen tener una presencia tan fuerte hacia la investigación turística en las universidades catalanas. Esta aparente

producción científica más reducida puede estar influida por un sistema educativo en el que se han organizado los estudios universitarios de turismo en esta zona geográfica, contando con una mayor división de instituciones públicas y privadas en su oferta académica.

En España hay un clamor permanente por aumentar la productividad científica de las universidades y aspirar a posiciones más privilegiadas en los ranking internacionales. El esfuerzo individual de muchos investigadores españoles es evidente en los últimos años, a nivel general, pero también en lo que respecta a la investigación turística. Se admite que la comunidad científica se divulgue principalmente en el idioma inglés. Pero a esto hay que añadir que las editoras y las sedes de las revistas están ubicadas en países anglosajones. Por tanto, se requiere de un mayor esfuerzo institucional para favorecer el aumento de productividad que se reclama política y socialmente. En este sentido, el hecho de haber tan solo una revista científica en el ámbito turístico, como es *Cuadernos de Turismo*, que está desde hace muy poco tiempo en Scopus, demuestra una situación muy precaria en relación con el gran esfuerzo de investigadores y, sobre todo, con el alto nivel de exigencia de los responsables políticos y sociales. *Cuadernos de Turismo* es una revista con carácter bianual creada en 1998, que patrocina la Universidad de Murcia y dirige el profesor Espejo Marín, y que ha sido recientemente incorporada a *Master Journal List*, perteneciente a *Thompson Reuters* y *Web of Science*. Sin embargo, esta misma escasez de revistas científicas ocurre igualmente en otros ámbitos de la investigación social, como puede ser en el ámbito de las disciplinas de sociología, psicología y economía y empresa.

Las instituciones universitarias y organismos públicos, implicando en lo posible al sector privado, deberían apoyar económicamente los proyectos de creación y/o desarrollo de revistas españolas dedicadas a la investigación, en general, y a la investigación turística, en concreto, en consonancia con la posición turística que ocupa España en el mundo. Un ejemplo de esta implicación, aunque a nivel general y no referida exclusivamente al ámbito turístico, es el esfuerzo de algunas áreas, como las del ámbito de la Geografía, para impulsar sus propias revistas e incorporarlas a las bases de datos internacionales.

Como limitación principal del presente estudio cabe señalar que su enfoque se ha planteado desde una perspectiva cuantitativa. Aunque la gran mayoría de estudios sobre autorías en la investigación turística, tanto a nivel nacional como internacional, se han basado especialmente en esta perspectiva

cuantitativa, es necesario realizar otros estudios de carácter más cualitativo que complementen los resultados obtenidos y aporten una visión más holística de las contribuciones académicas.

Por otra parte, hay que recalcar que en esta investigación se ha considerado a los autores prolíficos a partir de publicar un mínimo de seis trabajos en revistas científicas de difusión internacional durante un período de doce años. Esta cifra parece relativamente escasa, pero se ha justificado por el hecho de existir todavía pocos autores con muchas publicaciones en torno a la investigación turística. Por lo tanto, no cabe duda que se han detectado todos los autores prolíficos en el período de estu-

dio, en un sentido estricto, así como aquellos otros autores que pueden estar próximos a serlo, y que son, en definitiva, los autores adscritos a instituciones españolas que más publican en el ámbito de la investigación turística.

5. AGRADECIMIENTOS

Los autores agradecen sinceramente las evaluaciones de los revisores anónimos.

ACKNOWLEDGEMENTS

The authors sincerely thank for the valuable comments of anonymous reviewers.

6. REFERENCIAS

- Agencia Valenciana de Turismo (2006). Análisis de las tesis doctorales de turismo realizadas en universidades españolas. *Papers de Turisme*, 40, 111-116.
- Aguiló, E. (2010). Una panorámica de la economía del turismo en España. *Cuadernos de Economía*, 33 (91), 5-42. [https://doi.org/10.1016/S0210-0266\(10\)70056-X](https://doi.org/10.1016/S0210-0266(10)70056-X)
- Albacete-Sáez, C.A.; Fuentes-Fuentes, M.M. (2010): "Difusión de la investigación española sobre turismo en revistas internacionales". *Revista de Análisis Turístico*, 9, 14-29.
- Albacete, C.A., Fuentes, M.M.; Haro-Domínguez, M.C. (2013). La investigación española en turismo con impacto internacional (1997-2011). Una perspectiva de la economía y la dirección de la empresa. *Cuadernos de Economía y Dirección de la Empresa*, 16, 17-28. <https://doi.org/10.1016/j.cede.2012.05.003>
- Antón i Clavé, S.; López Palomeque, F.; Marchena Gómez, M.J.; Vera Rebollo, J.F. (1996). La investigación turística en España: aportaciones de la Geografía (1960-1995). *Estudios Turísticos*, 129, 165-208.
- Ardanuy, J. (2014). La publicación de artículos en las áreas de ciencias sociales y humanidades analizadas a partir de los sistemas de información sobre investigación: el caso de las Universidades de Barcelona y Girona. *Revista Española de Documentación Científica*, 37 (1), e035. <https://doi.org/10.3989/redc.2014.1.1074>
- Bigné Alcañiz, J.E.; Andreu Simó, L.; Sánchez García, I.; Alvarado Herrera, A. (2008). Investigación internacional en marketing turístico: análisis de contenido sobre temas y metodologías. *PASOS. Revista de Turismo y Patrimonio Cultural*, 6 (3), 391-398.
- Bote Gómez, V. (1996). La investigación en España sobre turismo y desarrollo económico. *Estudios Turísticos*, 129, 9-22.
- Camisón Zornoza, C.; Monfort Mir, V.M. (1996). La calidad en el turismo español: balance y prospectiva de la investigación. *Estudios Turísticos*, 128, 129-161.
- Camisón Zornoza, C.; Bou Llusar, J.C.; Roca, V.; Montesinos, C. (1997). Enlace de la investigación en turismo y en gestión de la calidad: balance de una década. *Papers de Turisme*, 20, 19-38.
- Ceballos, C.; Arias, C.; Ruiz, A.; Sanz, C.; Vázquez, I. (2010). La formación en turismo en España: pasado, presente y futuro en el nuevo Espacio Europeo de Educación Superior. *Cuadernos de Turismo*, 25, 45-67.
- Chang, C-M.; McAleer, M. (2012). Citations and impact of ISI tourism and hospitality journals. *Tourism Management Perspectives*, 1, 2-8. <https://doi.org/10.1016/j.tmp.2011.08.001>
- Corral-Marfil, J.A. (2013). Estudio bibliométrico de las Actas del Congreso de la Asociación Española de Expertos Científicos en Turismo (AECIT, 1994-2012). *Revista de Análisis Turístico*, 16 (2), 33-44.
- Corral, J.A.; Cànoves, G. (2013). La investigación turística publicada en revistas turísticas y no turísticas: análisis bibliométrico de la producción de las universidades catalanas. *Cuadernos de Turismo*, 31, 55-81.
- Corral, J.A.; Cànoves, G. (2014). La colaboración científica en turismo: grupos y redes de investigación en Cataluña. *Documents d'Anàlisi Geogràfica*, 60 (1), 57-86. <https://doi.org/10.5565/rev/dag.84>
- Corral, J.A.; San Eugenio, J. de (2013). La investigación de la geografía del turismo en España a través del análisis bibliométrico de las actas del Coloquio de Geografía del Turismo, Ocio y Recreación (1990-2012). *Boletín de la Asociación de Geógrafos Españoles*, 63, 535-549.
- Corral-Marfil, J.A.; Rodríguez, I.M.; Vargas, A.; Cànoves, G. (2015). Estudio de la investigación turística a través de las coautorías de artículos: cálculo de indicadores de colaboración y análisis de redes sociales. El caso de las universidades catalanas. *PASOS. Revista de Turismo y Patrimonio Cultural*, 13 (4), 789-803.
- Esteban, A. (2000). La investigación turística en la universidad española. *Estudios Turísticos*, 144-145, 155-180.

- Figuerola Palomo, M. (2000). Los trabajos de investigación de la economía del turismo en el Instituto de Estudios Turísticos: (1974-1993). *Estudios Turísticos*, 144-145, 17-45.
- García Dory, M.A. (1977). Covadonga National Park, Asturias, Spain. Its history, conservation interest and management problems. *Biological Conservation*, 11 (2), 79-85. [https://doi.org/10.1016/0006-3207\(77\)90030-1](https://doi.org/10.1016/0006-3207(77)90030-1)
- González-Albo, B.; Moreno, L.; Aparicio, J.; Morillo, F.; Bordons, M. (2009). *La investigación española sobre turismo a través de publicaciones científicas de difusión nacional e internacional (Web of Science e ISOC 1998-2008)*. IEDCYT, CCHS, CSIC. Informe.
- Granadino, B.; García-Carpintero, E.; Plaza, L.M. (2006). La presencia española en consejos y comités de redacción de revistas científicas internacionales: un instrumento para la promoción de nuestra ciencia. *Revista Española de Documentación Científica*, 29 (3), 398-408.
- Hernández, J.M.; Campón, A.M.; Folgado, J.A. (2011). La investigación en turismo en España a través de las revistas de dirección de empresas. *Revista de Análisis Turístico*, 12, 1-9.
- Ibáñez, A.; Bielsa, C.; Larrañaga, P. (2013). Análisis de la actividad científica de las universidades públicas españolas en el área de las tecnologías informáticas. *Revista Española de Documentación Científica*, 36 (1), e002. <https://dx.doi.org/10.3989/redc.2013.1.912>.
- Jiménez-Contreras, E.; Moya, F.; Delgado, E. (2003). The evolution of research activity in Spain. The impact of the National Commission for the Evaluation of Research Activity (CNEAI). *Research Policy*, 32, 123-142. [https://doi.org/10.1016/S0048-7333\(02\)00008-2](https://doi.org/10.1016/S0048-7333(02)00008-2)
- López-Bonilla, J.M.; López-Bonilla, L.M. (2004). Evolución y perspectivas del enfoque interdisciplinario en el estudio del turismo. *Estudios Turísticos*, 160, 31-44.
- López-Bonilla, J.M.; López-Bonilla, L.M. (2008). Producción científica española en marketing turístico. *Estudios Turísticos*, 177, 35-50.
- McKercher, B. (2008). A citation analysis of tourism scholars. *Tourism Management*, 29 (6), 1226-1232. <https://doi.org/10.1016/j.tourman.2008.03.003>
- Miralbell, O.; Lamsfus, C.; Gomis, J.M.; González, F. (2012). "Estudio de las TIC y el Turismo en España. Análisis de las ponencias del congreso TURITEC entre 1999 y 2010". *TuriTec 2012*, pp. 498-517. Facultad de Turismo, Universidad de Málaga. Málaga.
- Moreno, S.; Picazo, P. (2012). Difusión de la investigación científica en revistas de turismo realizada por instituciones españolas. *Revista de Análisis Turístico*, 14 (2), 33-52.
- Moreno-Pulido, A.; López-González, M.A.; Rubio-Garay, F.; Saúl, L.A.; Sánchez-Elvira-Paniagua, A. (2013). Evolución de las revistas españolas de Ciencias Sociales en el Journal Citation Reports (2006-2010) y su valoración como indicio de calidad en la normativa evaluadora española. *Revista Española de Documentación Científica*, 36 (3), 1-15.
- Ortega Martínez, E.; Rodríguez Herráiz, B. (2004). La investigación turística a través de tesis doctorales. Un análisis comparativo entre España y Francia. *Estudios Turísticos*, 159, 5-27.
- Park, K.; Phillips, W.J.; Canter, D.D.; Abbott, J. (2011). Hospitality and tourism research rankings by authors, university, and country using six major journals: the first decade of the new millennium. *Journal of Hospitality & Tourism Research*, 35 (3), 381-416. <https://doi.org/10.1177/1096348011400743>
- Ruiz, A.; Chávez, E.; Arias, C. (2006). Estado de la investigación de las publicaciones sobre yield management en las revistas de turismo y hostelería. *Papers de Turisme*, 40, 7-29.
- Sánchez Pérez, M.; Marín Carrillo, M.B. (2003). La investigación en turismo y economía de la empresa publicada en revistas especializadas españolas: 1996-2001. *Papers de Turisme*, 33, 6-39.
- Vargas Sánchez, A. (2011a): ¿Los principales destinos son también las principales potencias en la investigación en turismo? *Estudios turísticos*, 188, 91-111.
- Vargas Sánchez, A. (2011b). Tourism research in Spain: the state of the art. *Enlightening Tourism, A Pathmaking Journal*, 1, 93-110.
- Zhao, W.; Ritchie, J.R.B. (2007). An investigation of academic leadership in tourism research: 1985-2004. *Tourism Management*, 28 (2), 476-490. <https://doi.org/10.1016/j.tourman.2006.03.007>

ESTUDIOS / RESEARCH STUDIES

El uso de las licencias libres en los datos públicos abiertos

Luis Fernando Ramos-Simón*

*Facultad de Documentación. Universidad Complutense de Madrid
Correo-e: lframoss@ucm.es | ORCID iD: <http://orcid.org/0000-0003-2267-8405>

Recibido: 29-02-2016; 2ª versión: 19-09-2016; 3ª versión: 04-11-2016; Aceptado: 08-11-2016.

Cómo citar este artículo/Citation: Ramos-Simón, L. F. (2017). El uso de las licencias libres en los datos públicos abiertos. *Revista Española de Documentación Científica*, 40(3): e179. doi: <http://dx.doi.org/10.3989/redc.2017.3.1376>

Resumen: Los datos masivos o big data recopilados a través de distintas herramientas tecnológicas o extraídos de grandes colecciones de información en distintos formatos son capaces de generar nuevo conocimiento en los sectores más diversos, pero, a la vez, plantean problemas tanto de titularidad como de su uso posterior. El estudio analiza el concepto de análisis de datos, los distintos tipos de datos y su posible regulación, presentando un panorama sobre cómo en los últimos años los países más avanzados en apertura de datos han comenzado a aplicar unos esquemas de licencias abiertas que remueven los obstáculos legales para incorporar estos tipos de datos a la creación de nuevo conocimiento. La investigación realizada muestra cómo distintos gobiernos y otras corporaciones y organismos internacionales productores de datos públicos han encontrado en estas licencias —en particular, las modalidades PPDL y CC0 en las que se renuncia a todos los derechos sobre las bases de datos— un mecanismo adecuado para favorecer el uso y reutilización de estos grandes volúmenes de datos.

Palabras clave: Acceso abierto; datos públicos abiertos, licencias de información; derechos de autor; análisis de datos; derecho sui generis sobre bases de datos; minería de datos; datos masivos; reutilización de información del sector público (ISP); Creative Commons; Opendatacommons.

The use of free licenses in open public data

Abstract: *Big data*, the massive amounts of data collected using different technological tools or extracted from large collections of information in different formats, are able to generate new knowledge in diverse sectors. But problems are posed regarding both their ownership and subsequent use. This study analyzes the concept of data analysis, as well as the different types of data and their possible regulation, presenting an overview of how in recent years some open licensing schemes have begun to be implemented for avoiding the legal obstacles facing the inclusion of these data types in the creation of new knowledge. The study shows how different governments and international organizations that produce public data have found an appropriate mechanism for promoting the use and reuse of large volumes of public data through these licenses - in particular, the PPDL and CC0 modalities in which all rights to databases are waived.

Keywords: Open access; open governmental data; information licenses; copyright; data analysis; sui generis database rights; data mining; big data; reuse of public sector information (PSI); Creative Commons; Opendatacommons.

Copyright: © 2017 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

El uso de los datos se intensifica cada año. Estos datos sirven para ayudar en soluciones tan diversas como en el diagnóstico de daños cerebrales, a encontrar el lugar ideal para la instalación de parques eólicos, a evitar la congestión del tráfico o a prever el rendimiento de los cultivos en países en desarrollo (Comisión Europea, 2014). Al tiempo aumentan los dispositivos de recogida de datos en ámbitos públicos y privados (calles, comercios, estaciones, aeropuertos, carreteras...) y el uso de redes de datos (meteorológicos, espaciales, cultivos, fauna...), así como las redes sociales y controles de navegación en Internet proporcionan un enorme caudal de datos, fácilmente convertibles en conocimiento en distintos contextos (Fitzgerald y otros, 2009). Además, actividades como los proyectos de digitalización de colecciones bibliotecarias han puesto de manifiesto el gran valor que tienen esas páginas de texto escaneadas, digitalizadas y categorizadas después de haber sido convertidas en datos, valor que va mucho más allá de poner al alcance de la humanidad grandes colecciones bibliotecarias en dominio público. Todo ello ha convertido a los datos en la gran revolución de nuestro tiempo (Mayer-Schönberger y Cukier, 2013).

La eclosión de estos datos masivos recopilados a través de diferentes métodos y herramientas tecnológicas (tratamiento masivo de textos y datos de todo tipo, sensores, satélites, cámaras...) da lugar a nuevas actividades mediante las cuales se analiza la información resultante para obtener nuevo conocimiento de todo tipo (McKinsey Global Institute, 2013), lo que plantea problemas de legalidad referidos a la titularidad de los propios datos y cuestionan si socialmente tales datos deben seguir reglas de propiedad o más bien deben disponer de un estatuto abierto que permita un uso generalizado de los mismos. El estudio muestra cómo, frente a otras alternativas, las licencias se han convertido en un instrumento relativamente sencillo y práctico plenamente adecuado para satisfacer las expectativas de usuarios y reutilizadores. De este modo, en el ámbito de los portales gubernamentales de los países calificados con una mejor apertura de datos, las licencias libres estándar internacionales, ya sea adoptándolas directamente o bien haciéndolas compatibles con las licencias gubernamentales gozan de una casi plena aceptación como licencias por defecto, siendo las que más favorecen el libre uso de las bases de datos, tales como PDDL y CC0, las que parecen adaptarse mejor a ese entorno de datos públicos abiertos. A pesar de todo, todavía queda mucho camino por recorrer en este propósito de poner los datos fácilmente accesibles para todos, libres e interoperables.

El contenido del artículo está estructurado en las siguientes partes:

- El concepto de análisis de contenido o minería de datos.
- Protección legal de los datos.
- Aplicación de las licencias libres a los datos públicos abiertos.
- Conclusiones.

2. OBJETIVOS Y METODOLOGÍA

Los objetivos fundamentales de la investigación son los siguientes:

- Analizar los nuevos conceptos y técnicas de tratamiento de los datos mediante los cuales se obtienen nuevas posibilidades de uso y reutilización para generar nueva información y conocimiento.
- Estudiar las características básicas de las bases de datos y sus formas de protección por derechos de autor. La revolución de los datos pone de relieve la necesidad de abordar transformaciones en su marco regulatorio (en particular, en las bases de datos y las licencias), de modo que las bases de datos sean fácilmente consultables para cualquier propósito.
- Determinar las licencias que mejor se adecúan a la puesta a disposición de los datos abiertos organizados para su reutilización en forma de bases de datos.
- Examinar los tipos de licencias que emplean los portales gubernamentales mejor considerados a nivel internacional y las condiciones de reutilización de los datos públicos accesibles.
- Contrastar y comparar el uso de las licencias estándar y compatibles para determinar que licencias se adaptan mejor a la oferta de datos públicos abiertos.

Con este propósito, se ha llevado a cabo una pormenorizada búsqueda en los portales gubernamentales de once países (los diez primeros más España, 13^o) que están entre los más avanzados en la apertura de datos, conforme a los resultados del Open Data Barometer (WWW Foundation, 2015), en este indicador se analizan y comparan parámetros de unos noventa países. Sus resultados se presentan en dos cuadros comparativos en el que se analizan las licencias empleadas —por defecto y compatibles— en la puesta a disposición de los datos públicos abiertos. El estudio revela que los portales gubernamentales de los países más avanza-

dos en la gestión de los datos abiertos disponen de licencias que permiten la reutilización de todo tipo de bases de datos mediante una regulación propia que además es compatible con las licencias estándar internacionales más avanzadas orientadas al libre flujo de los datos sin restricciones.

Además del apartado empírico, este trabajo también se nutre de la experiencia obtenida en diversos proyectos, así como de estudios y trabajos sobre activos de información en el ámbito de los datos y contenidos abiertos llevados a cabo por el grupo de investigación Publidoc-UCM desde 2006 (MHAP, 2012). Desde esta perspectiva temporal, llama la atención que se ha pasado de una situación inicial de tímida apertura de datos con situaciones muy diversas y declaraciones de principios muy generales, al escenario actual en el que una licencia de datos abiertos por defecto es el eje central que convive con otros usos diferenciados, ya sea regulando el uso mediante pago con tarifas o bien distinguiendo uso comercial de no comercial.

3. DESARROLLO

3.1. El concepto de análisis de datos o minería de contenidos

La minería de datos es el término más tradicional para designar el uso de tecnología y arquitectura diseñadas para detectar información procesable y así obtener rendimiento económico y social del gran volumen de información depositado en bases de datos y que es accesible directamente o través de Internet, ámbito conocido como datos masivos o Big Data¹, para lo cual se utilizan herramientas para la recopilación, extracción y recombinación de distintos elementos con objeto de obtener nuevos conocimientos.

Frente a un concepto más clásico que se centraba sólo en la minería sobre los datos, las nuevas capacidades tecnológicas permiten incluir textos, imágenes, figuras y todo tipo de materiales digitales o elementos documentales, lo cual ha conducido a una pérdida de valor del concepto original y al surgimiento de denominaciones alternativas a minería de datos, tales como minería de textos, minería de contenidos, análisis de datos...

De tal modo que, en primer lugar debemos sistematizar los elementos que integran el concepto que ahora se conoce como análisis de datos o minería de contenidos:

La minería de datos es una técnica computacional que permite cruzar variables en un entorno informático para descubrir nuevo conocimiento a partir de la información contenida en las bases de datos. De una forma más precisa, "la minería de datos

es un conjunto de técnicas, métodos y algoritmos que permiten extraer información útil desde fuentes de distinto tipo", señala Mendoza (2011). Los métodos de la minería de datos (DM, Data Mining por sus siglas en inglés) "permiten descubrir esta información y transformarla en un valioso conocimiento tanto retrospectivo (histórico) como prospectivo (proyecciones) o comprensivo (entender lo que ocurre)" (Colle, 2002). Se emplean distintos métodos y modelos para la minería de datos, siendo el KDD (Knowledge Discovery in Databases) el primer modelo aceptado por la comunidad científica en 1996. A partir del 2000 surgen tres nuevos modelos el SEMMA (iniciales en inglés de Muestreo, exploración, modificación, modelado y valoración), Catalyst (P3TQ, producto, mercado, precio, tiempo y cantidad) y CRISP-DM, siendo este modelo el más aceptado actualmente, estructurando la minería de datos en fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implantación (Moine y otros, 2011).

En cuanto a la tipología de los datos, Hand y otros (2001) distinguen entre datos observacionales y datos experimentales para separar los datos que son recopilados sin más propósito que registrar algún hecho, mientras que los datos experimentales son resultado de la actividad programada, de ahí que desde su perspectiva, "la minería de datos es el análisis de (a menudo grandes) conjuntos de datos observacionales para encontrar relaciones desconocidas y resumir los datos en nuevas formas que sean a la vez comprensibles y útiles para el titular de los datos". Más recientemente, Melero y Hernández-San Miguel (2014), en el ámbito de los datos de investigación añaden otras dos categorías: datos computacionales (datos de entrada y derivados de ciertos programas) y datos de simulación (generados a partir de modelos de prueba).

Por su parte, la minería de textos es una actividad de más reciente aparición, surge con los procesos de digitalización masiva iniciados en 2004 con el proyecto Google Books, razón por lo que el término ha recibido menos atención que la minería de datos (Truyens y Van Eecke, 2014). Además de texto entran en el objeto de análisis las imágenes, vídeos, sonidos, figuras... La peculiaridad de esta nueva forma de análisis de los textos radica en que los procesos de indexación, tratamiento y/o uso de los textos e imágenes pueden estar más próximos a las aspiraciones de los autores por lograr una protección o retribución por este tratamiento que es subsiguiente al proceso creativo, como veremos más adelante.

Algunos autores (Triaille y otros, 2014) para estudiar este nuevo fenómeno, prefieren la expresión

análisis de datos, en lugar de minería de textos y datos, que definen como "el procesamiento automático de materiales digitales, los cuales pueden incluir textos, datos, sonidos, imágenes u otros elementos o una combinación de estos, con el fin de descubrir nuevo conocimiento o ideas estratégicas". El análisis de datos es un concepto más amplio que minería que se asocia a la extracción de información, mientras que el análisis implica "rastreo, procesamiento, comparación, copia, análisis, recuperación, interpretación, búsqueda, sintaxis, etc."

Por su parte, la expresión minería de contenidos es la preferida por otros autores (Guadamuz y Cabell, 2013), para definir el proceso de análisis de información mediante software que busca patrones interesantes o importantes que no podrían ser observados de otros modos, en ámbitos tales como los hospitales, el medio ambiente, el comercio o los sentimientos sociales. El análisis de contenido ha crecido debido al incremento de la capacidad de computación centrado en el descubrimiento de nuevo conocimiento en las grandes bases de datos, ámbito en el que la minería de datos sería una de sus ramas.

Como se observa, esta nueva actividad es el resultado de la capacidad de análisis que proporcionan los grandes sistemas de información, haciendo posible la obtención de nuevo conocimiento a partir del análisis de grandes volúmenes de datos. Desde el punto de vista legal, este nuevo fenómeno debe de ser estudiado en el marco regulatorio de las bases de datos. Es decir, las recopilaciones de obras, de datos o de otros elementos independientes dispuestos de manera sistemática o metódica y accesibles individualmente por medios electrónicos o de otra forma, es la realidad sobre la que se asienta toda esta actividad de análisis de datos que es nuestro objeto de estudio.

3.2. Tipos de datos y su protección como base de datos

En sí mismo, el dato es un componente que aislado no tiene ningún significado (Mendo y otros, 2013), pero "organizado en conjuntos de datos o bases de datos pueden constituir un activo de gran valor en actividades como la investigación, el marketing, la economía, y la gestión en todos los ámbitos: comercial, transporte, recursos humanos, hospitales..."

Como resulta evidente, todos los datos para su tratamiento y gestión automatizada están organizados en conjuntos o bases de datos, aspecto esencial para determinar las posibilidades de acceso y uso de la información allí contenida. La legislación europea (Parlamento Europeo, 1996), señala que "tendrán la consideración de «base de datos» las

recopilaciones de obras, de datos o de otros elementos independientes dispuestos de manera sistemática o metódica y accesibles individualmente por medios electrónicos o de otra forma" (art. 1.2).

Asimismo, están protegidas las bases de datos que no cumplen el requisito de la originalidad, pero en las que se ha realizado una inversión sustancial, ya sea financiera, de esfuerzo o de tiempo, para la "obtención, la verificación o la presentación de dicho contenido"(art. 7), es lo que la Directiva europea denomina bases de datos *sui generis* y que ha creado una notable controversia (Correa, 2007) por los efectos que tiene sobre la protección y apropiación del conocimiento en general, dado que si gran parte del conocimiento y memoria mundial está almacenada en bases de datos, todos esos contenidos pertenecerían a los titulares de esas bases de datos, con independencia de que sus contenidos sean creaciones originales, datos de dominio público o simples datos o representaciones obtenidas de la naturaleza.

De acuerdo a la definición y al contenido de la información de dichas bases de datos, podemos establecer las cuatro categorías que se especifican a continuación, si bien una misma base de datos puede tener protección jurídica en una o varias de esas categorías, aspecto a tener en cuenta en cuanto a su acceso, consulta y uso. Es especialmente relevante que las condiciones de la licencia seleccionada se adecúen al uso o reutilización pretendidos.

3.2.1. Bases de datos originales

Como decíamos más arriba, la protección de los derechos de autor se refiere a su estructura en cuanto forma de expresión o disposición de los contenidos, sin alcanzar a estos. Lo que se protege es la selección o disposición de tales contenidos, con independencia de éstos, ya sea por su diseño, formato o la sistemática de las anotaciones u ordenación de los elementos. Asimismo, los contenidos pueden estar también protegidos, en el caso de que por sí mismos constituyen obras protegibles (por ejemplo, un artículo académico incorporado a una base de datos científica). Sólo podrá consultarse sin autorización del autor con fines de ilustración de la enseñanza o de investigación científica, en la medida justificada, sin fines comerciales e indicando la fuente.

3.2.2. Bases de datos amparadas por un derecho "sui generis"

El segundo tipo de bases de datos mencionado en la Directiva de 1996, se refiere a las bases de datos "no originales" (López, 2002) en las que su

titular ha realizado una inversión sustancial para la "obtención, verificación o presentación de su contenido" (art. 7). Es importante subrayar que si bien ambas bases de datos (originales y amparadas por el derecho *sui generis*) están protegidas legalmente, la inclusión en uno u otro apartado es determinante, por cuanto en un caso se protege la obra como un todo, en el segundo caso, sólo la inversión sustancial y durante un período de quince años, si no se actualiza la base de datos (Cámara-Lapuente, 2007).

Como se ha señalado desde el lado latinoamericano, al igual que desde otros países y posiciones, el principal argumento en contra de la protección de las bases de datos fácticas no es que proteja las bases de datos en sí mismas, sino "la información en ellas contenida, con el consiguiente riesgo de que se establezcan límites a la circulación de aquella, incluso de la que hasta el presente se mantiene en el dominio público" (López, 2002). Añade el autor, en este estudio bien documentado, que hay otros dos argumentos más en contra de la protección, de un lado genera protecciones perpetuas —haciendo imposible el paso al dominio público— al ser su caducidad renovable y favorece la creación de mecanismos de rastreo y recopilación, en lugar de incentivar el esfuerzo creativo. Otros análisis de la directiva europea de bases de datos sugieren que la idea de dotar de una ventaja competitiva a los productores de bases de datos descuidó el necesario equilibrio de intereses entre editores, usuarios y autores (Bitton, 2008).

Además de estas críticas en contra del monopolio que suponen estas bases de datos, ha surgido también un movimiento igualmente poderoso a favor del libre uso de la información bajo el presupuesto de que su uso abierto y generalizado fomenta el bienestar social y crea riqueza.

3.2.3. Bases de datos públicos

Los datos públicos existentes en organismos públicos de la Unión Europea tienen un tratamiento específico en virtud de las Directivas de reutilización de información del sector público de 2003 y 2013, en las que se prevé la puesta a disposición del público para la reutilización de sus documentos "cualquier contenido sea cual sea el soporte (escrito en papel o almacenado en forma electrónica o como grabación sonora, visual o audiovisual —art. 2.3—)" conservados por organismos del sector público para fines comerciales o no comerciales. El principio general es la libre puesta a disposición de tales contenidos, en caso de aplicar alguna tarifa, ésta "se limitará a los costes marginales en que se incurra para su reproducción, puesta a disposi-

ción y difusión" (art. 6.1)². La mayoría de los países europeos están incorporando ambas directivas mediante la creación de portales de datos abiertos y los correspondientes mecanismos de licencias a través de los cuales se da acceso a las bases de datos públicas.

La información del sector público es tradicionalmente el principal proveedor de información sistematizada, estructurada y de buena calidad en la mayoría de los sectores (información geoespacial, meteorológica, catastral, etc.). Desde la década de los noventa del siglo XX esta información fue objeto de atención especial ante la visión de que en una sociedad basada en la información, la generada en el sector pública iba a ser una materia prima esencial para la innovación y el crecimiento económico (Comisión Europea, 1998).

Entonces, el panorama era muy distinto a un lado y otro del Atlántico, mientras que en Estados Unidos la Ley de Libertad de Información (U.S. Department of Justice, 2016) proclama desde 1966 el libre acceso de todos los ciudadanos a la información generada por el Gobierno Federal no existiendo derechos reservados sobre la información administrativa federal (Comisión Europea, 1998)³. Sin embargo, Europa hubo de hacer frente a una realidad muy fragmentada en cuanto a tradiciones administrativas, lenguas y territorios. Esa era la situación cuando se aprueba en 2003 la Directiva de información del sector público con el propósito de crear unas mínimas reglas comunes en cuanto a transparencia, igualdad de acceso a los documentos y al principio de tarifas basadas en los costes. La Directiva de 2013 (Parlamento Europeo, 2013), actualización de la de 2003, ha profundizado principalmente en la apertura de datos libres o a coste marginal, en el fomento de la reutilización y en la estandarización de formatos. Estas iniciativas y la política de datos abiertos del comienzo de la presidencia de Obama han impulsado en todo el mundo los portales de datos nacionales desde donde son accesibles miles de recursos en formatos de base de datos, aunque todavía queda un gran camino en cuanto a la actualización de datos, formatos e interoperabilidad, aspectos centrales para llevar a cabo actividades eficientes de análisis de datos.

El procedimiento que se pretende generalizar es el de acceso a través de licencias abiertas para permitir el libre uso de los datos, es decir, "el dato es legalmente abierto cuando o bien está en dominio público o con una licencia que permite el uso y reutilización comercial y no comercial sin restricciones", tal como señala la Guía de datos abiertos de la ONU (2013). A pesar de los avances que se dan en la implantación de licencias abiertas y a lo que luego nos referiremos, es necesario subrayar que la proli-

feración de condiciones de uso de los datos, la falta total de información de cómo usarlos y el empleo de licencias que no son adecuadas aparecen como las principales barreras de acceso a la información del sector público, tal como se ha señalado reiteradamente en los últimos años, así en el informe de la consultora Deloitte, encargado por el Gobierno del Reino Unido (Deloitte, 2013) o en el estudio de Vickery (2012) para la Comisión Europea.

3.2.4. Bases de datos personales

Estas bases de datos, como su nombre indica, contienen datos referidos a las personas⁴ y, por lo tanto, están sometidas a la legislación específica de protección de datos personales (Parlamento Europeo, 1995)⁵. La protección de los derechos de autor no puede "hacerse extensiva a su contenido y se entenderá sin perjuicio de los derechos que pudieran subsistir sobre dicho contenido" (Parlamento Europeo: 1996). En este sentido, el uso de bases de datos personales para fines distintos a los que fueron creadas exige un proceso previo de anonimización de los datos personales para cumplir la Directiva en cuanto a la libre circulación de esos datos personales. Es decir, estas bases de datos cuando los datos sean anonimizados tendrán la protección que les corresponda en función de sus características. Un tratamiento específico, pero que no reviste ninguna particularidad respecto a los derechos de autor, reciben las bases de datos de solvencia patrimonial y de prospección comercial (Rebollo y Serrano, 2014).

4. ACCESO Y USO DE LOS DATOS: EXCEPCIONES A LOS DERECHOS DE AUTOR Y LICENCIAS SOBRE LOS DATOS

Como queda explicado, el acceso, uso y posible reutilización de la información contenida en las bases de datos está regulado por las leyes, así como mediante contratos y/o licencias. Las licencias de información⁶ son una modalidad que se ha generalizado en Internet por ser un elemento sencillo para modular el acceso a la información y en particular por ofrecer mecanismos simples para regular las relaciones entre autores y usuarios de contenidos protegidos por derechos de autor, de hecho es la solución impulsada por la legislación europea ("Los Estados miembros alentarán a todos los organismos del sector público a que utilicen las licencias modelo", art. 8.2, Parlamento Europeo, 2013). Un aspecto interesante es si una licencia o contrato puede prevalecer sobre lo dispuesto en un texto legal, disponiendo así, por ejemplo, si puede o no ser usado un material para realizar minería de datos o textos, aspecto que no tiene una solución unánime en las legislaciones europeas (Truyens y

Van Eecke, 2014). Mayor interés para nuestro propósito reviste el debate sobre si es posible aplicar una excepción a los derechos de autor al análisis de los datos. Veamos a continuación este aspecto, antes de adentrarnos en las licencias para uso de datos públicos.

4.1. Las excepciones de derechos de autor en el análisis de datos

El debate sobre el acceso y uso de la información en Internet, es decir, qué tipo de protección de derechos de autor deben tener los contenidos, se ha extendido también a si deben estar protegidas las bases de datos en las actividades de minería de datos. Las posiciones no son muy diferentes a las que existen en torno al acceso abierto y al acceso comercial en los contenidos protegibles por derechos de autor; sin embargo, en el aspecto concreto del análisis de datos parece abrirse paso una excepción a los derechos de autor en favor de permitir ciertos usos, al menos los que tienen por finalidad la investigación.

Por lo descrito hasta ahora resulta difícil encontrar una base de datos cuyo acceso no pueda ser restringido por su titular, pues aun en los casos en que sean contenidos que están en dominio público o datos del sector público, sus recopiladores pueden ejercer la titularidad en base a la inversión económica efectuada, baste pensar en las colecciones digitales como Google Libros, u otras muchas colecciones digitales de las bibliotecas en donde aparece con frecuencia el símbolo de protección de los derechos de autor (©). De modo más rotundo, Tuyens y Van Eecke (2014) aseveran que fuera de la investigación no comercial es difícil crear un estatuto en favor de estas actividades de análisis de datos, especialmente en los casos de bases de datos comerciales y con textos archivados que tengan un mínimo nivel de creatividad. En este sentido, los autores comentan la diferencia entre Europa y Estados Unidos en donde no existe una legislación comparable sobre bases de datos y se aplican de modo más flexible las reglas del "fair use", esto es, un criterio interpretativo que se adapta a los cambios socioeconómicos y tecnológicos sin tener que actualizar constantemente la legislación formal.

Una cuestión clave en el debate sobre el análisis de datos es si los autores deberían estar autorizados a prohibir el uso de una obra para un propósito diferente al de la creación original, por ejemplo, en los casos en que 100 horas de grabaciones o de mil novelas que sean tratadas para analizar aspectos tales como la imagen de la mujer, el grado de uso de una tecnología o el vocabulario o la sintaxis empleada en un grupo social. La solución

es difícil que sea concluyente, en un primer momento la respuesta debería de ser negativa porque los derechos de autor protegen la expresión creativa, mientras que el análisis de datos de cada obra constituye una materia prima (*raw data*) que sólo adquiere valor junto con muchos otros textos. Por contra, la posición a favor de la protección reforzará el papel de quienes se aprovechan de la creatividad ajena a través de la mera acumulación de datos. Un aspecto no menor es si en tales actos de explotación la autorización le correspondería a cada autor por su aporte creativo o al titular de la base de datos, puesto que el bien explotado sería la mera colección de los datos. El asunto lejos de ser quimérico podría afectar a miles de editores, autores, colaboradores periodísticos, intérpretes, guionistas, compositores...

En este contexto se ha abierto un debate en Europa y en otras partes del mundo sobre si es posible la introducción de una excepción legal en el ámbito de la propiedad intelectual en las actividades de análisis de datos asegurando un equilibrio de los diferentes intereses implicados. La excepción ya está en aplicación o en estudio en algunos países (Japón o Reino Unido).

Un estudio muy sólido encargado para la Comisión Europea (Jean Paul Triaille y otros, 2014) aborda un análisis pormenorizado sobre el uso del análisis de datos en las bases de datos. En síntesis, tanto las bases de datos originales como las protegidas con derecho "sui generis", desde un punto de vista legal están protegidas, por lo tanto es necesario abrir vías alternativas para permitir el uso de las técnicas de análisis de datos, y que incluso la aplicación de la excepción de investigación —reconocida en la Directiva de bases de datos— encuentra dificultades interpretativas y, además, en dicha excepción se hace explícito que no está permitido el uso comercial. Así las cosas, los autores del informe proponen varias alternativas: facilitar acuerdos de entendimiento entre titulares y usuarios de las bases de datos, adoptar un documento interpretativo en donde la Comisión Europea clarifique los aspectos de la Directiva que dificultan una interpretación armonizada o proponer un cambio legislativo en el que se introduzca una excepción en toda la Unión Europea para facilitar las técnicas de análisis de datos, siendo ésta la vía por la que se inclina el documento⁷ y la que ha respaldado la Comisión Europea (2016).

5. APLICACIÓN DE LAS LICENCIAS LIBRES EN LOS DATOS PÚBLICOS ABIERTOS

La introducción de una excepción legal para permitir una utilización menos restringida de los datos ofre-

cería seguridad, sin embargo, esa solución es lenta, más en el entorno de la Unión Europea, en donde a la aprobación de las directivas hay que sumar otro período de aplicación en cada país. Por el contrario, las licencias de información ofrecen una solución sencilla en la circulación por Internet de contenidos protegidos por derechos de autor. La aparición en 2003 de Creative Commons y en 2004 de Open Knowledge Foundation (Murray-Rust, 2008), han servido para impulsar las licencias abiertas, llamadas también libres, porque son capaces de indicar a personas no expertas, mediante unos pocos símbolos, los derechos, permisos y restricciones que ofrece el titular de los derechos a sus posibles usuarios, lo que facilita la circulación por Internet de contenidos sin tener que adaptarse previamente a la singularidad de cada país y de su sistema jurídico. En el caso de los datos se han abierto vías para que algunas modalidades de licencias permitan cualquier tipo de uso o reutilización, al incluir estas licencias la renuncia a todos los derechos por parte de los titulares, incluido el derecho sobre las bases de datos no originales.

5.1. Las licencias libres aplicables a los datos

Como es sabido, mediante estas licencias abiertas o libres (Liang, 2004), (Rosen, 2005), los titulares de contenidos protegidos por derechos de autor limitan algunas formas de explotación de la obra (explotación comercial, creación de obras derivadas, reutilización...) y permiten otros usos (usos comerciales o no, compartir con otros las modificaciones, etc.). Estas condiciones de explotación se incorporan a la obra mediante un código digital al difundirla en Internet, de modo que el posible lector o usuario de la obra en línea conoce de antemano cuáles son los derechos que el autor le permite o no, sin demasiada ambigüedad. Desde hace años estas licencias se usan en algunas áreas de conocimiento (O'Boyle y otros, 2011), en gran parte de publicaciones científicas en acceso abierto, así como en otros sitios en los que se acogen contenidos producidos por los usuarios, tan populares como *Youtube* o *Flickr*, y cada vez más lo hacen algunos gobiernos con la información pública puesta a disposición de los reutilizadores, siendo esta práctica considerada razonablemente apropiada en el ámbito académico y comienzan a sugerir que las licencias vayan incorporadas a los "datasets" para facilitar su uso (Allen, 2014).

Señaladamente, en el ámbito del tratamiento de los datos científicos, un gran número de organizaciones del sector han optado por poner los datos y bases de datos a libre disposición de la comunidad científica mediante este tipo de licencias (Hrynaszkiewicz y Cockerill, 2012), adecuadas al

tratamiento de los datos, que pretenden soslayar las diferencias entre las legislaciones nacionales, con el objetivo central de que los posibles derechos de autor no actúen como una barrera en el progreso del conocimiento. En sectores donde predominan los datos públicos, como el de los geodatos se aboga claramente en favor de que los titulares de derechos promuevan un uso de los datos sin restricciones creando una comunidad de información (Information Commons) que beneficie a usuarios y titulares (CODATA, 2015), (GITA White Paper, 2005). No obstante, un estudio señala que estamos todavía lejos de ese propósito al mostrar que de 4.012 bases de datos analizadas, la mitad no informaba sobre las condiciones de uso del sitio (Guadamuz y Cabell, 2013). Según esta fuente, de las bases de datos con licencias de acceso, las utilizadas con más frecuencia forman parte de las familias de licencias Creative Commons (no se detalla por separado la licencia específica para datos u metadatos CC0, si bien en la fecha del estudio esta licencia estaba poco implantada) y de Open Data Commons. En Francia, parece también que ambas licencias son las más utilizadas, aunque el autor percibe que los datos se usan poco y se hace preciso crear unas condiciones propicias para impulsar el uso de los datos públicos (Boustany, 2013).

A continuación, reseñamos las características más relevantes de estos dos grupos de licencias en el uso de los datos abiertos:

Dentro del conjunto de licencias de Creative Commons (Ramos, 2014)⁸, la licencia CC0 1.0 Universal (CC0 1.0) (http://creativecommons.org/publicdomain/zero/1.0/deed.es_ES), está especialmente orientada al uso de datos y metadatos. De hecho, los más de cincuenta millones de metadatos de los objetos digitales de Europea son accesibles bajo esa licencia. La licencia CC0, como se indica en su título, es un "ofrecimiento al dominio público". De este modo el proveedor de los datos renuncia a todos los derechos de propiedad intelectual (incluido el derecho *sui generis* sobre las bases de datos) que recaigan sobre los datos suministrados; ello implica que el proveedor ha de obtener autorización de terceros en el caso de que no sea titular de los mismos. Asimismo, si la renuncia de derechos del proveedor sobre los metadatos no es legalmente posible a través de la licencia CC0, será de aplicación la licencia pública alternativa (Public License Fallback), lo que implica que otorga a cada persona afectada "una licencia exenta de cánones, intransferible, no sublicenciable, no exclusiva, irrevocable e incondicional para el ejercicio" de todos los derechos sobre la obra en todo el mundo y por la duración máxima prevista en la ley aplicable. En la actualidad, la página Web de Creative Commons

(https://wiki.creativecommons.org/CC0_use_for_data) relaciona 49 organizaciones que usan la licencia CC0, tan relevantes como la Organización Europea para la Investigación Nuclear (CERN), Europea o U.S. Government Data, el portal de datos del Gobierno de Estados Unidos.

En el ámbito de las licencias Creative Commons los otros dos tipos de licencia que se usan con frecuencia en la apertura de datos son la licencia CC-BY (Reconocimiento) y CC-SA (Compartir igual). En la primera, el titular de los datos exige que se reconozca su titularidad. En la segunda, CC-SA, el titular permite reutilizar y compartir los datos con tal de que la distribución de esos datos resultantes se haga con la misma licencia. Estas dos licencias no son específicas para compartir bases de datos *sui generis*, aunque siguen siendo las más utilizadas para estos propósitos.

La familia de licencias Open Data Commons (ODC), fue creada en 2006 a partir de un borrador elaborado por Talis Platform (Miller; Styles; Heath, 2008) están promovidas por la organización Open Knowledge Foundation (OKF) (<http://opendatacommons.org/licenses/>) que dispone también de tres tipos de licencias de datos abiertos para su aplicación a las bases de datos:

- Public Domain Dedication and License (PDDL) — "Public Domain for data/databases". La licencia es, como la licencia CC0, una cesión al dominio público, mediante la cual el titular de los derechos de autor ofrece la obra al beneficio público y renuncia a los derechos de autor y derechos *sui generis* que puedan recaer sobre la base de datos, por lo cual la obra es libre y está abierta para otros usos. Asimismo, la renuncia es para todo el mundo y para todos los derechos presentes y futuros en cualquier formato. Del mismo modo, la licencia prevé que si la renuncia resulta inaplicable en alguna jurisdicción, el licenciador autoriza todos los usos sobre la base de datos mediante una licencia no exclusiva, libre de derechos en todo el mundo y por todo el tiempo de protección de la obra. Además, la licencia contiene una renuncia expresa al ejercicio de los derechos morales.
- Attribution License (ODC-By) — "Attribution for data/databases". La licencia permite a los usuarios compartir libremente, modificar y usar la base de datos con el único requisito de atribuir la paternidad de la obra, libre de derechos, no exclusiva por todo el tiempo que dure la protección de la base de datos, con derecho de extracción y reutilización de la base de datos. La base de datos puede

estar protegida por otros derechos que no están contenidos en la licencia (datos personales, contratos privados o marcas...), asimismo recoge la renuncia a los derechos morales del titular. Es similar a la CC-BY (Reconocimiento de la autoría).

- Open Database License (ODC-ODbL) — “Attribution Share-Alike for data/databases”. La licencia es la tercera de las opciones y su finalidad es permitir a los usuarios compartir libremente, modificar y usar la base de datos manteniendo esa misma libertad para otros usuarios, es decir, con cláusula “copyleft”. Por lo demás, la licencia es similar a la anterior en cuanto cobertura temporal y derechos de acceso y uso de la base de datos, no afectando los contenidos individualmente. Esta licencia también es muy parecida a la licencia CC-SA, de modo que las obras derivadas deben disponer de la misma licencia para compartir las obras.

5.2. Resultados: uso de las licencias en los portales de datos públicos

Al igual que en el ámbito científico, en el sector público se toma conciencia de que la información es una fuente de progreso y riqueza, de ahí la necesidad de poner a disposición de los ciudadanos y de las empresas la información y datos generados por el sector público. Si la Unión Europea, como mencionamos arriba, ya había adoptado en 2003 la Directiva de reutilización de información del sector público (modificada en 2013), la OCDE adoptó sendos documentos para impulsar el uso de datos abiertos en todos los países miembros (OCDE, 2007 y 2008), al proponer una mejora del acceso y uso de la información del sector público para que se adoptan unos principios en tal sentido.

Interesa subrayar la recomendación que incide en favorecer las condiciones de acceso y transparencia en la reutilización, “haciendo disponible y desarrollando un sistema de licencias automatizadas en línea para dar cobertura a la reutilización cuando se aplican licencias, teniendo en cuenta los derechos de autor”. En este sentido se recomienda favorecer la reutilización para facilitar su acceso y uso, incluyendo mecanismos de licencias sencillas para hacer las obras ampliamente accesibles al público.

De hecho, a partir de 2009, coincidiendo con la declaración de Obama sobre datos abiertos (The Obama Administration’s... 2011), se han dado grandes avances en la apertura de datos públicos. Muchos gobiernos y organizaciones internacionales han puesto grandes colecciones de datos a disposición del público que, o bien se acogen a una licencia

estándar que permite un amplio uso de los datos, o adaptan su propias licencias nacionales⁹ a los estándares de licencias promovidos por organizaciones como la OKF (Ramos y otros, 2012). En todas estas iniciativas de datos abiertos se asumen unos principios básicos sobre los que se debe publicar la información y configurar las licencias. Estos principios (ONU, 2013, 37) son:

- Información accesible en Internet.
- En formatos procesables electrónicamente.
- Licencia abierta que permita la reutilización comercial.
- Libre de cargas y sin restricciones de reutilización.

De tal modo que las Licencias para un “Gobierno abierto” se asientan sobre estos principios básicos “Open Government”, tal como la impulsada por el gobierno británico (HM Government, 2012) en las que se establece:

- Libertad para copiar, publicar, distribuir y transmitir información.
- Adaptar la información.
- Hacer uso comercial, combinarla con otras informaciones o incluirla en los propios productos o aplicaciones.
- Con la obligación básica de incluir la fuente de información y la licencia empleada.

En este contexto, en la Tabla I, se muestran las licencias de referencia, por defecto, que utilizan los portales gubernamentales de datos de once países: los diez países mejor clasificados a nivel mundial (y España, 13ª posición) en el ranking Open Data Barometer (WWW Foundation, 2015).

A pesar de la variedad de situaciones se aprecia una clara tendencia en el predominio de la aplicación de las licencias libres cuyas características reseñamos en el epígrafe anterior. Aunque todavía falte mucho camino por recorrer en la homogeneización de los requisitos para el acceso a los datos y su reutilización —como demuestra con mucha precisión Morando (2013) en un estudio sobre interoperabilidad legal— debe quedar claro que la aplicación de estas licencias supone un gran avance en el panorama internacional de intercambio y uso de los datos.

Algunos informes nacionales de los organismos de datos abiertos, como el de Holanda, reconocen el liderazgo indiscutible que ejercen en esta materia Reino Unido y Estados Unidos, por cuanto hacen disponibles los datos en un portal nacional lo cual resulta adecuado para un gobierno abierto y transparente. En esta cuestión sobrevuela el de-

Tabla I. Licencias de referencia en los portales gubernamentales de los países con mejor apertura de datos

Puesto	País	Web Portal de datos	Licencia
1	Reino Unido	http://data.gov.uk/	<i>Open Government Licence</i> , por defecto: Recomendada para todos los organismos públicos que no aplican cargos. http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/ . Compatible con CC-BY 4.0 y ODC-BY.
2	Estados Unidos	http://www.data.gov/about	Los datos federales se ofrecen libres y sin restricciones. Los datos no federales pueden tener diferentes licencias. http://www.data.gov/privacy-policy#license . Recomendada atribución: <i>Data retrieved from Data.gov (https://www.data.gov/)</i>
3	Suecia	http://data.riksdagen.se/In-English/	Los datos públicos sobre el Gobierno están disponibles en la web Riksdag, "sus bases de datos están libremente disponibles para cualquiera que desea usarlas" http://data.riksdagen.se/In-English/
4	Francia	https://www.data.gouv.fr/fr/	https://www.data.gouv.fr/fr/terms/ , por defecto, compatible con estándares internacionales (ODC-BY, CC-BY 2.0)
5	Nueva Zelanda	http://www.data.govt.nz/	<i>Creative Commons Attribution 3.0</i> (CC-BY3.0), por defecto para los contenidos que están en el sitio https://www.data.govt.nz/copyright/ . Los datasets enlazados desde el portal se rigen por la licencia que establezca cada agencia.
6	Holanda	http://data.overheid.nl/	(CC-BY 4.0), a menos que se indique lo contrario. La mayor parte de los contenidos son geodatos (Open Data Trend Report 2015, p. 32).
7	Canadá	http://open.canada.ca/en	<i>Open Government License-Canada</i> , muy parecida a la británica. Licencia mundial, no exclusiva para cualquier propósito. No indica expresamente la compatibilidad con CC-BY y ODC-BY.
8	Noruega	http://data.norge.no/nlod/en/1.0	Licencias compatibles: para toda la información <i>Open Government Data</i> (NLOD). Para información que no constituye bases de datos: <i>Creative Commons Attribution Licence</i> (CC-BY 3.0). Para aquella parte de información que está en base de datos: <i>Open Data Commons Attribution License (version 1.0)</i> (ODC-BY)
9	Dinamarca	http://digitaliser.dk/	Licencia estándar de datos abiertos ("standard license which corresponds to CC-0 and CC-BY"). (https://digitaliser.dk/resource/2432531/artefact/Conditions_for_use_of_Danish_public_sector_data.docx). Corresponde con las licencias CC-0 and CC-BY, dependiendo de si el organismo público exige reconocimiento o no. Aclaración y enlace facilitados por correo electrónico.
10	Australia	http://data.gov.au/	http://data.gov.au/about . <i>Creative Commons Attribution 3.0 Australia licence</i> , por defecto.
13	España	http://datos.gob.es/	http://datos.gob.es/?q=aviso-legal . Asimilable a <i>Creative Commons Attribution</i> ("siempre que se ajusten a lo establecido en este real decreto", RD 1495/2011).

Fuente: <http://www.opendatabarometer.org> (2015) y elaboración propia. Datos consultados/actualizados en octubre de 2016.

bate de si la apertura de datos y la libertad de información son dos aspectos complementarios o en competencia. Sobre este aspecto, diversos análisis realizados en el ámbito europeo ponen de relieve que la apertura de datos repercute en las solicitudes de información en base a las leyes de transparencia, ya que si los gobiernos gestionan bien la información los costes de cada petición bajarán. En el mismo sentido, cuanto mayor es el coste de un servicio público mayor resulta el impacto que produce la transparencia y apertura de datos en ese servicio, según se pone de relieve en el informe holandés (Open Data Trend Report, 2015).

Los resultados revelan algunos aspectos peculiares en cuanto a las licencias y el uso de los datos:

- La licencia del gobierno británico sobre datos abiertos es la piedra angular que ha servido de referencia a todos los gobiernos, como se reconoce ampliamente y se recoge en el cuadro.
- El número de reutilizadores de información pública es muy escaso todavía, tal como ponen en evidencia algunos datos, por ejemplo, en Francia.

- Asimismo, la oferta de datos públicos, aunque crece a buen ritmo, es desigual y la calidad de los mismos es mediana (Francia, Holanda, Dinamarca, España). Como se pone de manifiesto en la excelente página de ayuda que contiene el portal francés, a veces se deben extraer los datos de documentos cerrados (en pdf o de tabulaciones en páginas web).
- Hay muchos organismos públicos que no ofrecen datos abiertos, como también se desprende de los datos manejados.
- En el aspecto positivo, con respecto a hace unos años (Ramos y otros, 2012) hay un notable incremento de buenas prácticas en aspectos tales como la oferta de tutoriales para transformar los datos, grandes mejoras en el uso de herramientas de visualización, así como para la conversión de datos de unos formatos a otros.
- Cada vez son más populares los encuentros para avanzar en la mejora de los datos y su reutilización (*hackatones, opendata camps*) de carácter festivo y competitivo que suelen durar un fin de semana (Francia, Reino Unido, España...).

Como se observa en la Tabla II, la mayoría dispone de una licencia propia elaborada por cada Gobierno que son muy parecidas entre sí (Reino Unido, Francia, Noruega, Dinamarca, Australia, Nueva Zelanda, incluso España), que es compatible con el libre uso de los datos o reconoce la compatibilidad de las condiciones de uso de la información pública con alguna de las licencias libres. Sirva como modelo el portal francés (data.gouv.fr) donde se contextualiza la licencia abierta en el marco de las licencias estándar internacionales y cita expresamente la del gobierno británico y las licencias estándar compatibles (ODC-BY, CC-BY 2.0). Asimismo, esta interpretación es la que cabe hacer en el marco de aplicación de estas licencias a la reutilización de la información del sector público en España. Como acertadamente señalan Clabo y Ramos (2015), estas licencias se adaptan "a cada uno de los usos definidos en la normativa española de reutilización y pueden servir para precisar y acotar el alcance de las condiciones fijadas en el aviso legal o para precisar los distintos usos de aquellos datos y documentos".

Parece interesante señalar la diferencia entre las versiones 3.0 y 4.0 de la licencia Creative Commons. Frente a las versiones anteriores que permitían adaptaciones nacionales y generaban confu-

Tabla II. Modalidades y compatibilidad de licencias libres empleadas en los portales gubernamentales seleccionados

Licencia por defecto	Países	Compatibilidad entre licencias	Volumen de datos en el portal
Sin licencia	Estados Unidos Suecia	-CC-0 y PDDL -(https://theunitedstates.io/licensing/) Datos libres de pago y de licencias	192.261 datasets 300.000 documentos aprox.
CC0 y CC-BY	Dinamarca	-Atribución del titular: CC-BY -Sin Atribución: CC-0	n.d.
CC-BY 4.0	Holanda	CC-BY 4.0	9.052 datasets
OGL	Reino Unido Francia Canadá España	CC-BY 4.0 y ODC-BY ODC-BY y CC-BY 2.0 No explícita No explícita	40.067 datasets 19.558 datasets 121.844 registros 12.021 resultados
CC-BY 3.0	Nueva Zelanda Noruega Australia	No se especifica -No base de datos: CC-BY 3.0 -Sí base de datos: ODC-BY No se especifica	4.387 datasets 757 datasets 23.371 datasets

Fuente: Elaboración propia a partir de los portales de datos mencionados en el cuadro I (octubre de 2016).

sión sobre si quedaba licenciado el uso de las bases de datos no originales, la versión 4.0, lanzada en 2013, considera expresamente incluidos en la licencia tanto la información como los datos protegidos por derechos de autor o por el derecho sui generis sobre bases de datos vigente en la Unión Europea y otros países (<https://creativecommons.org/licenses/by/4.0/legalcode>). Esa es la causa por la que la licencia Noruega excluye la compatibilidad de la licencia Creative Commons para el uso de los datos. Con la versión CC 4.0 se ha tratado de evitar el efecto de las llamadas licencias "portables", adaptadas a cada jurisdicción nacional con diferentes matices lo que impedía una interpretación uniforme en aspectos tan importantes como el de si se consideraban incluidos dentro de la licencia los derechos sobre las bases de datos no originales.

Mención aparte merece la puesta a disposición de la información y datos que hacen los portales de los gobiernos estadounidense y sueco. En la página dedicada al uso de los datos se pone muy de relieve que los datos del Gobierno de Estados Unidos se ofrecen, en dominio público, sin ninguna reserva de derechos, situación preferible a la información licenciada. Al abordar la compatibilidad se indica que las licencias CC0 y PDDL son muy similares, ambas son una renuncia a los derechos sobre los contenidos y allí donde no sea de aplicación la renuncia, permite una plena reutilización. En cambio, "*open licensing* presupone una protección por derechos de autor". Por su parte, en el portal sueco sólo se recoge la mención de que los datos abiertos pueden ser utilizados libre de cargos y sin necesitar licencias.

6. CONCLUSIONES

La investigación pone de relieve que los portales gubernamentales adoptan licencias por defecto compatibles con declaraciones internacionales sobre gobierno abierto o licencias libres estándar de uso internacional (Creative Commons y Open Data Commons) que permiten acceder y usar los datos en distintas modalidades con un mecanismo relativamente sencillo. Muchas de estas iniciativas asumen unos principios básicos que permiten la transformación de los datos, sin más requisitos que la cita de la fuente, e incluso se abren paso las licencias orientadas al libre flujo de datos y metadatos, más próximas al concepto de dominio público, como PDDL y CC0. El uso de estas licencias facilita el libre flujo de datos y su reutilización. Sin embargo, será necesario avanzar también en el uso de metadatos y en la expresión de derechos en Internet, puesto que en este aspecto la asignación de derechos y la combinación de contenidos (por ejemplo, que haya distinta li-

encia para los datos y el texto de un artículo), se convierte en un asunto de cierta complejidad tecnológica, pues hay que aplicar un lenguaje de derechos para cada tipo (REL, por sus iniciales en inglés). (Abelson y otros, 2012)

El concepto tradicional de minería de datos realizado sobre las bases de datos se ha extendido al análisis de textos, datos e imágenes, tratamiento que ahora se realiza sobre datos masivos o macrodatos (big data), de modo que estas actividades son conocidas como análisis de datos o análisis de contenido, fundamentales en todos los sectores sociales y económicos por cuanto son muy eficaces para analizar pautas históricas y, a partir de ahí, diseñar modelos y respuestas.

La gestión de estos grandes volúmenes de datos debe ser puesta en relación con su posible colisión con algunos derechos. Se suele afirmar que salvo en los datos en dominio público, el análisis de datos puede afectar a los derechos de los autores o de los titulares de las bases de datos no originales. Si bien los argumentos en torno a la protección de los datos por su originalidad son bastante débiles, la protección de tales bases de datos por razón de la inversión o el esfuerzo realizado supone que todas las bases de datos cuenten con algún tipo de protección legal, así sucede en todo el ámbito de la Unión Europea.

La cuestión clave en el debate sobre los datos y la autoría es dirimir si los autores deberían estar autorizados a prohibir el uso de una obra para un propósito diferente al de la creación original, en donde no se analiza la creatividad o la forma de expresión de las ideas sino, simplemente, a partir de grandes volúmenes de estos contenidos (grabaciones, novelas...) se obtienen datos de utilidad social y comercial que no están directamente relacionados con la actividad creativa. Por ahora el acceso a estos grandes volúmenes de datos muy demandados, sólo puede hacerse mediante la aprobación de una excepción a los derechos de los autores, o bien mediante el uso de licencias de acceso a dichos datos. El debate está muy condicionado por la evolución constante de la tecnología.

Por su parte, el uso de licencias de información resulta, en principio, una solución más fácil y práctica. Sin embargo, algunos estudios ponen de relieve que en muchos casos no se explicita la vía de acceso a las bases de datos y cuando sí existe no siempre la licencia resulta adecuada a los datos, por lo cual se aboga por una simplificación de las licencias y cesiones de los datos, así como por la formación de usuarios y proveedores, así como por modificar hábitos seculares (Boustany, 2013). Esa misma complejidad la hemos podido comprobar al

analizar los 11 portales gubernamentales, ya que son necesarias búsquedas pormenorizadas en cada sitio web para determinar cuál es la licencia aplicable por defecto, e incluso preguntar directamente en la institución responsable del portal de datos. Asimismo, a veces es muy tortuoso distinguir entre la licencia para usar-reutilizar las bases de datos que opera el portal y sus metadatos de la base de datos o *dataset* que contiene la referencia a los datos que se pretenden descargar o consultar.

Por último, aunque sea desde una perspectiva muy subjetiva, en el caso de los portales gubernamentales de datos abiertos se pone de relieve una clara evolución en estos últimos años desde unos términos de uso con declaraciones muy heterogéneas entre sí a la situación actual, algo más sencilla centrada en el uso de una licencia de datos abiertos por defecto. Pero no perdamos de vista que los conjuntos de datos públicos que se ofrecen a libre reutilización son relativamente pocos en algunos países, como queda claro en la tabla II. El tiempo nos dirá cuántos organismos públicos no se

acogen a la licencia por defecto de datos abiertos y a qué sectores pertenecen. El desconocimiento del impacto de los *big data* podría influir en la oferta de datos públicos.

7. AGRADECIMIENTOS

Este trabajo forma parte de los resultados del proyecto de investigación titulado "Organización del acceso, uso y reutilización de la información del sector público en España. Hacia la consolidación de una industria de la información", financiado por el Plan Nacional de I+D en España (Ref.: CSO2010-17451).

ACKNOWLEDGEMENTS

This study is part of the results of the research project entitled "Organization of access, use and reuse of public sector information in Spain. Towards the consolidation of information industry", funded by the National R + D Plan in Spain (Ref.: CSO2010-17451).

8. NOTAS

- 1 La consultora IDC (2012) define las tecnologías Big Data como una nueva generación de tecnologías y arquitecturas diseñadas para extraer valor económico en muy grandes cantidades de una amplia variedad de datos haciéndolo posible mediante captura a alta velocidad, exploración y/o análisis. Por su parte, Ubaldi (2013), en el estudio de la OCDE sobre datos públicos abiertos, define "Big data" como toda colección de *datasets* tan grande y compleja que es difícil usar herramientas tradicionales de gestión de bases de datos para su procesamiento lo que incluye captura, búsqueda, compartición, transferencia, análisis y visualización.
- 2 La Directiva 2013 ha sido incorporada al Derecho interno español mediante la Ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de información del sector público (BOE, 10/07/2015). En el texto legal se prevé la aplicación de tarifas por el suministro de documentos para su reutilización, "limitándose la misma a los costes marginales en que se incurra para su reproducción, puesta a disposición o difusión", literalmente igual que en el texto de la Directiva.
- 3 La Ley de Libertad de Información de Estados Unidos (FOIA, por sus siglas en inglés), ha tenido varias modificaciones desde su aprobación inicial en 1966. El texto actualizado con las modificaciones de 2016 está disponible en <https://www.justice.gov/oip/freedom-information-act-5-usc-552> (última consulta, 12/10/2016).
- 4 La Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, en su art. 2 a) entiende por a) «datos personales»: toda información sobre una persona física identificada o identificable (el «interesado»); se considerará identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un número de identificación o uno o varios elementos específicos, característicos de su identidad física, fisiológica, psíquica, económica, cultural o social;
- 5 La mencionada Directiva 95/46/CE, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, establece que en los Estados miembros no se podrá ni restringir ni prohibir el uso de los datos personales por la razón de garantizar el derecho a la intimidad en lo que respecta al tratamiento de los datos personales.
- 6 Preferimos este término "licencias de información" en lugar de recursos de información, para referirnos al amplio ámbito de licencias que regulan contenidos informativos sujetos a derechos de autor intercambiados a través de Internet, así lo utiliza en español el profesor Fernández Molina, véanse las referencias bibliográficas correspondientes.
- 7 Según se ha anunciado en septiembre de 2016, esta propuesta de facilitar el acceso a los datos de investi-

gación parece abrirse camino en la Unión Europea (ver propuesta de Directiva (COM (2016)593), presentada en septiembre de 2016: <http://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52016PC0593&rid=1>). La Comisión Europea propone entre las reformas a los derechos de autor, la creación de una excepción, sólo para organizaciones científicas "actuando en interés público", sólo para los contenidos a los que se haya accedido legalmente y cubriría la investigación comercial y no comercial, pero no se aplicaría a las firmas comerciales (*Nature News*, 15/09/2016).

- 8 Remitimos a los apartados del capítulo de esta obra sobre Europea en el que se explican las licencias para acceder al contenido y los metadatos accesibles desde Europea. Al igual que sucede en muchos portales de datos abiertos, es necesario distinguir los metadatos que contiene la base de datos del portal respecto de los contenidos a los que enlaza y que pertenecen a distin-

tas instituciones, con lo cual las condiciones de acceso pueden ser diferentes de unas bases de datos a otras e incluso entre las bases de datos accesibles dentro de la misma institución. Todo lo cual hace compleja la descarga y uso de los datos.

- 9 Ese ha sido el camino seguido por varios países. El primero fue el Reino Unido, país pionero en el uso de licencias en línea por la OPSI (Office Public Sector Information) hasta su desaparición y que después adoptó la licencia Open Government Licensing para ofrecer sus datos públicos, la cual es esencialmente equivalente a CC-BY u ODC Attribution. Esta misma práctica fue seguida por Francia e Italia que también adoptaron los principios de esas licencias libres a su propia licencia nacional (Morando, 2013, 443). España siguió ese mismo camino con la aprobación del RD1495/2011, en vigor, que regula la reutilización de los datos públicos del Gobierno central.

9. REFERENCIAS

- Abelson, H.; Adida, B.; Linksvayer, M.; Yergler, Y. (2012). CC REL: The Creative Commons Rights Expression Language. En: Dulong de Rosnay, M.; Martin, J.C. de (eds.). *The Digital Public Domain*. OpenBook Publishers; Cambridge (R.U.). <http://www.openbookpublishers.com/product/93/the-digital-public-domain-foundations-for-an-open-culture> [último acceso, 04/09/2016].
- Allen, B. (2014). *Access to and use of public sector information: the academic re-user perspective*. Gobierno de Australia: Information policy reports. <https://www.oaic.gov.au/resources/information-policy/information-policy-resources/access-to-and-use-of-public-sector-information-the-academic-re-user-perspective.pdf> [último acceso, 04/09/2016].
- Bitton, M. (2008). Exploring European Union Copyright Policy through the Lens of the Database Directive. *Berkeley Technology Law Journal*, 23 (4), 1411-1470. <http://dx.doi.org/doi:10.15779/Z38TH7W>
- Boustany, J. (2013). Accès et réutilisation des données publiques État des lieux en France. *Les Cahiers du numérique*, 9 (1), 21-37. <https://doi.org/10.3166/lcn.9.1.9-19>
- Cámara-Lapuente, S. (2007). Derecho sui generis sobre las bases de datos. En: Rodríguez Tapia, J.M. (Dir.). *Comentarios a la Ley de Propiedad Intelectual*. Editorial Thomson-Civitas; Cizur Menor (Na).
- Clabo, N.; Ramos, I. (2015). Reutilización de datos abiertos en la administración pública en España y uso de licencias-tipo. *Revista Española de Documentación Científica*, 38(3): e097. <http://dx.doi.org/10.3989/redc.2015.3.1206>
- CODATA (2015) *The Value of Open Data Sharing*. https://www.earthobservations.org/documents/dsp/20151130_the_value_of_open_data_sharing.pdf [último acceso, 20/10/2016].
- Colle, R. (2002). *Explotar la información noticiosa. Data mining aplicado a la documentación periodística*. Dpto. Biblioteconomía y Documentación de la Universidad Complutense, Madrid.
- Comisión Europea (2016). Propuesta de Directiva del Parlamento Europeo y del Consejo sobre los derechos de autor en el mercado único digital. Documento COM 2016 (593).
- Comisión Europea (Comunicado 2/7/2014). *La Comisión insta a los gobiernos a aprovechar el potencial de los macrodatos*. http://europa.eu/rapid/press-release_IP-14-769_es.htm [último acceso, 04/09/2016].
- Comisión Europea (1998). *La Información del Sector Público: Un Recurso clave para Europa. Libro Verde sobre la Información del Sector Público en la Sociedad de la Información*. Documento COM 1998 (585). http://cordis.europa.eu/pub/econtent/docs/gp_es.pdf [último acceso 12/09/2016].
- Correa, C. (2007) *Protección del dominio público: acceso a datos e información científica y su utilización*. Informe de Comité Nacional de Ética en la Ciencia y la Tecnología (CECTE). Argentina.
- Deloitte (2013). *Market Assessment of Public Sector Information*. Department for Business, Innovation and Skills; Londres. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198905/bis-13-743-market-assessment-of-public-sector-information.pdf [último acceso, 19/10/2016].
- Dulong de Rosnay, M.; Janssen, K. (2014). Legal and Institutional Challenges for Opening Data across Public Sectors: Towards Common Policy Solutions. *Journal*

- of *Theoretical and Applied Electronic Commerce Research*, 9 (3), 1-14. <https://doi.org/10.4067/s0718-18762014000300002>
- Fernández-Molina, J.C. (2004a). Licensing Agreements for Information Resources and Copyright Limitations and Exceptions. *Journal of Information Science*, 30 (4), 337-346. <https://doi.org/10.1177/0165551504045853>
- Fernández-Molina, J.C. (2004b). Los límites al derecho de autor en las licencias de información electrónica: la necesidad de una reforma legal. *Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas*, nº 8.
- Fitzgerald, A.M.; Fitzgerald, B.F; Pappalardo, K.M. (2009). The Future of Data Policy. En: Hey, T.; Tansley, S; Tolle, K. (Eds.). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, pp. 201-208. <http://eprints.qut.edu.au/28182/> [último acceso, 04/09/2016].
- GITA White Paper (2005). *Free or Fee: The Governmental Data Ownership Debate*. <http://www.calmapping.org/conferences/2005/ccma2005-GITA-free-or-fee.pdf> [último acceso, 04/09/2016]
- Guadamuz, A.; Cabell, D. (2014). Data mining in UK higher education institutions: Law and policy. *Queen Mary Journal of Intellectual Property*, 4 (1): 3-29. <https://doi.org/10.4337/qmjip.2014.01.01>
- Hand, D.; Mannila, H; Smyth, P. (2001). *Principles of Data Mining*. Massachusetts, USA, The MIT Press Cambridge.
- HM Government (2012). *Open Data White Paper | Unleashing the potential*. http://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf [último acceso, 04/09/2016].
- Hrynaszkiewicz, I.; Cockerill, M. (2012) Open by default: a proposed copyright license and waiver agreement for open access research and data in peer-reviewed journal. *BMC Research Notes* 2012, 5:1. <https://doi.org/10.1186/1756-0500-5-494>
- IDC/Comisión Europea (2012). *Worldwide Big Data Technology and Services*. <https://ec.europa.eu/digital-single-market/en/news/worldwide-big-data-technology-and-services-2012-2015-forecast> [último acceso, 04/09/2016].
- Liang, L. (2004). *The Guide to Open Content Licenses V.1.2*. Piet Zwart Institute; Amsterdam.
- López, A. (2002). *El Impacto de la Protección de las Bases de Datos no Originales sobre los Países de América Latina y el Caribe*. Comité permanente de derechos de autor y derechos conexos. Organización Mundial de la Propiedad Intelectual (OMPI).
- Mayer-Schönberger, V.; Cukier, K. (2013). *Big data. La revolución de los datos masivos*. Turner Publicaciones, Madrid.
- McKinsey Global Institute (2013) *Open data: Unlocking innovation and performance with liquid information*. McKinsey & Company. <http://www.mckinsey.com/business-functions/business-technology/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information> [último acceso, 04/09/2016].
- Melero, R.; Hernández-San-Miguel, J. (2014). Acceso abierto a los datos de investigación, una vía hacia la colaboración científica. *Revista Española de Documentación Científica*, 37(4): e066. <https://doi.org/10.3989/redc.2014.4.1154>
- Mendo Carmona, C.; Ramos Simón, L. F.; Arquero Avilés, R.; Valle Gastaminza, F. del; Botezán, I.; Sánchez, R.; Tejada, C.; Peón, J.L.; Cobo, S.; Sala, A. (2013). Del acceso a la reutilización, del dato al documento: una visión conceptual de la información pública. *Revista Española de Documentación Científica* 36(3): e013. <https://doi.org/10.3989/redc.2013.3.957>
- Mendoza Rocha, M. (2011). Minería de datos en la Web. En: Cacheda Seijo, F.; Fernández Luna, J.M.; Huete Guadix, M., *Recuperación de información. Un enfoque práctico y multidisciplinar*. Ra-Ma; Madrid.
- MHAP (Ministerio de Hacienda y Administraciones Públicas). (2012). *Funcionamiento, Mantenimiento y Actualización del Catálogo de Información Pública. Datos. gob.es*. MHAP. Centro de Publicaciones; Madrid.
- Miller, P.; Styles, R.; Heath, T. (2008). Open Data Commons. A License for Open Data. <http://ceur-ws.org/Vol-369/paper08.pdf> [último acceso, 04/09/2016].
- Moine, J.M.; Haedo, A.S.; Gordillo, S.E. (2011). Estudio comparativo de metodologías para Minería de Datos. En: *XIII Workshop de Investigadores en Ciencias de la Computación*. <http://hdl.handle.net/10915/20034>
- Morando, F. (2013). Legal Interoperability: Making Open (Government) Data Compatible with Businesses and Communities. *JLIS.it*. Vol.4, n.1 (Gennaio/January 2013).
- Murray-Rust, P. (2008). Open data in Science. *Serials Review*, 34 (1), 52-64. <https://doi.org/10.1080/00987913.2008.10765152>
- O'Boyle, N. M.; Guha, R.; Willighagen, E. L.; Adams, S. E.; Alvarsson, J.; Bradley, J.C.; Murray-Rust, P. (2011). Open Data, Open Source and Open Standards in chemistry: The Blue Obelisk five years on". *Journal of Cheminformatics*, 3(37). <https://doi.org/10.1186/1758-2946-3-37>
- OECD/OCDE (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. <http://www.oecd.org/science/sci-tech/38500813.pdf> [último acceso, 04/09/2016].
- OECD/OCDE (2008). *OECD Recommendation of the council for enhanced access and more effective use of public sector information*. [C (2008)36]. <http://www.oecd.org/sti/44384673.pdf> [último acceso, 04/09/2016].
- Open Knowledge Foundation (2012). *Open Data Handbook*. <http://opendatahandbook.org/guide/es/> [último acceso 04/09/2016].
- Open Data Trend Report (2015). Informe presentado al Parlamento de Holanda por *Algemene Rekenkamer* (Tri-

- bunal de Cuentas). http://www.courtsofaudit.nl/english/Publications/Audits/Introductions/2015/03/Open_Data_Trend_Report_2015 [último acceso, 14/10/2016].
- Organización de las Naciones Unidas (ONU), (2013). *Guidelines on open government data for citizen engagement*. <http://workspace.unpan.org/sites/Internet/Documents/Guidelines%20on%20OGDCE%20May17%202013.pdf> [último acceso, 04/09/2016].
- Parlamento Europeo (1995). Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. *Documento DOUE-L-1995-81678*. <http://www.boe.es/buscar/doc.php?id=DOUE-L-1995-81678> [último acceso, 04/09/2016].
- Parlamento Europeo (1996). Directiva 96/9/CE del Parlamento Europeo y del Consejo, de 11 de marzo de 1996, sobre la protección jurídica de las bases de datos. *DOUE-L-1996-80413*. <http://www.boe.es/buscar/doc.php?id=DOUE-L-1996-80413> [último acceso, 04/09/2016].
- Parlamento Europeo (2013). Directiva 2013/37/UE del Parlamento Europeo y del Consejo, de 26 de junio de 2013, por la que se modifica la Directiva 2003/98/CE relativa a la reutilización de la información del sector público. *DOUE-L-2013-81251*. https://www.boe.es/diario_boe/txt.php?id=DOUE-L-2013-81251 [último acceso, 04/09/2016].
- Ramos Simón, L.F. (2014). Declaraciones de Derechos sobre los metadatos y contenidos de Europeana. En Ramos Simón, L.F.; Arquero Avilés, R. (coords.). *Europeana. La plataforma del patrimonio cultural europeo*. Trea. Gijón.
- Ramos Simón, L.F.; Arquero Avilés, R.; Botezán, I.; Cobo Serrano, S.; Sala Jiménez, A.; Sánchez Jiménez, R.; Valle Gastaminza, F. del (2012). De la reutilización de la información del sector público a los portales de datos abiertos en Europa. *BiD: textos universitaris de biblioteconomia i documentació*, núm. 29. <http://bid.ub.edu/29/ramos2.htm> [último acceso 15/10/2016].
- Rebollo, L.; Serrano, M. (2014). *Manual de Protección de datos*. Dykinson, Madrid.
- Rosen, L. (2005). *Open Source Licensing: Software Freedom and Intellectual Property Law*. Upper Saddle River, New Jersey: Prentice Hall Professional Technical Reference.
- The Obama Administration's commitment to Open Government: a status report (2011). https://obamawhitehouse.archives.gov/sites/default/files/opengov_report.pdf
- Triaille, J.P.; Meeûs d'Argenteuil, J.; Francquen, A. (2014). *Study on the legal framework of text and data mining (TDM)*. Unión Europea. http://ec.europa.eu/internal_market/copyright/docs/studies/1403_study2_en.pdf [último acceso 04/09/2016].
- Truyens, M.; Van Eecke, P. (2014). Legal aspects of text mining. *Computer law & security review* 30 (2) 153-170. <https://doi.org/10.1016/j.clsr.2014.01.009>
- Ubaldi, B. (2013). Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. *OECD Working Papers on Public Governance*, No. 22, OECD. <https://doi.org/10.1787/5k46bj4f03s7-en>
- U.S Department of Justice (2016). The Freedom of Information Act 5 U.S.C.§552. <https://www.justice.gov/oip/freedom-information-act-5-usc-552> [último acceso, 8/11/2016].
- Vickery, G. (2012). Review of recent studies on PSI reuse and related market developments. París: Information Economics. Informe para la Comisión Europea (ejemplar mecanografiado). <https://ec.europa.eu/digital-agenda/en/news/review-recent-studies-psi-reuse-and-related-market-developments> [último acceso, 04/09/2016].
- WWW (World Wide Web Foundation) (2015). *Open Data Barometer Global Report (Second Edition), 2015*. <http://opendatabarometer.org/2ndEdition/> [último acceso 04/09/2016].

ESTUDIOS / RESEARCH STUDIES

Estudio métrico de ALFIN en Iberoamérica: de la bibliometría a las altmetrics

Alejandro Uribe-Tirado*, Joel Alhuay-Quispe**

* Escuela Interamericana de Bibliotecología, Universidad de Antioquia, Medellín – Colombia, Red ICONOS.

Correo-e: auribe.bibliotecologia.udea@gmail.com, auribe@correo.ugr.es | ORCID iD: <http://orcid.org/0000-0002-0381-1269>

** Centro de Investigación, Universidad San Ignacio de Loyola, Lima – Perú.

Correo-e: jalhuay@usil.edu.pe | ORCID iD: <http://orcid.org/0000-0002-1903-4687>

Recibido: 04-08-2016; 2ª versión: 19-10-2016; Aceptado: 22-12-2016.

Cómo citar este artículo/Citation: Uribe Tirado, A.; Alhuay-Quispe, J. (2017). Estudio métrico de ALFIN en Iberoamérica: de la bibliometría a las altmetrics. *Revista Española de Documentación Científica*, 40(3): e180. doi: <http://dx.doi.org/10.3989/redc.2017.3.1414>

Resumen: Este trabajo busca aproximarse a la identificación de la presencia, productividad e influencia de autores iberoamericanos sobre alfabetización informacional - ALFIN, y analizar el impacto y uso de sus contribuciones académicas en las plataformas sociales y científicas a través de indicadores bibliométricos y altmétricos. Partiendo de los resultados arrojados en estudios bibliométricos a nivel mundial e iberoamericano sobre ALFIN incluidos en las principales bases de datos y las publicaciones recopiladas en una wiki iberoamericana, se extraen los autores con mayor productividad (55 autores). Luego, a partir de la consulta directa a ocho plataformas científicas (Google Scholar, ResearchGate, Academia.edu, Mendeley, ORCID, IraLIS, E-LIS y EXIT), tres redes sociales (Facebook, Twitter y LinkedIn) y los datos proporcionados por un proveedor comercial (Altmetric.com), se realiza un análisis de indicadores bibliométricos y altmétricos a nivel de autores y publicaciones. Encontramos de manera general, que existe mayor presencia de autores en ResearchGate (58%), Academia.edu (51%) y Google Scholar (49%) que en Mendeley (25%) y ORCID (18%), y que la mayor influencia potencial en plataformas sociales por la cantidad de seguidores radica en Facebook (/ top 10 autores). Además, un análisis con el estadístico rho de Spearman, arroja entre algunas fuentes y plataformas, una correlación baja entre número de citas en Google Scholar y lecturas en Mendeley ($r=382$) y negativa baja para menciones en blogs ($r=0,-237$), Google+ ($r=0,-214$) y Twitter ($r=0,-183$). Finalmente, que tanto la productividad como el impacto-visibilidad se centran en unos determinados autores de ALFIN, y que, para estos autores, y, en general, para las perspectivas métricas asumidas, considerando distintos recursos de medición, existe un impacto positivo en doble vía, de lo bibliométrico a las altmétricas y viceversa.

Palabras clave: Alfabetización informacional; bibliometría; altmetrics; Web social; visibilidad; impacto; acceso abierto; ciencia abierta; ciencia 2.0.

Metric study of information literacy in Latin America: from bibliometrics to altmetrics

Abstract: This study identifies the presence, productivity and influence of Ibero-American authors that write about information literacy (InfoLit). Using bibliometric and altmetric indicators, it seeks to analyze the impact and subsequent use of their scholarly works on social and scientific platforms. Fifty-five authors with the highest productivity were identified, based on the results of bibliometric studies on InfoLit carried out on both an international and Ibero-American scale in searches of major databases as well as publications collected in a Latin American wiki. Subsequently an analysis of bibliometric and altmetric indicators at the author and publication level was carried out, based on the results of searches on eight scientific platforms (Google Scholar, ResearchGate, Academia.edu, Mendeley, ORCID, IraLIS, E-LIS and EXIT), three social networks (Facebook, Twitter and LinkedIn), and data provided by a commercial supplier (Altmetric.com). Overall we found a greater presence of authors in ResearchGate (58%), Academia.edu (51%) and Google Scholar (49%) as opposed to Mendeley (25%) and ORCID (18%). Furthermore, as to social platforms, the greatest potential influence lies with Facebook, due to its high number of followers (/ top 10 authors). In addition, an analysis with the Spearman rho statistic, shows among some sources and platforms, a low correlation between the number of citations in Google Scholar and readings in Mendeley ($r = 382$) and low negative for mentions in blogs ($r = 0,-237$), Google+ ($r = 0,-214$) and Twitter ($r = 0,-183$). In conclusion, both the productivity and the impact-visibility center on specific authors writing about InfoLit, and various measurement resources show that for these authors there is a positive two-way impact from bibliometric to altmetric and vice versa.

Keywords: Information literacy; bibliometrics; altmetrics; social web; visibility; impact; open access; open science; science 2.0.

Copyright: © 2017 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

Sobre Alfabetización Informacional - ALFIN (*Information Literacy - INFOLIT*), se puede decir que hay en la actualidad una amplia literatura publicada que va desde **lo científico** (i.e. trabajos investigativos respecto a este concepto y sus implicaciones, especialmente en una disciplina como la Bibliotecología y Ciencias de la Información); a **lo académico** (i.e. trabajos enfocados a cómo enseñar a los futuros bibliotecólogos y profesionales de la información a ser unos mejores formadores y gestores de programas de ALFIN, para diferentes tipos de usuarios, a partir del desarrollo de competencias informacionales); y a **lo profesional** (i.e. trabajos orientados a compartir diferentes casos de éxito y experiencias concretas de desarrollo de programas, estrategias y actividades de formación en ALFIN, desde distintos tipos de unidades de información, niveles e instituciones educativas y contextos sociales).

Dentro de estas tres tendencias de la literatura existe cierto número de trabajos que, desde una perspectiva histórica, bibliométrica y de lecciones aprendidas, han dado cuenta de gran parte de la producción científica sobre ALFIN en los últimos años. Dichos trabajos han permitido comprender que la ALFIN ha tenido diferentes períodos de desarrollo, tanto en el contexto mundial como en el contexto iberoamericano (Uribe-Tirado, 2013, 2014), y que determinados autores, artículos, revistas y países han sido los de más presencia e influencia.

No obstante, la orientación bibliométrica no permite entender cómo dichos autores, artículos, revistas y países en el contexto de Internet, y específicamente del acceso abierto y la Web social, están teniendo presencia e influencia, o si desde estos lugares digitales son otros autores, artículos, revistas y países los que están teniendo más presencia e influencia, o qué interrelaciones y tendencias se están dando entre lugares tradicionales de la comunicación científica (representados en las grandes bases de datos) y espacios más actuales representados en diferentes herramientas y recursos (plataformas) de acceso abierto y Web social.

1.1. Altmetrics: Métricas complementarias

El concepto que actualmente se tiene sobre *altmetrics* tiene su origen y está interrelacionado con los métodos de análisis y alcance teórico de disciplinas de estudios métricos de la información, con años de tradición como la bibliometría y más recientes como la cibermetría. Estos son transversales y orientadores de este estudio, pero también lo es la dinámica de la comunicación científica actual, desde nuevas perspectivas epistemológicas, disciplinares y prácticas.

Esta dinámica conduce a la necesidad de conocer toda la influencia y transformaciones que Internet y la Web social han generado y seguirán generando en el entorno académico, y es desde allí, que se ha producido la aparición de un concepto como *altmetrics*.

Desde su aparición en 2010, el concepto que se tiene por *altmetrics* (altmétricas, en español) o métricas alternativas o complementarias, se refiere a la creación y estudio de nuevas formas de medición del alcance e impacto de las publicaciones científicas en Internet y, más específicamente, en la *social media*. Estas, originalmente denominadas métricas alternativas, surgieron para proporcionar una visión complementaria al análisis de citas, ocupando un papel en la evaluación de la ciencia (Borrego, 2014), asumidas sea como un área métrica nueva o como una especificidad de la cibermetría, iniciada hace dos décadas, pero a nivel académico (Orduña-Malea y Aguillo, 2014).

Para este trabajo retomando aspectos claves de distintas definiciones^{1,2} entendemos la ***altmetrics* (altmétricas-métricas complementarias)**, en sentido general, como:

El estudio y toma de decisiones en la actividad científica y académica, a partir de la creación y análisis de nuevos indicadores y fuentes de información y medición, basados especialmente en la Web social, que consideran el uso (acceso y descarga), la evaluación por pares (opinión de especialistas e interesados), las citaciones, y las interacciones y recomendaciones (almacenamiento, links, favoritos, conversaciones y comentarios) de diferentes publicaciones y contenidos tanto por canales, formatos y/o herramientas: formales-informales, tradicionales-actuales, físicos-digitales.

Este concepto y las distintas prácticas en su medición están generando diferentes estudios que conllevan a una cada vez más creciente publicación desde esta perspectiva (Ortega, 2015a; González-Valiente y otros, 2016), y la interrelación con otros conceptos y prácticas (Uribe-Tirado, 2015), considerando la mayor preocupación por la formación de los profesores y/o investigadores en este campo (Alfabetización científica) y la importancia que conlleva para la visibilidad personal/profesional y de universidades/centros de investigación (Visibilidad científica)³ en el marco de lo que se está llamando Ciencia 2.0, todo ello en interacción directa con las políticas y reglamentos institucionales sobre el acceso abierto y la cultura de lo *open* (*Open Science*), entre dichos profesores, investigadores e instituciones.

Actualmente se rescata el valor de *altmetrics* como una herramienta complementaria más no

alternativa, en un sentido excluyente, a los métodos de análisis tradicionales de la bibliometría (Melero, 2015; Alonso-Arévalo, 2016) como el análisis de citas (Costas y otros, 2015). Y, aunque originalmente *altmetrics* nace con el objeto de estudio del impacto en medios sociales de los artículos científicos (*almetrics*), nuevos enfoques se orientan al estudio del uso de los indicadores altmétricos más allá del manuscrito para dar lugar a el análisis integrado de datos altmétricos, a nivel de organismos e instituciones académicas y de investigación (Institution-Level-Metrics), de revistas (Journal-Level-Metrics), y de autores o investigadores (Author-Level-Metrics).

1.1.1. Altmétricas a nivel de institución

El uso de datos altmétricos⁴ tomando como objeto de estudio las instituciones académicas y entidades vinculadas a la investigación (de carácter nacional, regional y supranacional), tiene su base de utilidad en los más de 100 indicadores a nivel de autor (Wildgaard y otros, 2014) empleados y propuestos en el ámbito de la bibliometría, y que pueden ser representados en las plataformas sociales.

Si bien, aún son pocos los estudios que proponen el uso de indicadores altmétricos para medir el impacto en la Web social de las instituciones, como instrumentos de evaluación de la ciencia, resulta prometedor observar que actualmente existen varios espacios en línea a partir de los que analizar el impacto a nivel de instituciones, entre ellos, los repositorios; los perfilan como espacios Web institucionales por excelencia para el monitoreo de datos altmétricos a nivel de instituciones (Konkiel y otros, 2015).

Incluso plataformas de gestión de repositorios como DSpace u Open Repository, incluyen módulos para el conteo de número de descargas de archivos y vistas de los documentos (PDF, HTML, etc.), datos considerados dentro del indicador altmétrico de uso.

Otras plataformas científicas como Academia.edu, que permite el conteo del número de investigadores y documentos agrupando a los autores por departamentos dentro de la institución, y ResearchGate, que muestra número de miembros, cantidad de publicaciones y un RG Score total de los miembros. Por otro lado, Altmetric.com, una de las plataformas comerciales proveedoras de datos altmétricos, posee una interfaz de búsqueda (Altmetric Explorer) orientada al monitoreo de altmetrics para instituciones⁵.

1.1.2. Altmétricas a nivel de revista

Al igual que a nivel de instituciones, el uso de los datos altmétricos proporcionados desde plata-

formas científicas (Academia.edu y ResearchGate), y marcadores bibliográficos (Mendeley), permite enfocar un análisis alrededor de las revistas científicas y sus publicaciones. Por ejemplo, Zhao y Wolfram (2015) a partir de un estudio de los perfiles de usuario de revistas en Bibliotecología y Ciencias de la Información, proponen un indicador altmétrico basado en Twitter que serviría como métrica complementaria para conocer el impacto de una revista en esta plataforma social de *microblogging*; o diferentes estudios, como recopilan Uribe-Tirado, Vallejo-Echavarría y Betancur-Marín (2016), han ido realizando análisis de la visibilidad e impacto de una determinada revista desde datos altmétricos en interrelación con datos bibliométricos, e incluso proponiendo un indicador como D/T Metrics (*Descargas-Downloads / Tiempo-Time*), que ya ha comenzado a generar interés y utilización en próximos estudios de las revistas desde esta perspectiva altmétrica.

1.1.3. Altmétricas al nivel de autor

Sin embargo, no solo las instituciones y revistas deben ser capaces de realizar un seguimiento al impacto en la Web de la producción intelectual de sus académicos, sino los investigadores individualmente deben ser capaces de saber hacia dónde va el impacto de sus propias contribuciones (Liu y Adie, 2014) y cómo están siendo utilizadas en la Web social y académica, para así responder a las tendencias de la Ciencia 2.0, de la Ciencia Abierta (Tabla I).

Existen algunas plataformas que incluyen un perfil de investigador para rastrear las altmétricas de sus publicaciones como es el caso de ImpactStory y PlumX. ImpactStory⁶ es una plataforma que proporciona, de manera gratuita, datos altmétricos a nivel de autores con solo registrar un usuario, mientras que PlumX⁷ es una plataforma comercial que rastrea las publicaciones y construye perfiles por investigador.

1.2. Producción bibliométrica sobre ALFIN

La revisión de la literatura sobre ALFIN permitió identificar trabajos previos (Uribe-Tirado, 2013) que han dado cuenta del desarrollo de la alfabetización informacional⁸ desde su origen hasta la primera década del siglo XXI. En lo que va de esta segunda década, la producción ha sido abundante con aportes desde diferentes perspectivas: histórica (Tewell, 2015), geográfica (China: Wang y otros 2011; América Central: González-Valiente, 2015; Brasil-España: Cuevas, 2014; Sudáfrica: Moyo y Mavodza, 2016; Alemania: Piloiu, 2016; etc.), temática (Suri y otros, 2014) y metodológica (Bruce, 2011), entre otros.

Tabla I. Estudios sobre presencia de autores en plataformas de investigación y de la web social

N°	Fuente de datos		Muestra	Estudio
	Bibliométricos	Altmétricos		
1	Google Scholar, WoS, Scopus	LinkedIn, Mendeley, CiteULike	57 conferencistas del STI 2012	Bar-Ilan y otros (2012)
2	WoS	Twitter	447 investigadores WoS	Holmberg y Thelwall (2014)
3	Scopus	Mendeley	100 astrofísicos europeos	Bar-Ilan J. (2014)
4	Google Scholar, WoS	Mendeley, LinkedIn, Slideshare, Twitter	248 autores EPI	Torres-Salinas y Milanés-Guisado (2014)
5	Microsoft Academic Search, Google Scholar	ResearchGate, Academia.edu y Mendeley	7106 investigadores CSIC	Ortega (2015b)
6	Google Scholar, WoS	ResearchGate, Mendeley, Twitter	814 investigadores en Bibliometría	Martín-Martín y otros (2016)
7	Google Scholar	ResearchGate	1400 investigadores colombianos	Aguillo, Uribe-Tirado, López (2016)

La literatura especializada en estudios bibliométricos sobre ALFIN arroja un total de 16 textos, los cuales se clasificaron entre aquellos de alcance mundial (Tabla II) y de Iberoamérica (Tabla III).

No obstante, para los efectos de este estudio, nuestro interés se centró en aquellos textos que han reportado, desde la perspectiva bibliométrica, la situación de ALFIN a partir del 2010, como se detalla en el apartado de Metodología.

1.3. Estudios altmétricos en Iberoamérica

Basados en que sobre ALFIN no hay ningún estudio desde la perspectiva altmétrica y dado que lo contextual es muy importante, como lo evidencia Alperin (2013, 2014), quien indica la necesidad de que estas métricas integren conteos de medios sociales como Wikipedia o Blogs, no sólo de habla inglesa, y de ese modo, construir una métrica más neutral que reduzca el sesgo geográfico, se realizó una indagación sobre qué trabajos han publicados, autores iberoamericanos sobre altmétrica, para tener un marco mayor a este trabajo.

Una revisión en bases de datos considerando el término *altmetrics*⁹ y filtrándolos por la procedencia geográfica de los autores, se encuentra que la gran mayoría de trabajos se centran en España y provienen del Grupo EC3 (Universidad de Granada)¹⁰ y del CSIC (Consejo Superior de Investigaciones Científicas)¹¹; además de Juan Pablo Alperin¹² (Simon Fraser Univer-

sity y PKP, Canadá), cuyos trabajos se centran principalmente en Latinoamérica. A su vez, algunos otros autores iberoamericanos, también han trabajado el tema, directa o indirectamente (Tabla IV), incluso aportando un estado del arte (2005–2015) sobre *altmetrics* (González-Valiente y otros, 2016).

En síntesis, aún los estudios altmétricos para realidades de la comunicación científica desde autores iberoamericanos y desde mediciones centradas en nuestros contextos de comunicación científica (autores, publicaciones, universidades, disciplinas, etc.), son muy incipientes ya que en la lista anterior la mayoría son de una cobertura mundial, sin énfasis en nuestros países, por lo que este artículo pretende aportar tanto a la ALFIN como a la *altmetrics* en Iberoamérica.

Por tanto, como síntesis, este trabajo busca aproximarse a la identificación de la presencia, productividad e influencia de autores iberoamericanos sobre ALFIN, y analizar el impacto y uso de sus contribuciones académicas en las plataformas sociales y científicas a través de indicadores bibliométricos y altmétricos; ya que como indica el estudio de Ortega (2015), la presencia y el grado de visibilidad e impacto, en dichos espacios Web, depende de los contextos, disciplinas y nos atreveríamos a decir, hasta líneas de investigación, como es el caso del presente estudio alrededor de la ALFIN.

Tabla II. Estudios bibliométricos sobre producción científica de ALFIN en el mundo

Nº	Autor	Título	Fuente	País*	Año
1	Adithya Kumari, H., Madhusudhan, C. M., y Ali, H.	A bibliometric study of world research output on information literacy in the field of library and information science during 1999-2013.	e-Library Science Research Journal	India	2015
2	Majid y otros	Analyzing publishing trends in information literacy literature: A bibliometric study	Malaysian Journal of Library & Information Science	Malaysia	2015
3	Pinto, M.	Viewing and exploring the subject area of information literacy assessment in higher education (2000-2011)	Scientometrics	Hungría	2015
4	Pinto, M.; Escalona, M.I.; Pulgarín, A.	Viewing information literacy concepts: a comparison of two branches of knowledge	Scientometrics	Hungría	2014
5	Hsieh, P. N., Chuang, T. M., y Wang, M. L	A Bibliometric Analysis of the Theses and Dissertations on Information Literacy Published in the United States and Taiwan	Proceedings of the International Computer Symposium ICS 2012 Held at Hualien, Taiwan, December 12-14, 2012	Taiwan	2013
6	Panda, I., Maharana, B., y Chhatar, D. C.	The Journal of Information Literacy: A Bibliometric Study	International Journal of Scientific and Research Publications	India	2013
7	Pinto, M., Escalona-Fernández, M. I., y Pulgarín, A.	Information literacy in social sciences and health sciences: a bibliometric study (1974-2011)	Scientometrics	Hungría	2013
8	Sproles, C.; Detmering, R.; Johnson, A.M.	Trends in the literature on library instruction and information literacy, 2001-2010	Reference Services Review	USA	2013
9	Taşkın, Z.; Doğan, G.; Şencan, I.	Analyzing the Intellectual Structure of World Information Literacy Literature through Citations and Co-Citations	Worldwide Commonalities and Challenges in Information Literacy Research and Practice	USA	2013
10	Uribe-Tirado, Alejandro and Pinto, María and Gómez Díaz, Raquel and Cordón, José Antonio	La producción científica internacional sobre competencias informacionales e informáticas: tendencias e interrelaciones.	Revista Información, Cultura y Sociedad	Argentina	2011
11	Park, M. K., y Kim, H. J.	A Bibliometric Analysis of the Literature on Information Literacy	Journal of the Korean Society for information Management	Corea del Sur	2011
12	Aharony, Noa	Information literacy in the professional literature: an exploratory analysis	Aslib Proceedings: New Information Perspectives	USA	2010
13	Pinto, M.; Cordón, J.A.; Gómez Díaz, R.	Thirty years of information literacy (1977-2007): A terminological, conceptual and statistical analysis	Journal of Librarianship and Information Science	UK	2010

* Se considera como país al lugar de publicación de la revista.

Tabla III. Estudios bibliométricos sobre producción científica de ALFIN en Iberoamérica

Nº	Autor	Título	Fuente	País*	Año
1	Ponjuán, G., Pinto, M. y Uribe-Tirado, A.	Conceptualización y perspectivas de la alfabetización informacional en Iberoamérica: un estudio Delphi	Information Research	Suecia	2015
2	Pinto, M.; I Escalona, M.I.; Pulgarín, A.; Uribe-Tirado, A.	The scientific production of Ibero-American authors on information literacy (1985–2013)	Scientometrics	Hungría	2015
3	González-Valiente, C. L.	La investigación cubana sobre alfabetización informacional: un análisis de su literatura publicada.	Revista Cubana de Información en Ciencias de la Salud	Cuba	2014

* Se considera como país al lugar de publicación de la revista.

2. METODOLOGÍA

2.1. Método

Estudio observacional y descriptivo, de carácter transversal, que analiza la presencia, productividad e influencia de autores iberoamericanos sobre ALFIN y el impacto y uso de sus contribuciones académicas en las plataformas sociales y científicas a través de indicadores bibliométricos y altmétricos.

2.2. Muestra

Se analiza a los autores más productivos en materia de alfabetización informacional de Latinoamérica, España y Portugal: 55 autores, identificados a partir de los datos consignados en estudios bibliométricos previos y en una wiki sobre ALFIN en Iberoamérica.

2.3. Recolección de datos

Como **primer paso**, se identificaron estudios bibliométricos previos sobre ALFIN con fecha de publicación a partir del 2010, año en que se publicó el Manifiesto Altmetrics⁴. Los documentos fueron recuperados de bases de datos comerciales (Web of Science, Scopus, Wiley, Emerald) y repositorios abiertos (Dialnet, E-lis, Scielo, RedAlyc), lo que permitió ubicar un total de 16 estudios: 13 de alcance mundial y 3 iberoamericanos (Tablas II y III). En los estudios de alcance mundial, se buscó si eran mencionados autores de países iberoamericanos, y para el caso Iberoamericano, al ser textos directamente relacionados con el contexto de interés, se tomaron los que se encontraron. Finalmente, de ambas fuentes se identificaron un total de 109 autores con duplicidad de nombres, que fueron luego depurados.

Un **segundo paso**, se centró en la búsqueda en la Wiki Alfin/Iberoamérica (<http://alfiniberoamerica.wikispaces.com>) para obtener los datos de autores (nombres, países) y sus trabajos publicados (artículos, tesis, libros, ponencias), lo que arrojó un total de 187 autores. Luego se excluyó a autores con menos de cuatro trabajos incluidos en la wiki, teniendo en cuenta que por el alcance de este estudio se buscaba identificar los de mayor trayectoria.

Como **tercer paso**, a partir de los dos listados, se excluyeron los datos de autores duplicados y se procedió a normalizar los nombres. Con el listado integrado y depurado (55 autores) se procedió a buscar a cada autor en las siguientes plataformas:

- Buscador académico: Google Scholar.
- Redes sociales académicas: ResearchGate y Academia.edu.
- Redes sociales y profesionales: Facebook, Twitter y LinkedIn.
- Marcador bibliográfico: Mendeley.
- Identificadores de autor: ORCID, IraLIS.
- Repositorio especializado: E-LIS.
- Directorio especializado de investigadores: EXIT.

La recolección de los datos de los autores en los medios sociales se llevó a cabo durante el mes de marzo del 2016.

Tras estos tres pasos, para completar la recopilación y el análisis, se realizó un **cuarto paso**, a partir de la interfaz API de CrossRef (<http://api.crossref.org>) y mediante consultas con los nombres de los 55 autores (`works?query={nombre%de%autor}`) se recuperaron 372 publicaciones. A partir de es-

Tabla IV. Estudios relacionados con altmetrics por autores de Iberoamérica

Título	Autores	Año	País-Autor*
A review of altmetrics as an emerging discipline for research evaluation	González-Valiente, C. L., Pacheco-Mendoza, J.; Arencibia, J.	2016	Cuba
Availability of digital object identifiers (DOIs) in Web of Science and Scopus.	Gorraiz, J., Melero-Fuentes, D., Gumpenberger, C., y Valderrama-Zurián, J.	2016	España
¿Qué papel juegan los bibliotecarios en las altmetrics?	González-Fernández-V, et al	2015	España
Percepciones, políticas y formación hacia la visibilidad académica y científica entre profesores-investigadores de la Universidad de Antioquia	Uribe-Tirado, A.	2015	Colombia
Challenges of library and information science journals: editor's opinions. A survey approach.	Abadal, E.; López-Borrull, A. y Ollé-Castellà, C.	2015	España
El papel de las bibliotecas en la evaluación de la investigación. Indicadores bibliométricos y altmétricos	García-Romero, A. et al	2015	España
Criterios, clasificaciones y tendencias de los indicadores bibliométricos en la evaluación de la ciencia	Peralta González, M.J. et al	2015	Cuba
Marketing científico digital e métricas alternativas para periódicos: da visibilidade ao engajamento	Ferreira de Araújo, R.	2015	Brasil
Traditional and alternative metrics: The full story of impact	Donato, H.	2015	Portugal
Managing academic profiles on scientific social networks	Campos, F., Valencia, A.	2015	España
Presencia de las universidades españolas en las redes sociales digitales científicas: Caso de los estudios de comunicación	González-Díaz, C., Iglesias-García, M., y Codina, L.	2015	España
Are the most highly cited articles the ones that are the most downloaded? A bibliometric study of IRRODL.	Martínez, R. A., y Anderson, T.	2015	Cuba
Altmetrics – A complement to conventional metrics.	Melero, R.	2015	España
Citas, revistas y autores: claves para la evaluación de la producción científica en ciencias de la salud	Medino Muñoz, J.; Estrada Lorenzo, J.M.	2015	España
Altmétria: métricas alternativas do impacto da comunicação científica	Souza, I.	2014	Brasil
Alfabetización en Comunicación Científica: Acreditación, OA, redes sociales, altmetrics, bibliotecarios incrustados y gestión de la identidad digital.	Alonso-Arévalo, J.	2014	España
Challenges to the publishers of humanities in scientific journalism and social networks: Reflections and experiences	Benchimol, J. L., Cerqueira, R. C., y Papi, C.	2014	Brasil
A webometric analysis of travel blogs and review hosting: The case of catalonia.	Marine-Roig, E.	2014	España
Edición y comunicación científica: evolución y tendencias actuales.	Fresco-Santalla, A.	2013	España

* Se considera el país de filiación firmada por el primer autor.

tas 372 publicaciones con DOI, se buscaron los datos altmétricos desde la plataforma más reconocida para la obtención de estos datos, como es Altmetrics.com.

2.3.1. Técnicas

Se emplearon dos técnicas: 1) revisión de la literatura científica para la identificación de los autores con producción científica sobre estudios bibliométricos en ALFIN, 2) revisión de datos secundarios, a partir de una wiki, bases de datos científicas, y plataformas sociales y científicas, para identificar los datos bibliométricos y altmétricos sobre la producción científica de los autores.

2.3.2. Instrumentos

Los instrumentos de recolección de datos empleados en el estudio se distinguieron según dos fines: bibliométricos y altmétricos. Para la recuperación de los datos altmétricos, se recuperaron directamente desde plataformas sociales y científicas, y en algunos casos, fue necesario emplear recursos de terceros, aunque se prefirieron los directos, por dar datos más precisos e incluyentes de cada uno de esos 55 autores seleccionados:

- Bibliométricos: Los datos bibliométricos de los perfiles de autores registrados en Google Scholar se recuperaron empleando directamente el buscador académico, y del mismo modo se procedió con ResearchGate y Academia.edu. Para corroborar el número de citas recibidas por cada artículo se empleó el software *Publish or Perish* (Harzing, 2007).
- Altmétricos: Para identificación de los perfiles de los autores en ResearchGate, Academia.edu, ORCID, Facebook, LinkedIn y Mendeley, se empleó Google usando el comando de consulta *site:* y comillas (") como delimitador de cadena de caracteres. En el caso de Twitter, para recuperar los indicadores de *tweets*, *following*, *followers* se empleó la herramienta *Search Bios Follower Wonk* de MOZ (<https://moz.com/followerwonk>). Los datos altmétricos fueron provistos por la plataforma comercial (Altmetric.com).

2.4. Análisis de datos

Los instrumentos de análisis de datos que se emplearon fueron el software de cálculo MS Excel (Figura 1) y el paquete estadístico IBM SPSS.

Se identificaron indicadores bibliométricos y altmétricos a nivel de autor, por lo que las métricas adoptadas se consideran desde estas dos cate-

gorías (Ortega, 2015b). Es decir, el grupo de indicadores bibliométricos: número de citas, cantidad de documentos e índice H; sumado con los indicadores de uso y sociales: cantidad de lecturas, vistas de perfil, número de seguidores, que permitieron el análisis a partir de dos enfoques: los autores y sus publicaciones.

3. RESULTADOS

3.1. Análisis de presencia, productividad e influencia de autores

3.1.1. Procedencia de los autores

La procedencia geográfica de un autor se determinó a partir de la filiación indicada por cada autor en dos fuentes: los datos de correspondencia de las publicaciones, y la institución de pertenencia consignada en las plataformas sociales y científicas. En ese sentido, un análisis para conocer la procedencia de la muestra final de 55 autores, permitió observar que la mayor procedencia geográfica de los autores de ALFIN estudiados, corresponde a España, Brasil y México, con el 29% (16), 19% (10) y 15% (18), respectivamente (Figura 2).

3.1.2. Presencia en las plataformas sociales y científicas

En la Figura 3 se observa que la presencia y ausencia en las distintas fuentes consultadas varía teniendo en los extremos a ResearchGate como la plataforma con mayor presencia (32 autores) y ORCID como la de menor preferencia (10 autores), lo que da cuenta de que los autores más reconocidos en la publicación sobre ALFIN, deben realizar una mayor gestión de su visibilidad en Internet (Identidad digital académico-científica). Además, resulta llamativo identificar que, siendo autores con recorrido y reconocimiento, no estén aprovechando todas las herramientas y plataformas clave de la Ciencia 2.0, especialmente Mendeley y ORCID, que como presenta la Figura 3 alcanzan una presencia menor que las otras, y que en otros estudios se ha reconocido su potencial para la visibilidad y posteriores impactos (citas y reconocimiento académico).

Por otro lado, para el caso de los espacios de redes sociales, los resultados muestran que existen más autores con un perfil de usuario creado en LinkedIn (21) y en Facebook (31) que en Twitter (11), pese a que esta plataforma de *microblogging* es considerada como la de mayores posibilidades de interacción para fines académicos. Para efectos de este estudio, ante estos datos y la preferencia por las plataformas científicas, para este momen-

Figura 1. Tabla para el registro y análisis de la información de cada autor y plataforma

Nº	NOMBRE	AMIENTA RESEAR	IG Score	H-Ind	H-index sin autotitac	Percentil e	PUBLICACION	Lectur	Citacion	Impact Poir	texto compl	Vistas de perff
1	Angulo Marcial, Noel	SI	1,82	1	1	5%	9	48	4	0	7	152
2	Area Moreira, Manuel	SI	11,63	11	10	47,50%	81	14087	372	2,66	42	732
3	Barbosa Chacón, Jorge Winston	NO										
4	Belluzzo, Regina Célia Baptista	NO										
5	Calderón-Rehecho, Antonio	SI	3,03	1	1	12,50%	17	279	4	0,9	5	145
6	Calzada Prado, Francisco Javier	NO										
7	Campal-García, María-Felicidad	NO										
8	Campello, Bernadete	NO										
9	Castañó-Muñoz, Wilson	SI	0	0	0	0%	5	38	7	0	5	123
10	Ceretta-Soria, María Gladys	SI	1,7	0	0	5%	3	20	0	0,47	1	35
11	Ramalho Correia, Ana Maria	NO										
12	Cuevas-Cerveró, Aurora	SI	9,07	4	3	40%	54	1003	50	1,52	36	165
13	Domínguez-Aroca, María-Isabel	SI	1,41	2	1	2,50%	17	254	7	0,33	10	130
14	Vinciane Doucet, Anne	SI	11,85	7	6	47,50%	13	293	84	10,85	4	89
15	Adriana Dudziak, Elisabeth	SI	5	4	4	25%	50	1978	55	0,49	46	191
16	Durban Roca, Glòria	NO										
17	Fernández Marcial, Viviana	SI	9,04	4	4	40%	38	962	69	5,17	17	501
18	Fernández Valdés, María de las M.	SI	1,05	3	3	2,50%	21	273	27	0	11	90
19	Fialho, Janaina Ferreira	NO										
20	Freire, Isa Maria	NO										
21	Gonçalves Dias Gasque, Kelley Cristine	SI	8,72	2	2	37,50%	24	266	22	0,41	17	68
22	Gómez-Hernández, José-Antonio	SI	12,98	6	6	52,50%	135	1669	165	1,75	125	313
23	González Fernández-Villavicencio, Nieves	SI	9,08	3	3	40%	36	783	27	2,7	23	666
24	Hernández-Hernández, Carmen-Julia	NO										
25	Hernandez Salazar, Patricia	SI	6,29	4	4	30%	18	299	48	1,31	12	98
26	Lau, Jesús	SI	8,03	4	4	35%	35	1196	140	2,72	16	159
27	Licea de Arenas, Judith	SI	14,97	8	7	57,50%	79	668	167	12,64	5	189
28	Machin-Mastromatteo, Juan D.	SI	13,76	2	1	55%	43	934	28	4,42	34	483
29	Pinto, María	SI	28,5	14	12	85%	141	4,09	619	56,12	45	804
29	Malheiro da Silva, Armando	SI	7,91	4	4	35%	23	209	88	7,09	1	369
30	Marciales-Vivas, Gloria Patricia	SI	7,08	5	5	32,50%	17	461	56	2,32	10	115
31	Martí Lahera, Yohannis	SI	7,16	3	3	32,50%	23	1015	36	0,83	10	277

Figura 2. Procedencia geográfica de los autores

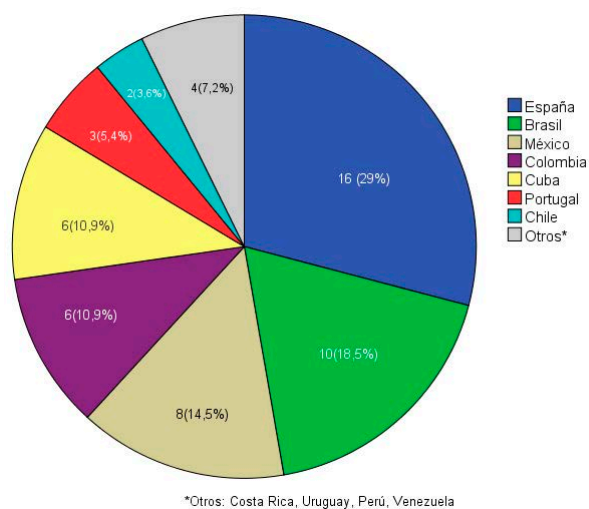
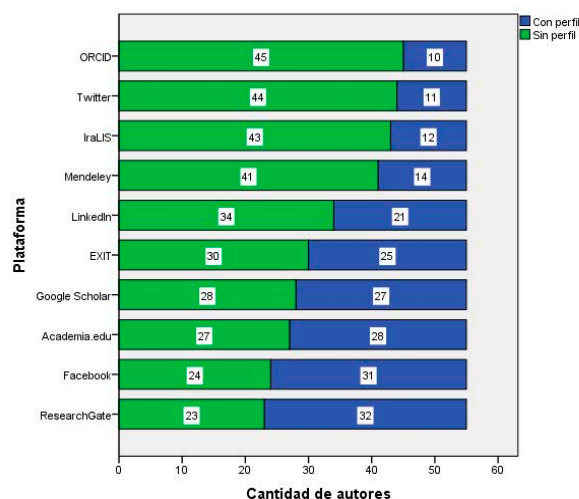


Figura 3. Presencia de autores iberoamericanos que publican sobre ALFIN en plataformas sociales y de investigación



to investigativo, se prefirió profundizar más en las que hemos denominado plataformas científicas.

3.1.2.1. Autores con mayor presencia

Cuando se integran los datos de los autores que fueron listados desde estas distintas plataformas, y se organizaron los datos en orden descendente, identificando los *top 10* en cada fuente, se identi-

có que de los 55 autores seleccionados para el estudio, 35 aparecen en alguno de los top 10 de esas fuentes, por lo tanto, 20 autores no tendrían ninguna o escasa presencia significativa en las fuentes consultadas. Es decir, se hace evidente que son un número específico de autores los que tienen mayor presencia y/o gestión de esa identidad digital académica-científica en Internet (Figura 4).

3.1.3. Productividad de autores

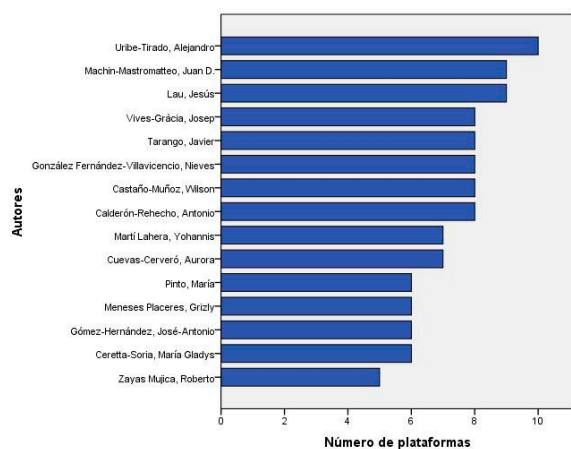
El tener presencia en una plataforma social o científica no implica necesariamente que el autor está gestionando su presencia en dichas fuentes, es decir, el estar incluyendo la mayor cantidad de sus producciones y aprovechando su potencial como vías para la visibilidad y posterior impacto. La Figura 5 muestra que la cantidad de publicaciones incluidas por los autores estudiados, según las seis plataformas útiles para esos fines, se centra en Google Scholar y ResearchGate.

En la Figura 6 observamos que las plataformas científicas con mayor preferencia para incluir sus publicaciones académicas por los 10 autores más prolíficos son Google Scholar y ResearchGate. Sin embargo, resulta importante destacar que los tres primeros autores de cada una de estas dos plataformas, no aparecen en ambas, sino que son autores distintos, lo que supone dos escenarios: la diferencia en las preferencias o la facilidad de uso de la plataforma.

3.1.4. Influencia de los autores

Para determinar la influencia que poseen los usuarios en las plataformas sociales, uno de los indicadores que se toman en consideración es la cantidad de seguidores (Cha y otros, 2010; Bakshy y otros, 2011; Xia y otros, 2016). De las plataformas empleadas para el análisis en este estudio, solo 6 presentan datos sobre *followers*, 3 de los cuales de visibilidad científica y 3 de redes sociales.

Figura 4. Autores iberoamericanos sobre ALFIN con presencia en plataformas de visibilidad investigativa y espacios sociales



Ya que no todos los autores analizados poseen presencia en las 9 fuentes estudiadas, un análisis a partir de un diagrama de cajas sobre los 10 autores con más seguidores en las 6 plataformas, nos muestra un mayor recorrido intercuartílico de la cantidad de seguidores en Facebook frente a los seguidores en Mendeley (Figura 7).

3.2. Análisis de impacto y uso de los artículos

3.2.1. Impacto de las publicaciones

Para conocer el impacto de las publicaciones de los autores analizados a partir de los indicadores de citas e índice H, se presentan los datos de estos autores (que alcanzaron a estar en la mayoría de los *top 10*) respecto a dos fuentes que ofrecen ambos datos métricos: Google Scholar y ResearchGate.

Un primer análisis según el índice H proporcionado por las dos plataformas, permite observar que Google Scholar reporta mayor impacto de las publicaciones frente a ResearchGate; sin embargo, a diferencia de lo que se indicó en párrafos anteriores, se muestra una incidencia de dos de los tres primeros autores en ambas plataformas, lo que presenta a Manuel Área Moreira y María Pinto como los autores con más impacto según su índice H en ResearchGate y Google Scholar (Figuras 8a y 8b), aunque es importante indicar que para el primero, su línea de investigación es más amplia (educación), mientras el segundo se centra en la ALFIN como línea principal de investigación, lo cual hay que considerar al pensar en la cantidad de citas.

Figura 5. Cantidad de publicaciones según plataformas

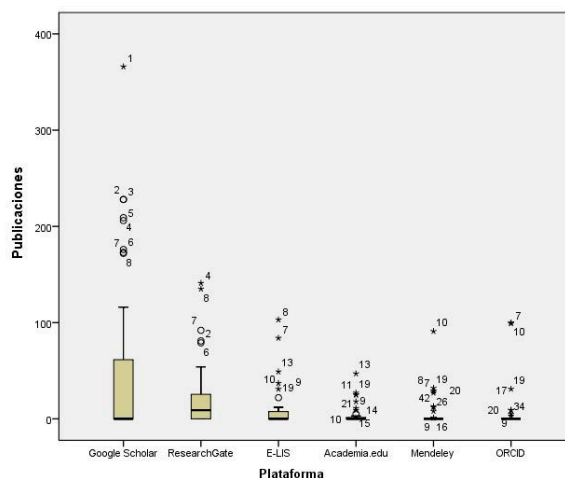
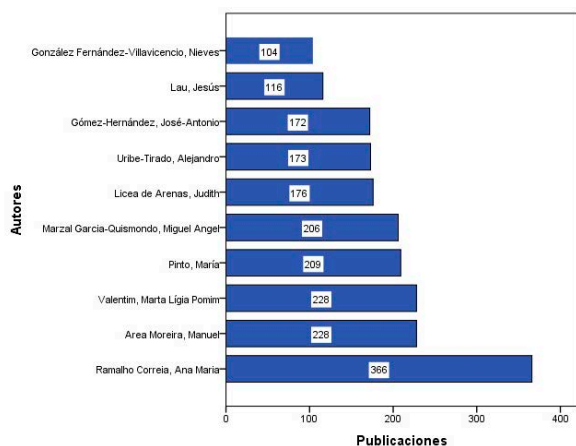
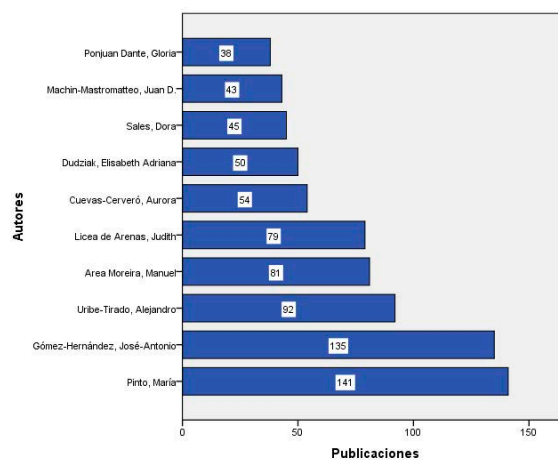


Figura 6. Autores iberoamericanos con más publicaciones en plataformas de visibilidad investigativa

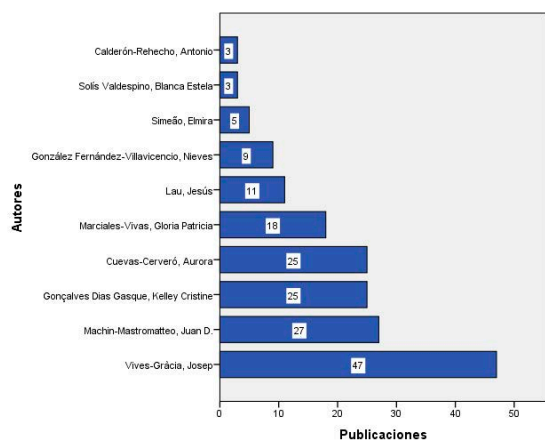
6a. Top 10 autores con más publicaciones en Google Scholar



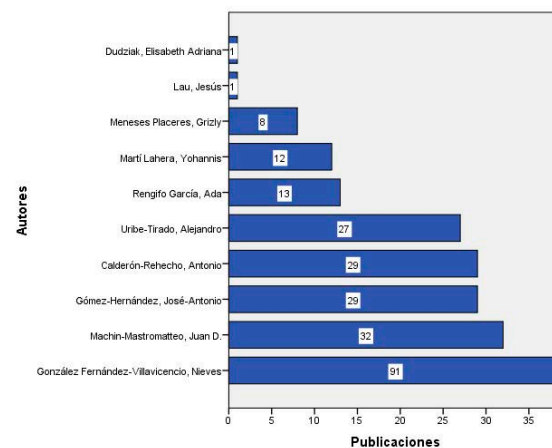
6b. Top 10 autores con más publicaciones en ResearchGate



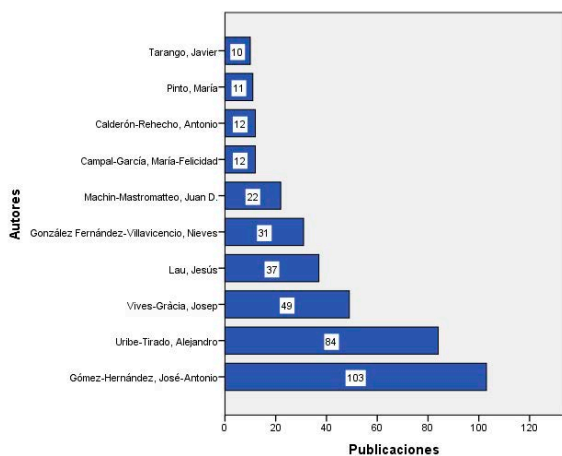
6c. Top 10 autores con más publicaciones en Academia



6d. Top 10 autores con más publicaciones en Mendeley



6e. Top 10 autores con más publicaciones en ELIS



6f. Top 10 autores con más publicaciones en ORCID

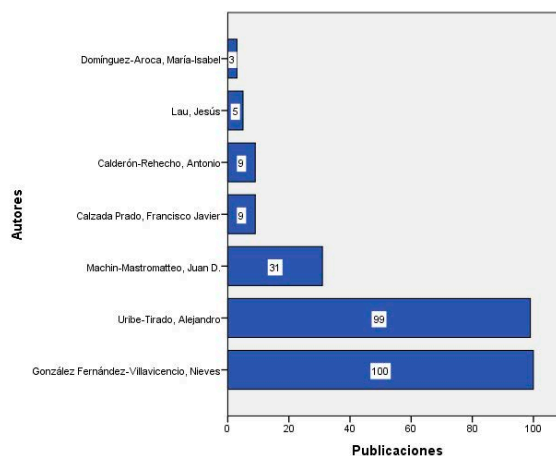


Figura 7. Influencia en plataformas sociales de 10 autores con más seguidores

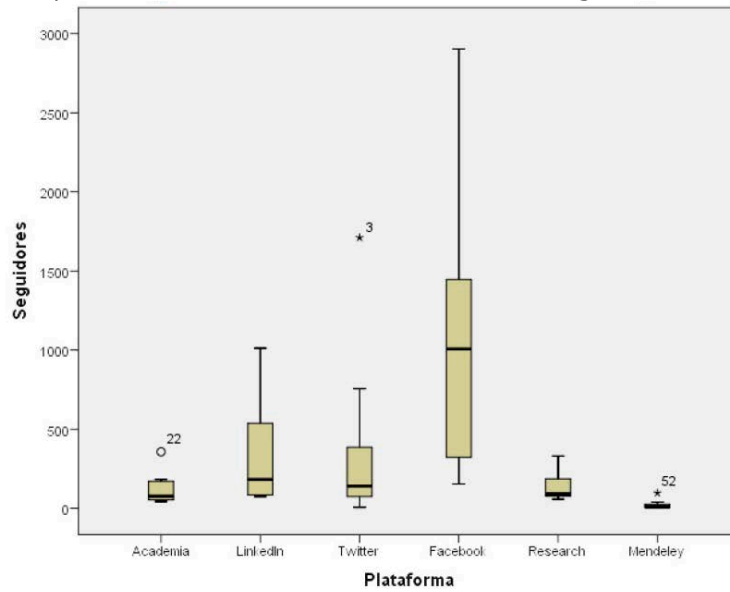
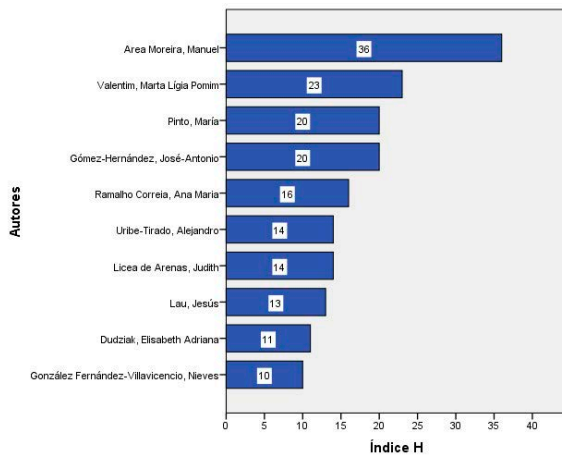
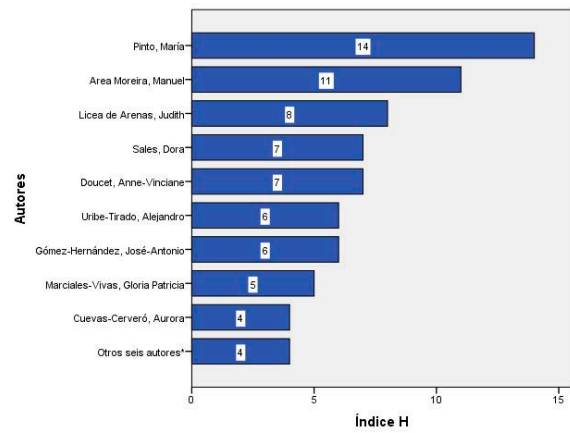


Figura 8. Autores con mayor impacto en Google Scholar y ResearchGate

8a. Top 10 autores con mayor Índice H de Google Scholar

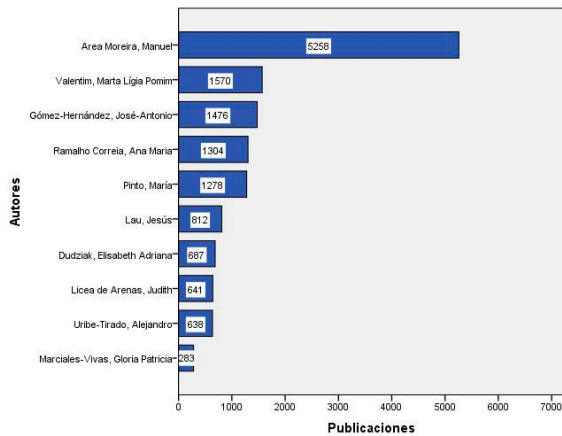


8b. Top 10 autores con mayor Índice H de ResearchGate

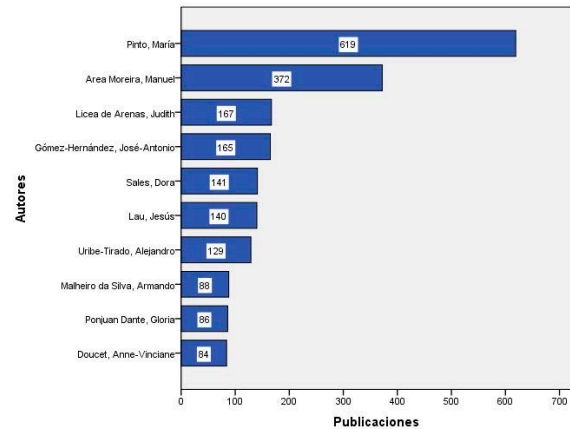


*Seis autores restantes: Dudziak, Elisabeth Adriana, Fernández Marcial, Viviana, Hernández Salazar, Patricia, Lau, Jesús; Malheiro da Silva, Armando; Ponjuan Dante, Gloria

8c. Top 10 autores más citados en Google Scholar



8d. Top 10 autores más citados en ResearchGate



Un segundo análisis a partir de la cantidad de citas recibidas, en las mismas dos plataformas analizadas para el índice H, muestra mayor impacto en Google Scholar para los 10 autores más citados frente a los datos encontrados para ResearchGate (Figuras 8c y 8d), lo que corrobora lo encontrado en el análisis anterior.

3.2.2. Uso de las publicaciones

Los indicadores de uso actualmente entendidos como lectura, vistas y descargas son datos ofreci-

dos por dos plataformas analizadas (ResearchGate y Mendeley), por lo que un análisis adicional se centró en comparar el indicador de uso (lecturas) frente al de impacto (citas) para los autores con publicaciones en las dos plataformas mencionadas (Figura 9 y 10).

La interfaz de búsqueda de altmetric.com (Altmetric Explorer) permitió recuperar de manera preliminar las publicaciones sobre ALFIN con mayor altmetric score (Figura 11).

Figura 9. Autores con mayor cantidad de citas y lecturas en ResearchGate

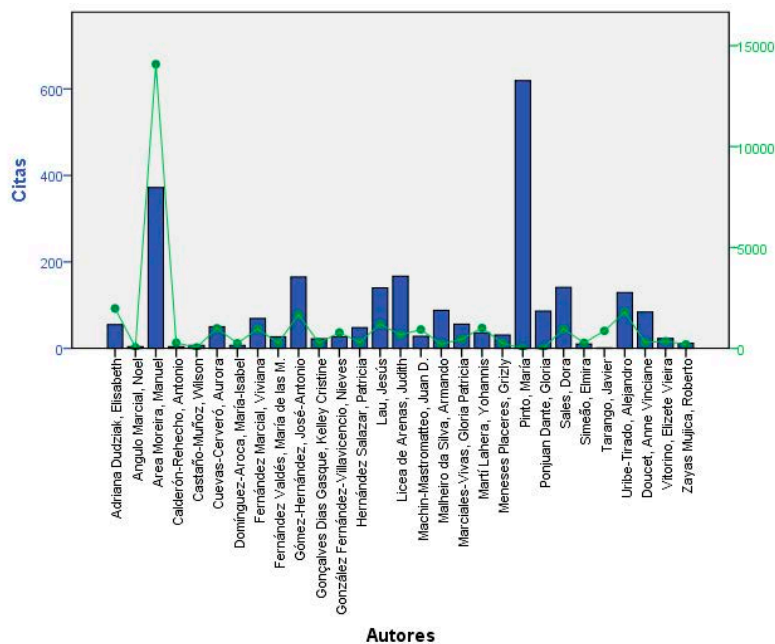


Figura 10. Autores con mayor cantidad de citas y lecturas en Mendeley

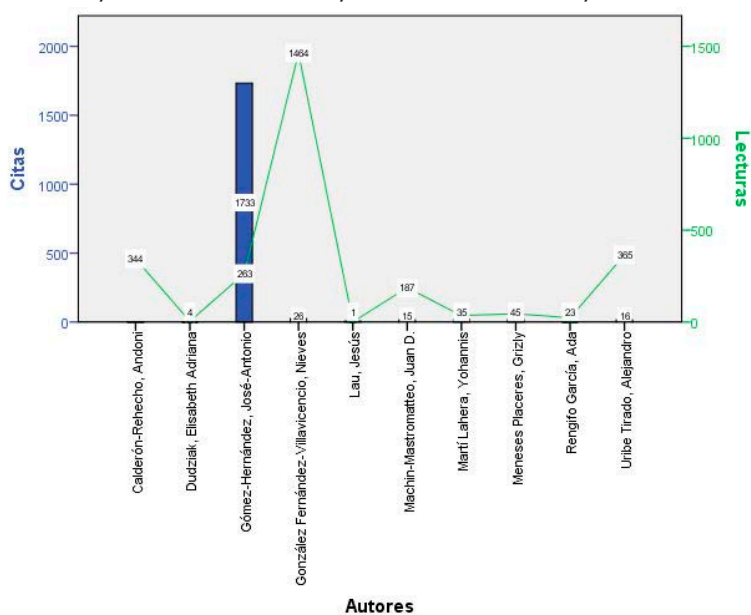
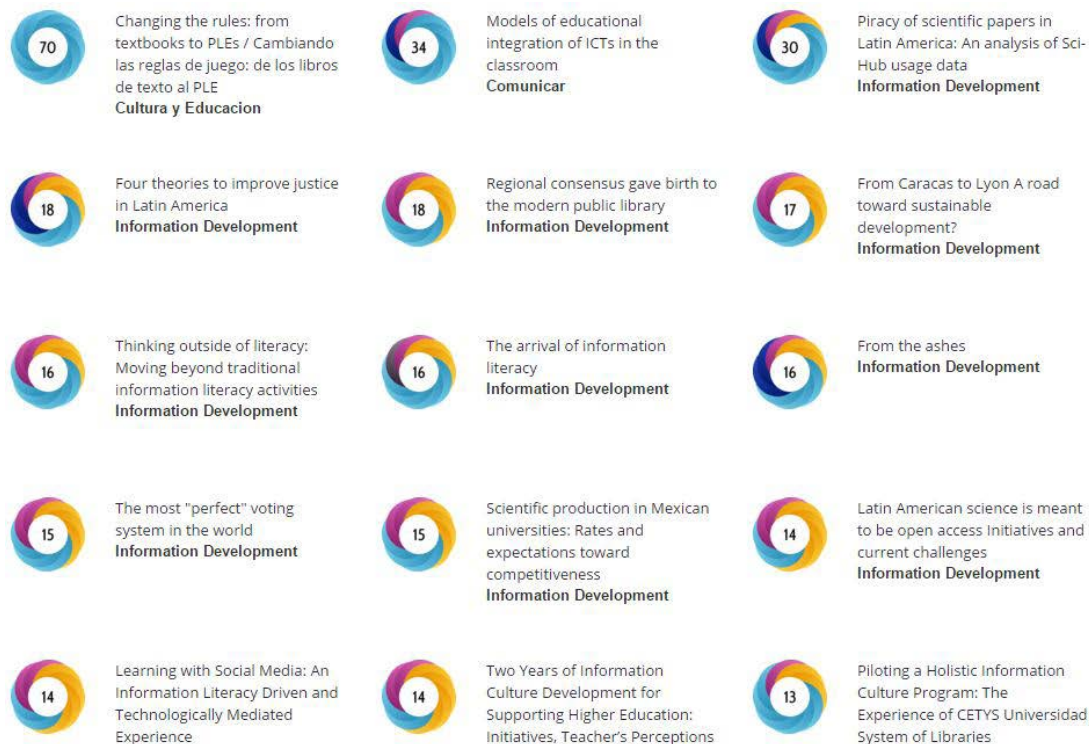


Figura 11. Interfaz Altmetric Explorer con los manuscritos de mayor altmetric score



3.2.3. Correlación altmétrica y bibliométrica

Un análisis adicional, como se mencionó, indicaba que sólo alrededor del 14 % (52) de los manuscritos con DOI, del total de 372 publicaciones, contaba con al menos una actividad altmétrica. Mediante el estadístico rho de Spearman, se evidenció además, de manera específica para algunas fuentes y plataformas (Tabla V), que existe una baja correlación entre el número de citas recibidas en Google Scholar y lecturas en Mendeley (r=382), y negativa, relativamente baja, para menciones en blogs (r=0,-237), Google+ (r=0,-214) y Twitter (r=0,-183).

Esto implica, como han evidenciado otros estudios, que hay que seguir indagando, para encontrar relaciones directas entre distintas fuentes y plataformas, bibliométricas y altmétricas, pero en ello, serán claves dos procesos: 1) la formación de los autores-investigadores para la creación y gestión de una identidad digital y, (2) que las publicaciones deban contar con DOI, lo cual como indicó Alperin (2015), en nuestros contextos es prioritario, ya que solo hasta hace muy poco, se está convirtiendo en una práctica presente en cada revista para sus artículos, y por eso, la invisibilidad también en el contexto altmétrico cuando se utilizan los recur-

Tabla V. Correlación entre citas e indicadores bibliométricos

	ALM score	Citas en GS	Blog	Twitter	Facebook	Wikipedia	Google+	Mendeley
ALM score	1,000	-,091	,733**	,839**	,013	,156	,735**	-,004
Citas en GS		1,000	-,237	-,183	,003	-,038	-,214	,382**
Blog			1,000	,513**	-,074	,037	,786**	-,024
Twitter				1,000	-,153	,064	,641**	-,009
Facebook					1,000	-,125	-,088	-,192
Wikipedia						1,000	,308*	,101
Google+							1,000	-,042
Mendeley								1,000

*La correlación es significativa en el nivel 0,01 (bilateral).
 **La correlación es significativa en el nivel 0,05 (bilateral).

sos habituales para ese tipo de medición y estudios como Altmetrics.com u otros:

"The differences uncovered by this study point to the need of careful sample selection when making generalizable claims about altmetrics. In particular, it draws attention to the importance of considering and further studying altmetrics in developing and emerging country contexts. The levels of coverage by the major altmetrics sources are generally lower than what has been previously reported. Most sources showed zero or almost zero coverage for all of the SciELO collection. Large differences were also found between the metrics that showed some coverage (Mendeley, Twitter, and Facebook)". Alperin (2015).

Por eso, la necesidad, por ahora, de ir a identificar los datos a cada fuente y plataforma, como ha hecho en este estudio.

4. CONCLUSIONES Y PERSPECTIVAS

En primer lugar, es necesario indicar que con los datos recopilados y otras opciones, es posible poder ampliar más el estudio considerando cada autor, cada fuente, o utilizando diferentes herramientas para su recopilación y comparación, pero este artículo, buscó una primera aproximación para la ALFIN, que diera cuenta de las posibilidades de estudios de este tipo, para tener alternativas complementarias (altmétricas), sin ser estudios totalmente abarcanes o finalizados, ya que en *altmetrics*, se está cada día probando y aproximando esas otras alternativas.

Por ende, su crecimiento en estudios y publicaciones, harán que tarde o temprano sean usuales, integrados con la bibliometría y la cibermetría, y por ello, *no sean solo alternativas*, como dice Alonso-Árevalo (2016):

Todos estos signos indican que altmetrics pueden no seguir siendo alternativa por mucho tiempo, lo que es especialmente atractivo para aquellas disciplinas y áreas que tradicionalmente han sido excluidas del análisis bibliométrico, ya que no están suficientemente representadas en bases de datos como Web of Science (WoS) o Scopus, como las ciencias sociales y las humanidades.

"La naturaleza del impacto de la investigación es cada vez más amplia y multifacética que nunca. La incorporación junto a la medición ortodoxa de medidas alternativas presenta una oportunidad para obtener una visión más holística del impacto de la investigación y su influencia, mediante la factorización de nuevos métodos de capturar e informar sobre la comunicación en línea y la actividad que rodea la investigación. La combinación de altmetrics y bibliometría presenta una gran oportunidad para obtener una imagen más precisa del alcance y la influencia de los documentos de investigación en la era digital".

Ya concretamente, para el caso de la ALFIN y los autores seleccionados, a partir de las plataformas estudiadas, concluimos de manera general, que es común que según las disciplinas de procedencia de los autores, de los investigadores, haya mayores preferencias en determinadas plataformas¹³.

De acuerdo con nuestro estudio, es ResearchGate la plataforma de carácter científico, entre las indagadas, la que se perfila como la más usada por los autores de este tema en dicho contexto geográfico, lo cual no es de extrañar considerando sus facilidades de uso, el ofrecer ORCID y la citación que indica, que aunque con cuestionamientos (Orduña-Malea y otros, 2016), es un complemento a la medición tradicional y que ha tenido un impacto mundial.

Por otro lado, LinkedIn se destaca como el espacio social con mayor preferencia por los autores estudiados, con un 62% de la muestra analizada, lo que concuerda con los hallazgos de Haustein (2014), en una muestra de investigadores bibliométricos (perfil afín a los estudiados en este trabajo), y desde lo indagado, es posible profundizar más adelante, en las que hemos denominado plataformas sociales para autores e interesados en ALFIN.

Otra de las principales conclusiones que deja este estudio, considerando los 35 autores que aparecen en la Figura 5, es que la presencia en las diversas plataformas sociales y científicas, se centra en un número limitado de autores (*top 10*), quienes resultan ser los mismos autores, que tuvieron más visibilidad e impacto, además de productividad, desde lo bibliométrico. Además, observamos cierta afinidad entre el indicador bibliométrico de número de citas con el indicador altmétrico de uso como la cantidad de lecturas en Mendeley, tal como lo reportaron otros estudios (Bar-Ilan, 2014).

Esto conlleva a identificar que hay un **impacto positivo en doble vía, de lo bibliométrico a las altmétricas y viceversa**, es decir, cuando un autor tiene una productividad importante y lo une con una buena gestión de su visibilidad en Internet, genera una adecuada identidad digital académica-científica, lo que produce que tanto las mediciones bibliométricas como altmétricas, se ven impactadas positivamente.

Por tanto, esto refuerza la "fuerte" afirmación que expresa Kuchner (2012) en su libro *"Marketing for Scientist"* (que también se había identificado en otro estudio relacionado con *altmetrics* e investigadores colombianos –Aguillo y otros, 2016–), donde afirma, parafraseándolo, que en el contexto del siglo XXI: **ser un buen científico es mitad ciencia y mitad divulgación**, y este estudio con autores iberoamericanos de ALFIN, también lo identifica.

Por tanto, queda como perspectiva, que hay una relación positiva de mutua reciprocidad entre la bibliometría y las altmétricas, por lo cual debemos seguirla profundizando desde diferentes casos y disciplinas; a la par, que es necesario seguir aprovechando esas distintas plataformas como medios para alcanzar una mayor difusión y divulgación, un mayor impacto científico y social, especialmente cuando están en sintonía con el acceso abierto y su potencial, a nivel citación¹⁴ y hacia medios de comunicación, políticos y gobernantes, profesionales de distintas áreas y comunidad en general¹⁵.

5. AGRADECIMIENTOS

Este trabajo se enmarca en el proceso investigativo: "Desarrollo de la red interinstitucional de conocimiento en CTIS -Red ICONOS- (2015-2017)" donde participa la Universidad de Antioquia-UdeA y otras instituciones, cuya acción 1 como Red es: "Identificación y análisis de alternativas de concepción y formas de diseño que tengan como misión fundamental observar, medir, analizar, evaluar y comunicar los estados y las dinámicas sociales asociadas a la Ciencia, la Tecnología, la Innovación, las Humanidades y las Artes". A su vez, este

artículo investigativo se enmarca en la Estrategia de Sostenibilidad Universidad de Antioquia-CODI 2016-2017 para el grupo de investigación de la EIB: "Información, Conocimiento y Sociedad". Investigador colaborador grupo INFOSFERA, Universidad de Granada-España.

5. ACKNOWLEDGEMENTS

This work fits into the research process: "Development of the inter-institutional knowledge network in CTIS - ICONOS network - (2015-2017)", where the Antioquia University UdeA and other institutions are participated and whose action 1 as a network is: "Identification and analysis of conception and design forms alternatives which have as essential task to notice, measure, analyze, evaluate and communicate the states and social dynamics associated to the Science, the Technology, the Innovation, the Humanities, and the Arts". At the same time, this research article fits into the support strategy of the Antioquia University - CODI 2016-2017, for the EIB research group: "Information, Knowledge and Society". INFOSFERA group Researcher, Universidad de Granada, Spain.

6. NOTAS

1 <http://www.infotecarios.com/altmetrics-para-todos/>

2 <http://altmetrics.org/manifesto/>

3 **Alfabetización científica (universitaria):** Es el **proceso de enseñanza-aprendizaje** que busca que los diferentes directivos y niveles de investigadores (o estudiantes, según el alcance) adquieran los conocimientos, habilidades y actitudes (competencias) que les permitan difundir y divulgar su quehacer científico y de innovación, utilizando de manera eficiente y eficaz, crítica y ética, distintos canales, formatos y/o herramientas (formales - informales, tradicionales - actuales, físicos - digitales/Web social); para lograr así una mayor **visibilidad científica** local, nacional o internacional, de sus procesos y desarrollos investigativos y de su producción (publicaciones, contenidos, otros), frente a las comunidades científicas, las instituciones de promoción y medición de la ciencia, las entidades gubernamentales y de políticas públicas, los medios de comunicación masivos o alternativos, las empresas con y sin ánimo de lucro, las organizaciones sociales y ONG's, y la sociedad en general.

Visibilidad científica (universitaria): Es el **nivel de reconocimiento** a nivel local, nacional o internacional, frente a las comunidades científicas, las instituciones de promoción y medición de la ciencia, las entidades gubernamentales y de políticas públicas, los medios de comunicación masivos o alternativos, las empresas con y sin ánimo de lucro, las organizaciones

sociales y ONG's, y la sociedad en general; que tiene una universidad, centro-grupo de investigación y/o sus diferentes niveles de investigadores (o estudiantes, según el alcance), por su **quehacer científico y de innovación** (extensión productiva o social), gracias a la difusión y divulgación, de sus procesos y desarrollos investigativos y de su producción (publicaciones, contenidos, otros) y medición, tanto por canales, formatos y/o herramientas: formales - informales, tradicionales - actuales, físicas - digitales/Web social.

4 A partir de este momento del texto, usaremos los términos en español: altmétrica(s), altmétrico(s).

5 <https://www.altmetric.com/products/explorer-for-institutions/>

6 <https://impactstory.org/>

7 <https://plu.mx/>

8 Ver referencias completas en: Uribe-Tirado (2013); Virkus (2003); entre otros.

9 En inglés, ya que permite encontrar trabajos en inglés, español o portugués, por el uso común de las *palabras clave*.

10 <https://ec3metrics.com/repositorio/#PAPERS-1>

11 <https://scholar.google.es/citations?user=jYqQRWgAAAAJ&hl>

12 <http://alperin.ca/publications/>

13 Por ejemplo, Ortega (2015a) reporta a ResearchGate como el más popular entre biólogos españoles frente a Academia.edu, como red social académica más usada por investigadores españoles de Humanidades y Ciencias Sociales.

14 <http://www.1science.com/oanumbr.html>

15 <http://blogs.lse.ac.uk/impactofsocialsciences/book/>

7. REFERENCIAS

- Aguillo, I.; Uribe-Tirado, A.; López, W. (2016). Comparative study of Colombian Researchers according to data from Google Scholar, ResearchGate and the National System for Measurement Science (Colciencias). *Proceedings of the 21st International Conference on Science and Technology Indicators*. València (España), 14-16 de Septiembre.
- Alonso-Arévalo, J. (2016). Altmetrics pueden no seguir siendo alternativa por mucho tiempo. *Revista ORL*, 7(3), 169-177. <http://dx.doi.org/10.14201/orl201673.14190> [10/10/2016]
- Alperin, J. P. (2013). Ask not what altmetrics can do for you, but what altmetrics can do for developing countries. *Bulletin of the American Society for Information Science and Technology*, 39(4), 18-21. <http://dx.doi.org/10.1002/bult.2013.1720390407> [10/10/2015]
- Alperin, J.P. (2014). Altmetrics could enable scholarship from developing countries to receive due recognition. *LSE Impact Blog*. <http://blogs.lse.ac.uk/impactofsocialsciences/2014/03/10/altmetrics-for-developing-regions/> [10/12/2015]
- Alperin, J.P. (2015). Geographic variation in social media metrics: an analysis of Latin American journal articles. *Aslib Journal of Information Management*, 67(3), 289-304. <https://stacks.stanford.edu/file/druid:sr068mj0031/AlperinGeographicVariationAltmetrics.pdf> [15/3/2016]
- Bakshy, E.; Hofman, J. M.; Mason, W. A.; Watts, D. J. (2011). Everyone's an influencer: quantifying influence on twitter. *Proceedings of the 4th ACM International Conference on Web Search and Data Mining, WSDM 2011*, pp. 65-74. Hong Kong: ACM. <http://dx.doi.org/10.1145/1935826.1935845> [10/10/2016]
- Bar-Ilan, J.; Haustein, S.; Peters, I.; Priem, J.; Shema, H.; Terliesner, J. (2012). Beyond citations: Scholars' visibility on the social Web. *Proceedings of the 17th International Conference on Science and Technology Indicators (STI)*, 5-8 September. Montreal, Quebec, Canada: Science-Metrix and OST, pp. 98-109.
- Bar-Ilan, J. (2014). Astrophysics publications on arXiv, Scopus and Mendeley: a case study. *Scientometrics*, 100(1), 217-225. <http://dx.doi.org/10.1007/s11192-013-1215-1> [26/09/2016]
- Bruce, C. (2011). Information Literacy Programs and Research: An International Review. *Australian Library Journal*, 60(4), 326-333. <http://doi.org/10.1080/00049670.2011.10722652> [30/11/2016]
- Borrego, Á. (2014). Altmétricas para la evaluación de la investigación y el análisis de necesidades de información. *El profesional de la información*, 23(4), 352-357. <http://dx.doi.org/10.3145/epi.2014.jul.02> [15/11/2015]
- Cha, M.; Haddadi, H.; Benevenuto, F.; Gummadi, P. K. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, pp. 10-17. Washington, D.C.: ICWSM.
- Costas, R.; Zahedi, Z.; Wouters, P. (2015). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, 66(10), 2003-2019. <http://dx.doi.org/10.1002/asi.23309> [07/10/2016]
- Cuevas-Cerveró, A.; Soriano, M. A. A. (2014). Producción científica internacional sobre alfabetización informacional para inclusión social de 2009 a 2013. *III Seminario Hispano Brasileño de Investigación en Información, Documentación y Sociedad*. <http://soac.unb.br/index.php/SHBPIDS/3/paper/view/3489> [15/12/2015]
- González-Valiente, C. L. (2015). Una aproximación al impacto de la investigación cubana sobre alfabetización informacional. *Revista Cubana de Información en Ciencias de la Salud*, 26, 53-70. <http://acimed.sld.cu/index.php/acimed/article/view/720> [10/12/2015]
- González-Valiente, C. L., Pacheco-Mendoza, J.; Arencibia, J. (2016). A review of altmetrics as an emerging discipline for research evaluation. *Learned Publishing*, 29(4), 229-238. <http://doi.org/10.1002/leap.1043> [26/09/2016]
- Harzing, A.W. (2007). *Publish or Perish*. <http://www.harzing.com/pop.htm> [10/10/2016]
- Haustein, S.; Peters, I.; Bar-Ilan, J.; Priem, J.; Shema, H.; Terliesner, J. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics*, 101(2), 1145-1163. <http://doi.org/10.1007/s11192-013-1221-3> [30/06/2016]
- Holmberg, K.; Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics*, 101(2), 1027-1042. <http://dx.doi.org/10.1007/s11192-014-1229-3> [26/09/2016]
- Konkiel, S.; Dalmou, M.; Scherer, D. (2015). *Altmetrics and analytics for digital special collections and institutional repositories*. <https://dx.doi.org/10.6084/m9.figshare.1392140.v1> [08/10/2016]
- Kuchner, M. J. (2012). *Marketing for Scientist*. Washington: IslandPress, 236 p.
- Liu, J.; Adie, E. (2014). Realising the potential of altmetrics within institutions. *Ariadne*, (72). <http://www.riadne.ac.uk/issue72/liu-adie> [09/10/2016]

- Loach, T. V.; Evans, T. S. (2015). Ranking Journals Using Altmetrics. *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference, 29 June to 3 July, 2015*. Istanbul, Turkey: Bogaziçi University Printhouse. <https://arxiv.org/abs/1507.00451>
- Melero, R. (2015). Altmetrics – a complement to conventional metrics. *Biochemia medica*, 25(2), 152-160. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4470104/>
- Martín-Martín, A.; Orduna-Malea, E.; Ayllón, J.M.; Delgado López-Cózar, E. (2016). The counting house: measuring those who count. Presence of Bibliometrics, Scientometrics, Informetrics, Webometrics and Altmetrics in the Google Scholar Citations, ResearcherID, ResearchGate, Mendeley & Twitter. *EC3 Working Papers*, (21). <https://arxiv.org/ftp/arxiv/papers/1602/1602.02412.pdf> [26/09/2016]
- Moyo, M.; Mavodza, J. (2016). A comparative study of information literacy provision at university libraries in South Africa and the United Arab Emirates: A literature review. *Library Review*, 65(1-2), 93-107. <http://dx.doi.org/10.1108/LR-06-2015-0069> [10/02/2016]
- Orduña-Malea, E. y Aguillo, I. (2014). *Cibermetría. Midiendo el espacio red*. Barcelona: EPI-OUC. 190 p.
- Orduña-Malea, E.; Martín-Martín, A.; López-Cózar, E. D. (2016). Researchgate como fuente de evaluación científica: desvelando sus aplicaciones bibliométricas. *El profesional de la información*, 25(2), 303-310.
- Ortega, J. L. (2015a). Disciplinary differences in the use of academic social networking sites. *Online Information Review*, 39(4), 520-536. <http://dx.doi.org/10.1108/OIR-03-2015-0093> [11/12/2015]
- Ortega, J. L. (2015b). Relationship between altmetric and bibliometric indicators across academic social sites: The case of CSIC's members. *Journal of Informetrics*, 9(1), 39-49. <http://dx.doi.org/10.1016/j.joi.2014.11.004> [26/09/2016]
- Piloiu, R.G. (2016). Rethinking the concept of "information literacy": a German perspective. *Journal of Information Literacy*, 10(2), 78-93. <https://doi.org/10.11645/10.2.2126> [10/12/2016]
- Pinto, M.; Escalona, M.I.; Pulgarín, A.; Uribe-Tirado, A. (2015). The scientific production of Ibero-American authors on information literacy (1985-2013). *Scientometrics*, 102(2), 1555-1576. <http://dx.doi.org/10.1007/s11192-014-1498-x> [10/10/2015]
- Sproles, C.; Detmering, R.; Johnson, A.M. (2013). Trends in the literature on library instruction and information literacy, 2001-2010. *Reference Services Review*, 41(3), 395-412. <http://doi.org/10.1108/RSR-03-2013-0014> [12/11/2016]
- Suri, V.R.; Chang, Y.K.; Majid, S.; Foo, S. (2014). Health information literacy of senior citizens -a review. *ECIL 2014: Information Literacy. Lifelong Learning and Digital Citizenship in the 21st Century*. Communications in Computer and Information Science, vol. 492. Springer, Cham.
- Tewell, E. (2015). A decade of critical information literacy: A review of the literature. *Communications in Information Literacy*, 9(1), 24-43. <http://www.comminfolit.org/index.php?journal=cil&page=article&op=view&path%5B%5D=v9i1p24> [30/01/2016]
- Torres-Salinas, D.; Milanés-Guisado, Y. (2014). Presencia en redes sociales y altmétricas de los principales autores de la revista El profesional de la información. *El Profesional de la Información*, 23(4), 367-372. <http://dx.doi.org/10.3145/epi.2014.jul.04> [26/09/2016]
- Uribe-Tirado, A. (2013). *Lecciones aprendidas en programas de Alfabetización Informacional en universidades de Iberoamérica*. Doctorado en Bibliotecología y Documentación Científica thesis, Universidad de Granada (España). <http://eprints.rclis.org/22416/> [30/11/2015]
- Uribe-Tirado, A. (2015). Percepciones, políticas y formación hacia la visibilidad académica y científica entre profesores-investigadores de la Universidad de Antioquia, 2015. *15a Conferencia Internacional BIREDIAL-ISTEC*, Barranquilla-Colombia: Universidad del Norte, 17 al 21 de noviembre. <http://eprints.rclis.org/28499/> [10/10/2015]
- Uribe-Tirado, A.; Pinto, M. (2014). 75 lecciones aprendidas en programas de alfabetización informacional en universidades iberoamericanas. *Revista Española de Documentación Científica*, 37 (3). <http://dx.doi.org/10.3989/redc.2014.3.1118> [30/11/2015]
- Uribe-Tirado, A.; Vallejo-Echavarría, J.C.; Betancur-Marín, D.A. (2016). Somos visibles y tenemos impacto. Análisis desde datos de acceso abierto, altmetrics y otros de la Revista Interamericana de Bibliotecología. *Revista Interamericana de Bibliotecología*, 39 (3), 243-275. <http://aprendeonline.udea.edu.co/revistas/index.php/RIB/article/view/323672/20782703> [28/9/2016]
- Virkus, S. (2003). Information literacy in Europe: A literature review. *Information Research*, 8(4). <http://www.informationr.net/ir/8-4/paper159.html> [01/11/2015]
- Wang, H.; Ya-hong, C.; Yuan, X.; Xu, Q.; Xia, J. (2011). A Study on Information Literacy of Higher Vocational College Students by the Literature Bibliometrics Analysis and Counterplan. *Journal of Taizhou Polytechnic College*, 6, 003. http://en.cnki.com.cn/Article_en/CJFDTOTAL-TZZY201106003.htm [05/11/2015]
- Wildgaard, L.; Schneider, J. W.; Larsen, B. (2014). A review of the characteristics of 108 author-level bibliometric indicators. *Scientometrics*, 101(1), 125-158. <http://dx.doi.org/10.1007/s11192-014-1423-3> [05/10/2016]
- Xia, D.; Mankad, S.; Michailidis, G. (2016). Measuring Influence of Users in Twitter Ecosystems Using a Counting Process Modeling Framework. *Technometrics*, 58 (3), 360-370. <http://dx.doi.org/10.1080/00401706.2016.1142906> [10/10/2016]
- Zhao, Y.; Wolfram, D. (2015). Assessing the popularity of the top-tier journals in the LIS field on Twitter. *Proceedings of the Association for Information Science and Technology*, 52(1), 1-4. <http://dx.doi.org/10.1002/pr2.2015.145052010092> [09/10/2016]

ESTUDIOS / RESEARCH STUDIES

Determinación de grupos de usuarios de bibliotecas digitales mediante el análisis de ficheros log

Juan-Antonio Martínez-Comeche

Universidad Complutense de Madrid, Facultad de Ciencias de la Documentación
Correo-e: juaamart@ucm.es | ORCID iD: <http://orcid.org/0000-0001-9074-8116>

Recibido: 25-09-2016; 2ª versión: 10-12-2016; Aceptado: 22-12-2016.

Cómo citar este artículo/Citation: Martínez-Comeche, J. A. (2017). Determinación de grupos de usuarios de bibliotecas digitales mediante el análisis de ficheros log. *Revista Española de Documentación Científica*, 40(3): e181. doi: <http://dx.doi.org/10.3989/redc.2017.3.1420>

Resumen: En este estudio se analiza el modo en que los usuarios realizan tareas de búsqueda y recuperación de información mediante consulta en la Biblioteca Digital Hispánica, distinguiendo grupos de usuarios en función de su distinto comportamiento informacional. Para ello se emplean los ficheros log recopilados por el servidor durante un año y se cotejan distintos algoritmos de agrupamiento. Se observa que el algoritmo k-means es un procedimiento de agrupamiento adecuado al análisis de extensos ficheros log de consultas en bibliotecas digitales. En el caso de la Biblioteca Digital Hispánica se distinguen tres grupos de usuarios cuyo comportamiento informacional distintivo se describe.

Palabras clave: Agrupamiento; algoritmo k-means; bibliotecas digitales; ficheros log; análisis de ficheros de transacciones; Biblioteca Digital Hispánica.

Clustering of users of digital libraries through log file analysis

Abstract: This study analyzes how users perform information retrieval tasks when introducing queries to the Hispanic Digital Library. Clusters of users are differentiated based on their distinct information behavior. The study used the log files collected by the server over a year and different possible clustering algorithms are compared. The k-means algorithm is found to be a suitable clustering method for the analysis of large log files from digital libraries. In the case of the Hispanic Digital Library the results show three clusters of users and the characteristic information behavior of each group is described.

Keywords: Clustering; k-means algorithm; digital libraries; log files; Transaction Log Analysis (TLA); Biblioteca Digital Hispánica.

Copyright: © 2017 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 presencia de las bibliotecas en la Web es cada vez más frecuente, pues permite el acceso remoto a un número creciente de colecciones digitalizadas y a los metadatos incorporados a sus documentos, favoreciendo en gran medida la difusión de sus fondos. Sin embargo, la puesta a disposición de recursos en línea plantea el reto de la adecuación entre el proceso técnico llevado a cabo internamente en las bibliotecas y las necesidades de los usuarios externos a la misma. El tratamiento documental comporta una organización y disposición de los fondos que suele diferir de la estructura del conocimiento que posee el usuario, pudiendo dificultar en ocasiones la tarea de localización y acceso a la documentación deseada.

Para superar este desajuste, además de poner a disposición del usuario diversas herramientas que le ayuden en su tarea, como los buscadores, es muy importante comprender el modo de actuar de los usuarios en su interacción con las bibliotecas digitales. De este modo podremos no solo eliminar o minimizar los obstáculos que encuentren al tratar de acceder a la documentación deseada, sino que podremos mejorar el funcionamiento y eficiencia de los sistemas de recuperación. Con interfaces mejor adaptadas al modo de proceder de los usuarios, los sistemas pueden recuperar la información buscada de manera más rápida y precisa.

El análisis que presentamos aquí se enmarca, pues, dentro del campo del comportamiento informacional, entendido como el estudio de cualquier experiencia de un individuo o grupo de individuos relacionada con la necesidad, búsqueda, gestión, difusión y uso de la información en diferentes contextos (González Teruel y Barrios Cerrejón, 2012; Fisher y otros, 2005).

Dentro de este marco general, en este estudio abordaremos específicamente el modo en que los usuarios realizan búsquedas en una biblioteca digital concreta, la Biblioteca Digital Hispánica (<http://bibliotecadigitalhispanica.bne.es>), portal desarrollado por la Biblioteca Nacional de España que actualmente facilita el acceso a miles de documentos digitalizados de una treintena de colecciones de contenido muy diverso: material cartográfico, dibujos, discos, libros, manuscritos o música, entre otros.

Emplearemos para ello los ficheros log de uno de los servidores utilizados por la Biblioteca Digital Hispánica (BDH en adelante). Existe una amplia literatura que avala la utilidad del análisis de ficheros log como procedimiento para comprender el comportamiento informacional de los usuarios.

En cuanto que registran las interacciones producidas entre las personas y los sistemas de información accesibles en Internet (Peters, 1993), el estudio de los ficheros log es un procedimiento útil para observar cómo actúan los usuarios en la Web: análisis de redes sociales (Rechavi y Rafaeli, 2014), marcadores sociales (Borrego y Fry, 2012), gobierno electrónico (Wang y otros, 2014), weblogs (Priya y Vadivel, 2012), organismos públicos y privados (Ortega Priego, 2005) o juegos online (Wang y otros, 2011a) pueden ser objeto de investigación por este medio.

Cuatro ámbitos pueden destacarse en la literatura sobre este tema: el sanitario, el académico y el empresarial, además de las unidades de información o sistemas de recuperación de información – que abordaremos después-. En cuanto al primero, son objeto de interés reciente el comportamiento de enfermos y pacientes (Yom-Tov y otros, 2014; Van Gemert-Pijnen y otros, 2014) o el empleo de directorios de temática sanitaria (Zhang y An, 2010).

En relación al segundo, se han realizado estudios sobre sedes electrónicas de universidades (Wang y otros, 2003), revistas electrónicas (Ortega Priego, 2004), repositorios digitales (Asunka y otros, 2011), uso de materiales docentes (Hershkovitz y otros, 2014), redes de investigación (Kahlon y otros, 2014) o la influencia de conflictos bélicos en el comportamiento del personal académico (Gul y otros, 2013). Dentro del marco empresarial, por su parte, interesa desde el comercio electrónico o las transacciones web (Ma, 2013) hasta las comunicaciones entre los participantes en un determinado negocio (Stuit y Wortmann, 2012).

En estos estudios se hace hincapié en los patrones de navegación (Guerbas y otros, 2013) y en el tráfico web (Dick y otros, 2014), en ocasiones con datos recopilados con herramientas como Google Analytics (Ozen y otros, 2014) o integrando su uso con datos del ámbito concreto analizado, recopilados mediante otros medios (Iyer y Raman, 2011).

En relación al empleo de ficheros log para la observación del comportamiento de los usuarios en unidades de información o en sistemas de recuperación de información, este tipo de datos ha sido importante fuente de información en un amplio abanico de estudios que van desde los buscadores genéricos (Jansen y Spink, 2006) hasta los catálogos de bibliotecas (Villén-Rueda y otros, 2007).

En el ámbito bibliotecario, de especial interés para este trabajo, destaca el análisis de los servicios ofrecidos en las unidades correspondientes (Leeder y Lonn, 2014), como el préstamo interbibliotecario (Munson y Otto, 2013), el servicio de referencia

(Rozaklis y MacDonald, 2011) o el servicio de atención al usuario (Arnason y Reimer, 2012).

La implantación de nuevos servicios aprovechando las posibilidades tecnológicas ocupa también un lugar destacado en la literatura. Entre estos estudios, pueden apuntarse el desarrollo de un chat (Berndt-Morris y Minnis, 2014), de un sistema de recomendación personalizada (Lai y Zeng, 2013), de un catálogo interactivo (Spiteri y Tarulli, 2012) o del préstamo bibliotecario mediante el teléfono móvil (Wang y otros, 2012).

Las colecciones y los recursos disponibles son igualmente objeto de interés, como las bases de datos accesibles (Mbabu y otros, 2013), la sección de libros digitales (Ahmad y otros, 2014) o la diferencia de empleo entre las colecciones digitales y las colecciones impresas (Kapoor, 2010).

Los objetivos finales de estos artículos inciden en la mejora de los servicios y de la experiencia del usuario (Tobias y Blair, 2015) o la mejora de la usabilidad y el diseño del sitio web (Shieh, 2012).

Referente al proceso de búsqueda y recuperación de información por parte de los usuarios, objeto específico de este estudio, sobresale el análisis de las consultas (Dempsey y Valenti, 2016) con fines muy diversos: mejorar la interfaz de búsqueda (Brett y otros, 2015), rediseñar sistemas de información (Waller, 2010) o servicios (Malliari y otros, 2010), mitigar las búsquedas fallidas (Moulaison y Stanley, 2013), descubrir la clase de información más consultada (Lambert, 2013) o las estrategias de búsqueda seguidas (Shiri, 2011).

También a raíz de los datos proporcionados por los ficheros log de las consultas realizadas por los usuarios se han tratado de inferir aspectos ajenos estrictamente al proceso de recuperación, como son los objetivos o las tareas en los que se enmarcan las consultas (Strohmaier y Kroll, 2012), ciclos económicos (Chen y Tsai, 2012) o incluso la posible orientación del voto de los ciudadanos (Borra y Weber, 2012).

Para alcanzar los objetivos señalados, estas investigaciones analizan muy diversos aspectos del uso: términos de búsqueda (Zhang y Zhao, 2013), clasificación de las consultas (Maabreh y otros, 2012), duración de las sesiones de búsqueda (Park y Lee, 2016), datos demográficos de los visitantes, tráfico, detección de problemas en las páginas web o transacciones comerciales primordialmente, aspectos que los programas de análisis web (Google Analytics, por ejemplo) han ido incorporando a los informes que generan.

Sin embargo, pocos han sido los esfuerzos por descubrir grupos de usuarios en base a su com-

portamiento con los sistemas en línea (Chapman, 1981; Hunt y otros, 2013), más allá de los factores geográficos o del tipo de dispositivo o programas de acceso que suelen incluir las herramientas de analítica web (Clifton, 2012).

Así pues, el objetivo principal de este estudio consiste en descubrir grupos de usuarios en función del comportamiento mostrado al afrontar tareas de búsqueda y recuperación de información en la Biblioteca Digital Hispánica mediante formulación de consultas, a partir de los ficheros log de tales actividades correspondientes al año 2013. La interfaz no ha sufrido cambios desde entonces y el sistema apenas ha incorporado algunas colecciones nuevas (rollos de pianola, por ejemplo), por lo que los resultados son totalmente aplicables en la actualidad. El hecho de considerar información relativa a un año en su integridad evita la incorporación de efectos estacionales indeseados en el análisis.

Los ficheros log recopilados incluyen, tras el proceso de limpieza, un total de 195.497 sesiones de consulta, cantidad equiparable a la empleada en otros estudios sobre comportamiento de búsqueda en bibliotecas digitales: en la New Zealand Digital Library se analizaron 251.878 consultas, considerando que una sesión puede abarcar varias consultas (Mahoui y Cunningham, 2000); en el caso de la Biblioteca de la Universidad de Granada se analizaron 200.000 transacciones (Villén-Rueda y otros, 2007). En otras investigaciones el número de sesiones es muy inferior, con 2509 sesiones de búsqueda (Brett y otros, 2015), aunque puede alcanzar también cifras muy superiores cuando se recopilan datos de varios años (Park y Lee, 2016). El número de sesiones consideradas en este estudio es, pues, cantidad suficiente para avalar la representatividad de los resultados alcanzados.

Como se ha indicado anteriormente, el empleo de ficheros log referentes a episodios de recuperación de información comporta tres posibles niveles de análisis (Jansen, 2006): términos, consultas y sesiones. Dado que el objetivo del estudio consiste en delimitar grupos de usuarios con un comportamiento semejante en las tareas de búsqueda de información mediante introducción de consultas, nos centraremos inicialmente en las sesiones que incluyan consultas.

El concepto de sesión y la delimitación de su duración puede ser diverso en función del contexto y del objetivo del análisis (Mahoui y Cunningham, 2000). En nuestro caso, conforme al objetivo expuesto, concebimos una sesión como el conjunto de actividades de consulta realizadas en el sistema por un cierto usuario con una duración máxima de

24 horas. Estos límites relativos al tiempo y al tipo de actividad vienen impuestos por el método de recopilación de la información seguido en el servidor de la BDH, puesto que la visualización de documentos es gestionada por un servidor distinto al empleado en las consultas y porque el sistema estructura los ficheros log por periodos de 24 horas.

Aunque no podamos analizar las acciones realizadas por los usuarios durante la visualización, debido a la estructuración seguida en la BDH para la recopilación de datos log, el tiempo ocupado en estas tareas se añade a la sesión de consulta correspondiente, pues esta acción no cierra la sesión en curso. Este hecho implica que la sesión de consulta considerada aquí aporta información añadida sobre el proceso de búsqueda y recuperación en su totalidad, más allá de la formulación de la consulta y la consulta de la lista de resultados.

Por su parte, partiremos de un concepto de consulta (Jansen y Pooch, 2001) que incluye las dos modalidades que permite el sistema Solr utilizado en la BDH: la consulta sencilla (mediante la introducción de términos en un área de texto) y la consulta avanzada (mediante la cumplimentación de un formulario con diversos campos relativos a los principales puntos de acceso: autor, título, edición, fechas o tipo de documento, entre otros). Estas dos modalidades de consulta fueron ya descritas entre los estudios iniciales de los catálogos en línea (Hancock-Beaulieu, 1989).

Una particularidad del sistema de recuperación de información de la BDH consiste en la posibilidad de filtrar los resultados obtenidos por cualquier modalidad de consulta, gracias a la inclusión de seis filtros en el sistema: acceso temático, tipo de material, colecciones destacadas, autor, lengua y año. Ello permite al usuario seleccionar características añadidas a la documentación aportada inicialmente por el sistema en respuesta a la consulta formulada.

El análisis del empleo de filtros por parte de los usuarios es de especial interés para los responsables de la BDH con el fin de complementar los datos que sobre ellos aporta el software de análisis web utilizado. En consecuencia, será objetivo prioritario de este estudio el análisis del empleo de los distintos filtros disponibles, incluyendo los diversos filtros como los factores principales para la constitución de los diversos grupos de usuarios.

En función de los objetivos expuestos anteriormente y de la definición de sesión adoptada, será precisamente la sesión el eje nuclear que nos permitirá delimitar comportamientos distintos en los usuarios. En consecuencia, nos detendremos inicialmente en aquellos aspectos cuantitativos relativos a las sesiones detectadas que configurarán, a

su vez, los factores que permitan caracterizar los distintos grupos de usuarios: duración, modalidad de consulta y filtros empleados.

Con el objeto de conocer al mismo tiempo las características generales del sistema analizado, mostraremos también el número de sesiones y su distribución por día y mes del año, aunque estos datos no se hayan empleado en el estudio de grupos de usuarios, objetivo primero de este estudio.

Una vez determinados los grupos de usuarios de la BDH, analizaremos los términos presentes en las consultas formuladas por los miembros de cada grupo, con el fin de mostrar algunas opciones de mejora del sistema de recuperación en función de su modo de proceder.

2. METODOLOGÍA

La metodología empleada para desvelar grupos de usuarios con características semejantes en función del comportamiento mostrado al realizar consultas, tal como el sistema los ha registrado en los ficheros log, es de carácter cuantitativo, entendiendo bajo esta denominación la combinación de procedimientos esencialmente matemáticos y estadísticos que permiten alcanzar respuestas, entre otros ámbitos, en la investigación social y del comportamiento (Vogt, 2011). El motivo que justifica esta elección es la conveniencia de aplicar algoritmos matemáticos comúnmente utilizados para afrontar el problema del agrupamiento o clustering de datos en relación a un conjunto de factores concreto y conforme a la noción de similitud o distancia adoptada (Kaufman y Rousseeuw, 2005).

Pueden distinguirse cuatro fases en el desarrollo de este tipo de estudios (Adèr y otros, 2008):

1. Recopilación de datos. Como se ha señalado previamente, los responsables de la BDH han facilitado los 365 ficheros de texto plano recogidos día a día por el servidor Solr que gestiona el sistema de recuperación de información de este organismo durante el año 2013. Dado que no es el único servidor involucrado durante una visita de un usuario al sitio web de la BDH, no es posible analizar el proceso íntegro seguido por los usuarios, sino tan solo las actividades de consulta en el buscador.
2. Procesamiento de datos. Con el fin de preservar el anonimato en la identidad de los usuarios, los responsables de la BDH sometieron los ficheros a un proceso de modificación de las direcciones IP en cada uno de los archivos que recoge la actividad producida en el servidor durante un día, al tiempo que se han reunido en una única línea todos los valores adoptados

por las variables durante la consulta. Así, cada sesión de consulta queda reflejada en un único estado final. Ello impide abordar las sesiones de consulta como un proceso compuesto de diversas acciones y, por otra parte, implica que puede aparecer una misma IP en dos días distintos sin que ello necesariamente suponga un mismo usuario, equipo o lugar de conexión con la BDH. De ahí que el análisis sea de carácter estático, sin entrar en la dinámica del proceso de consulta, y que nuestro concepto de sesión incorpore un límite temporal de 24 horas. De esta manera, cada dirección IP durante un cierto día permite identificar al mismo usuario que ha accedido al sistema.

3. Limpieza de datos. Una vez procesados los datos, cada uno de los ficheros con la actividad diaria se incorporó, mediante el software libre de tratamiento estadístico R (<https://www.r-project.org>), a tablas donde las filas recogen cada una de las sesiones de consulta realizadas al servidor por parte de los usuarios y las columnas las distintas variables incorporadas a las peticiones con sus valores respectivos. Se han desarrollado diversos scripts en el propio lenguaje R, con el objeto de estructurar cada sesión en las filas y de discernir en las columnas los valores que presenta cada variable (correspondientes básicamente a cada uno de los seis filtros, tipo de consulta –sencilla o avanzada–, duración de la sesión, dirección IP y día del año) en las respectivas consultas. Una vez incorporados los datos a tablas legibles en R, durante esta tercera fase se procedió a limpiar los datos, eliminando los casos de sesiones con datos incorrectos (duraciones negativas o nulas), las peticiones duplicadas (las consultas avanzadas recogidas inicialmente como "AdvancedSearch.do" y posteriormente como "Search.do") o las peticiones no consideradas en este estudio (como las solicitudes de visualización de datos, recogidas mediante la acción "CompleteSearch.do"). No se consideraron datos atípicos, dado el número significativo de sesiones con cualquier duración, incluyendo las muy cortas y las extremadamente prolongadas. El conjunto final empleado en este estudio incluye un total de 195.497 sesiones de consulta producidas durante el año.
4. Análisis de datos. Esta cuarta fase abarca tanto un análisis inicial del número de sesiones producidas por día y mes del año, con el objeto de tener una visión global del sistema, como el estudio de los ocho factores o parámetros cuantitativos disponibles para establecer posteriormente los distintos grupos de usuarios:

- a. Duración de la sesión.
- b. Tipología de consulta empleada en la sesión: sencilla o avanzada.
- c. Filtros empleados en la sesión, de los seis posibles: acceso temático, tipo de material, colecciones destacadas, autor, lengua y año.

Para ello se elaboraron diversos scripts en lenguaje R. Por último, se sometieron los valores de estos ocho parámetros para cada una de las 195.497 sesiones de consulta al algoritmo k-means de agrupamiento, analizándose los resultados y características de cada uno de los grupos obtenidos. Posteriormente se cotejó este algoritmo de agrupamiento con otros posibles procedimientos de clustering para observar sus diferencias cuando manejamos volúmenes grandes de datos, como es el caso de las consultas realizadas a una biblioteca digital de carácter nacional. Para el procesamiento de los distintos algoritmos de agrupamiento se emplearon los scripts disponibles en diversos paquetes del lenguaje R.

3. RESULTADOS

Los resultados del estudio se organizan en los siguientes apartados: en primer lugar, se presentan las características generales de las sesiones de consulta producidas durante el año, a fin de tener una visión general del sistema; a continuación, se muestran los valores obtenidos –con carácter global– en los distintos factores o parámetros empleados posteriormente en el cálculo de los grupos de usuarios (duración de las sesiones, tipología de consulta empleada en las mismas y porcentajes de utilización de los filtros por parte de los usuarios). Por último, se aborda la determinación de los grupos de usuarios y sus características más destacadas.

3.1. Número de sesiones por día y por mes

De las 195.497 sesiones de consulta efectuadas en la BDH durante el año, se han obtenido los siguientes valores medios:

- Media de sesiones por día: 535,61
- Media de sesiones por mes: 16291,42

Los usuarios mantienen un nivel de actividad parecido durante todo el año, con un cierto aumento durante los últimos cuatro meses del año. La máxima actividad se produce en el mes de septiembre (un 11,03%) y el mes de julio presenta la mínima actividad del año con un 6,47%.

Considerando las sesiones durante la semana, se produce un nivel de actividad semejante durante todos los días, con un ligero descenso durante el

fin de semana. Los lunes registran el nivel máximo de actividad (un 16,12%), mientras que los sábados presentan el nivel mínimo con un 10,83%.

3.2. Duración de las sesiones

Las sesiones presentan, en cuanto a su duración, valores en todo el intervalo comprendido entre un mínimo de 1 segundo y un valor máximo de 24 horas. La mitad de las sesiones presentan una duración menor o igual a 7,12 minutos, con una duración media de 102,20 minutos. Este valor medio tan elevado se justifica por el número significativo de sesiones con duraciones muy elevadas (1057, en concreto, duran entre 23 y 24 horas). En la Tabla I se resumen los tiempos de las sesiones distribuidos en intervalos, junto con su porcentaje.

3.3. Tipología de consulta empleada en las sesiones

Son posibles dos modalidades de consulta en las sesiones consideradas: consulta sencilla y consulta avanzada. Predominan en el corpus, con un 64,90%, las consultas sencillas en las que el usuario se limita a introducir una cadena de caracteres en el área de texto desarrollada al efecto, frente al 35,10% que ha optado por las consultas avanza-

das durante la sesión, en las que el usuario señala valores en relación a algunos de los puntos de acceso del catálogo (autor, título o tipo de documento, por ejemplo).

Como dato complementario, en la tabla II se recogen los campos más frecuentemente utilizados por los usuarios cuando cumplimentan el formulario de la búsqueda avanzada (sumando los valores de las tres variables "field" que las acogen).

3.4. Empleo de filtros en las consultas

El sistema de recuperación de información empleado en la BDH permite restringir los resultados iniciales mostrados ante una consulta del usuario (ya sea sencilla o avanzada), mostrando en el lateral izquierdo de la pantalla seis filtros desplegable: "acceso temático", "tipo de material", "colecciones desatadas", "autor", "lengua" y "año". De este modo, el usuario puede –sin necesidad de reformular la consulta– discernir estas características añadidas en los documentos hallados por el sistema en respuesta a su pregunta.

Los datos muestran que, de las 195.497 sesiones consideradas, en 116.216 sesiones (un 59,45% de los casos) los usuarios no utilizaron esta he-

Tabla I. Duración de las sesiones

Duración de la sesión	Frecuencia	Porcentaje
<=1 min.	45.961	23,51
>1 y <=2 min.	16.435	8,41
>2 y <=3 min.	11.020	5,64
>3 y <=4 min.	8.072	4,13
>4 y <=5 min.	6.333	3,24
>5 y <=6 min.	5.138	2,63
>6 y <=7 min.	4.342	2,22
>7 y <=8 min.	3.703	1,89
>8 y <=9 min.	3.357	1,72
>9 y <=10 min.	2.882	1,47
>10 y <=20 min.	18.093	9,26
>20 y <=30 min.	9.692	4,96
>30 y <=40 min.	6.575	3,36
>40 y <=50 min.	4.898	2,51
>50 y <=60 min.	3.782	1,93
>60 y <=70 min.	3.053	1,56
>70 y <=80 min.	2.468	1,26
>80 y <=90 min.	2.100	1,07
>90 y <=100 min.	1.760	0,90
>100 y <=1440 min.	35.833	18,33

Tabla II. Frecuencia de selección de campos en búsqueda avanzada

Campo seleccionado	Frecuencia	Porcentaje de empleo
Todos los campos	55.441	49,95
Autor	36.107	32,53
Materia	8.526	7,68
Lugar de publicación	4.214	3,80
Colección	3.971	3,58
Título	1.393	1,26
Otros	1.340	1,21

rramienta, mientras que en las 79281 sesiones restantes –lo que supone un 40,55% del total– los usuarios se sirvieron de esta herramienta para precisar los resultados deseados. En la Tabla III se puede consultar el número de sesiones en que se utilizó cada uno de los filtros, junto con el porcentaje del total de sesiones que representa.

De la tabla se concluye que el filtro más empleado por los usuarios fue el de “tipo de material” (en un 19,71% de las sesiones), seguido a bastante distancia por el filtro de “colecciones destacadas” (un 10,46%) y el filtro de “acceso temático” (un 7,93%). De igual forma, resalta la muy escasa utilización de los filtros de “año”, “autor” y “lengua”, que en ningún caso alcanzan un 2,5% de las sesiones.

3.5. Grupos de usuarios

La partición de las 195.497 sesiones en grupos se realizará mediante la aplicación del algoritmo k-means, debido a que no existen factores o características irrelevantes o que introduzcan ruido (Amorim y Mirkin, 2012), a que se trata de un algoritmo de poca complejidad, por lo que es adecuado para conjuntos de datos grandes (Celebi y otros, 2013), y a que es una técnica no supervisada, por lo que es adecuado para colecciones de datos no etiquetados o clasificados a priori como sucede en la BDH (Verma y otros, 2012).

Sus ventajas lo han convertido en uno de los algoritmos más frecuentemente empleados (Steinbach y otros, 2000). Entre los inconvenientes señalados en relación a este procedimiento destaca la decisión previa sobre el número de grupos que desean formarse (Amorim y Hennig, 2015; Huang, 1998). Para resolver este problema, se ha aplicado uno de los criterios más conocidos disponibles estadísticamente, denominado Sum of Squared Errors (SSE) (Celebi y otros, 2013; Schalkoff, 2001). En consecuencia, se establecen los siguientes apartados en esta sección, correspondientes a las sucesivas fases llevadas a cabo: Determinación del número de grupos, validación del agrupamiento y análisis de los grupos.

Determinación del número de grupos

SSE calcula la suma de los cuadrados de las distancias entre cada elemento de un grupo y su centroide, de manera que un número apropiado de grupos sería aquel en el que la disminución del SSE es más drástica (Peeples, 2011). La fórmula empleada es la siguiente, donde ‘n’ es el número de elementos de cada grupo:

$$SSE = \sum_{i=1}^n (x_i - \bar{x})^2$$

Tabla III. Frecuencia de empleo de los filtros en las consultas

Filtro	Frecuencia	Porcentaje de sesiones
Acceso temático	15.498	7,93
Tipo de material	38.536	19,71
Colecciones destacadas	20.449	10,46
Autor	3.531	1,81
Lengua	1.162	0,59
Año	4.334	2,22

En nuestro caso, como puede observarse en la Figura 1, el codo -esto es, el punto donde la reducción del valor de SSE es mayor- corresponde a N= 3 grupos.

En efecto, la diferencia entre N=2 y N=3 (en relación a la suma de cuadrados de las distancias intra-grupos) es mayor que la diferencia entre N= 3 y N= 4 (las cifras exactas pueden consultarse en la Tabla IV: la diferencia de SSE entre N=2 y N=3 es de 1,54 e+09, mientras que la diferencia de SSE entre N=3 y N=4 es de 4,74 e+08).

Una vez determinado que tres es el número de grupos inicialmente más adecuado para nuestro corpus de datos, sometemos los valores binarios (aplicación -1- o no -0-) de los 7 factores (6 filtros y la consulta avanzada) más el tiempo ocupado por cada una de las 195.497 sesiones al algoritmo k-means conforme al modelo clásico (MacQueen, 1967).

Validación del agrupamiento

Emplearemos dos criterios internos (intra-grupos) y dos criterios externos (inter-grupos) para comparar diversas posibilidades de particiones próximas entre sí (dos, tres, cuatro y cinco grupos mediante k-means):

- Criterios internos: índice Calinski-Harabasz e índice Silhouette
- Criterios externos: índice Pearson Hubert's Gamma e índice Single Link Average.

Estos criterios, a los que se añade el índice SSE empleado anteriormente, nos permitirán validar la elección adoptada. Los criterios internos proporcionan información sobre el grado de cohesión, relación o semejanza entre los elementos dentro de cada grupo. Los criterios externos, por su parte, indican en qué medida los grupos poseen límites cla-

Figura 1. Determinación del número de grupos mediante SSE

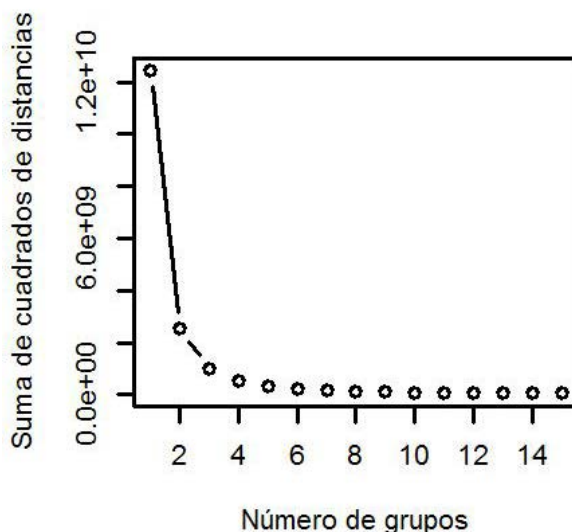


Tabla IV. Criterios intra-grupos e inter-grupos de validación del agrupamiento

K-medias	Criterios intra-grupos			Criterios inter-grupos	
	Calinski-Harabasz	Silhouette	SSE	Pearson Hubert Gamma	Single Link (Media)
2 grupos	757603	0,74	2,55 e+09	0,88	0,1
3 grupos	1107739	0,68	1,01 e+09	0,82	187,35
4 grupos	1443773	0,64	5,36 e+08	0,76	269,04
5 grupos	1714638	0,61	3,44 e+08	0,76	309,76

ros con respecto a los demás, cuantificando el nivel de desemejanza o dispersión entre los grupos. Un determinado agrupamiento será mejor cuanto mayor cohesión o semejanza exista entre los elementos de cada grupo y, a la vez, presente una mayor desemejanza entre los grupos. En consecuencia, cuanto mayor valor posea un cierto agrupamiento, en relación a cualquiera de los criterios internos o externos, mejor partición se considera en relación a dicho índice (Maulik y Bandyopadhyay, 2002).

El índice Calinski-Harabasz se basa en la relación entre la varianza entre los grupos y la varianza dentro de los grupos, de manera que un valor mayor del cociente indica una mejor partición (Calinski y Harabasz, 1974). El índice Silhouette compara las distancias de cada elemento con el resto de puntos de su grupo y con los demás puntos pertenecientes a los grupos restantes. Su valor oscila entre 1 (los elementos están apropiadamente situados en sus grupos) y -1 (los elementos deberían adscribirse a otros grupos) (Rousseeuw, 1987; Liu y otros, 2010).

En cuanto a los criterios externos, la versión de Pearson del coeficiente Gamma de Hubert halla la correlación entre las distancias entre todos los pares de elementos y el vector 0-1, donde 0 significa mismo grupo y 1 significa grupos o clusters diferentes (Halkidi y otros, 2001). El índice Single Linkage señala la distancia más corta entre dos elementos pertenecientes a dos grupos o clusters diferentes (Zhu y otros, 2015). En la Tabla IV se incluye la media de dichas distancias inter-grupos.

Como puede observarse en la tabla, las particiones en 2, 4 y 5 grupos presentan un mínimo en al menos un criterio: el agrupamiento en 2 conjuntos posee el valor mínimo en Calinski-Harabasz y en Single Link; el agrupamiento en 4 conjuntos o clusters posee el valor mínimo en el índice Pearson Hubert Gamma; la partición en 5 grupos posee el valor mínimo en el índice Silhouette y en el índice Pearson Hubert Gamma (junto al agrupamiento en 4 clusters).

De estos resultados se concluye que la partición en tres grupos no solo posee el máximo valor en el índice SSE, sino que no presenta tampoco ningún mínimo en los cuatro índices restantes. En consecuencia, se confirma la partición en tres grupos como una solución apropiada al corpus de datos.

Análisis de los grupos

El análisis cuantitativo de los tres grupos en relación a los ocho parámetros considerados en el algoritmo de agrupamiento se estructura en los siguientes puntos esenciales, pudiéndose consultar un resumen en la Tabla V.

Grupo 1

El primer grupo (que podemos denominar de usuarios experimentados) se caracteriza por los siguientes aspectos:

- Es el grupo más numeroso (172.402 sesiones) al acoger el 88,19% del total de 195.497 sesiones y, en consecuencia, agrupa la mayoría de los usuarios de la BDH.
- Este grupo engloba los usuarios cuyas sesiones de consulta ocupan el menor tiempo de conexión. En relación a este factor, los usuarios se conectan una media de 24 minutos con el buscador de la BDH, con un máximo de 237 minutos.
- Necesidad informativa centrada en la localización de documentos.
- Aproximadamente un 35% de este grupo utiliza la búsqueda avanzada, lo que supone un porcentaje igual a la media general.
- El filtro más frecuentemente utilizado por los usuarios de este grupo es el filtro "tipo de material" (un 20% de las sesiones), seguido del filtro de colecciones (11%) y del filtro temático (8%), en porcentajes semejantes a la media general.

Grupo 2

En el segundo grupo (que podemos denominar de usuarios poco experimentados) pueden destacarse los siguientes aspectos:

- Es el grupo menos numeroso (7.781 sesiones) con un 3,98% del total de 195.497 sesiones, por lo que agrupa a una minoría de los usuarios de la BDH.
- Este grupo incluye las sesiones más prolongadas; los usuarios se conectaron un mínimo de 799 minutos (13 horas) y un máximo de 1440 minutos (24 horas), con una media de 1147 minutos (19 horas).
- Necesidad informativa centrada en la visualización de documentos (para su análisis, copia y/o impresión).
- Aproximadamente en un 40% de estas sesiones se utilizó la búsqueda avanzada, en un porcentaje superior al resto de los grupos y a la media general.
- El filtro más frecuentemente utilizado por los usuarios de este segundo grupo es el filtro "tipo de material", aunque con un porcentaje inferior al grupo 1 (un 18% de las sesiones frente al 20% de las sesiones del grupo 1),

Tabla V. Factores y aspectos distintivos en los grupos de usuarios de la BDH

	Nº de sesiones	Duración media de sesiones (min)	Porcentaje de consultas avanzadas	Filtros más empleados (porcentaje de sesiones)	Aspectos distintivos
Valores globales	195497 (100%)	102,2	35,10	*Tipo de material (19,71) *Colecciones (10,46) *Temático (7,93)	
Grupo 1 (usuarios experimentados)	172402 (88,2%)	24	35,17	*Tipo de material (19,88) *Colecciones (10,77) *Temático (8,02)	*Grupo más numeroso (88,2%) *Menor tiempo de conexión * Necesidad informativa centrada en la localización de documentos *Valores semejantes a la media global en el empleo de filtros y consultas avanzadas
Grupo 2 (usuarios poco experimentados)	7781 (3,98%)	1147	39,35	*Tipo de material (18,35) *Autor (10,67) *Temático (10,53)	*Grupo menos numeroso (3,98%) *Mayor tiempo de conexión *Necesidad informativa centrada en la consulta o visualización de documentos *Mayor empleo de filtros de autor y temático
Grupo 3 (usuarios muy experimentados)	15314 (7,83%)	451,3	32,17	*Tipo de material (18,51) *Colecciones (8,15) *Temático (5,59)	*Menor empleo de consulta avanzada *Menor empleo de filtro de colecciones *Sesiones prolongadas *Necesidad informativa centrada en la consulta o visualización de documentos *Menor empleo de filtro temático

seguido de los filtros temático y de autor (un 11%) y del filtro de colecciones (8%). Hay, en consecuencia, un aspecto distintivo en este grupo consistente en un aumento en el empleo de los filtros de autor (que pasa del 2% general al 11%) y temático (que pasa del 8% general al 11%) y en una disminución del filtro de colecciones con respecto al primer grupo y a la media general (pasando del 10% al 8%). También se observa un aumento del empleo del filtro de lengua (0,9%, frente a un 0,6% en el total del corpus).

Grupo 3

El tercer grupo (que podemos denominar de usuarios muy experimentados) se caracteriza por los siguientes aspectos distintivos:

- Es un grupo poco numeroso (15314 sesiones), con un 7,83% del total de sesiones, lo que muestra la uniformidad general en el comportamiento de los usuarios de la BDH en tareas de consulta, agrupados mayoritariamente en el primer grupo.
- Los usuarios de este tercer grupo se caracterizan por sesiones prolongadas, entre 238

minutos (casi 4 horas) y 798 minutos (aproximadamente 13 horas), con una media de 451,3 minutos (7,5 horas).

- Necesidad informativa centrada en la visualización de documentos (para su análisis, copia y/o impresión).
- Los usuarios de este grupo son los que menos utilizaron la búsqueda avanzada durante sus consultas (un 32% de las sesiones, frente a un 40% en el grupo 2 y a un 35% en el grupo 1).
- El filtro más frecuentemente utilizado por los usuarios de este grupo es también el filtro "tipo de material", en un porcentaje (18,5%) parecido al resto de los grupos y a la media general (19,7%). A continuación figuran el filtro de colecciones (8%) y el filtro temático (un 5,6%). Este grupo se caracteriza, pues, por una disminución en el empleo de los filtros temático y de colecciones frente a la media general (8% y 10,5% respectivamente).

Una explicación para las sesiones prolongadas de los grupos segundo y tercero consiste en que las necesidades informativas de estos usuarios no se limitan a la búsqueda y localización de determinados textos o ejemplares de documentos conservados en la Biblioteca Nacional de España (como en el primer grupo), sino que se centra en la posterior visualización o consulta de las copias digitalizadas de tales documentos, bien para su análisis o estudio, bien para su copia y/o impresión, o para ambas tareas. Como se señaló en la introducción, la visualización o acceso a los documentos es gestionada por un servidor distinto al empleado en las consultas, pero siempre que el usuario regresa a la página de resultados tras la visualización (para ver otros documentos del listado o para acceder al contenido de otro u otros documentos relevantes recuperados), dicha acción quedará reflejada en el servidor empleado en la consulta como parte de su sesión, englobando así el tiempo empleado en la visualización. Por otra parte, Park y Lee (Park y Lee, 2013) -en el análisis de las consultas al sistema de recuperación del Instituto de Ciencia y Tecnología de Corea- han detectado sesiones de 10 horas, concluyendo que las sesiones en un sistema de recuperación de información son más prolongadas que en un buscador genérico.

En la Tabla VI se resumen algunas características de los tres grupos de usuarios de la BDH relativas a consultas, términos, colecciones y filtros. En el caso de las colecciones más consultadas, debe tenerse en cuenta que la solicitud de

la colección "Cartas náuticas" queda reflejada en los ficheros log dentro del área de texto, motivo por el cual figura como una consulta más. Tampoco las solicitudes de las colecciones "Fotografía" y "Grabados de Rembrandt" se registran como el resto, motivo por el que no se incluyen en los datos mostrados.

Por otra parte, el análisis mediante clustering jerárquico de los filtros ha permitido obtener el nivel de correlación entre los filtros en cada uno de los tres grupos de usuarios, indicador de la utilización simultánea de los mismos. De dicho análisis se desprende también que la búsqueda avanzada no está relacionada ni con la utilización de ciertos filtros ni con la consulta de determinadas colecciones, por lo que su empleo puede considerarse independiente de ellas.

4. DISCUSIÓN

Se ha destacado en diversas ocasiones que los usuarios, cuando se enfrentan a un sistema de recuperación de información, suelen realizar consultas sencillas con un número de términos pequeño (Markey, 2007). En el caso de los motores de búsqueda genéricos, Spink y Jansen (Spink y Jansen, 2004) confirman que el empleo de características avanzadas en los buscadores genéricos es muy baja (en torno al 5%), resaltando el caso de Altavista, que en 2002 alcanzó un 27,6% de utilización de operadores de búsqueda en las consultas efectuadas.

En el caso del OPAC de una biblioteca universitaria, por el contrario, se ha señalado una utilización similar de ambas modalidades de consulta, fenómeno explicado porque los usuarios, cuanto mayor conocimiento de la información que están buscando y más familiaridad con el OPAC poseen, optan en mayor medida por las búsquedas sencillas (Villén-Rueda y otros, 2007).

El hecho de que las colecciones de la BDH sean especializadas y posean un alto valor histórico restringe en buena medida los usuarios potenciales, que en mayor proporción saben con precisión qué documentación buscan y qué biblioteca la conserva. En consecuencia, el carácter más especializado de la documentación conservada y la mayor proporción de usuarios con un conocimiento previo de sus necesidades informativas y del sistema de recuperación entre los que acceden a esos fondos permite justificar la preponderancia de las consultas sencillas (casi un 65%) sobre las consultas avanzadas (un 35%), tanto con carácter general como específicamente en el grupo 1 que mayoritariamente las acoge.

Tabla VI. Características de los grupos relativas a consultas, términos, colecciones y filtros

	Consultas más frecuentes	Longitud media de consultas (nº de términos)	Términos más frecuentes (eliminadas las palabras vacías)	Co-términos más frecuentes (eliminadas las palabras vacías)	Colecciones más frecuentes	Filtros de mayor utilización simultánea
Valores globales	"Avicena" "guitarra" "partituras" "portulano" "Filipinas"	1,8	"Juan" "historia" "España" "Antonio" "san"	"carta"+"náutica" "Puerto"+"Rico" "Johann"+"Strauss" "Josef"+"Strauss" "Carlos"+"Gardel"	*Obras maestras *Carteles *Grabados de Durero *Ephemera *Viajes	* Tipo de material * Lengua
Grupo 1	"Avicena" "guitarra" "partituras" "portulano" "carta náutica"	1,8	"Juan" "historia" "España" "Francisco" "Antonio"	"carta"+"náutica" "Puerto"+"Rico" "Johann"+"Strauss" "Josef"+"Strauss" "Carlos"+"Gardel"	*Obras maestras *Carteles *Grabados de Durero *Ephemera *Libros de caballería	* Tipo de material * Lengua
Grupo 2	"calzada romana" "Puente la Reina" "Nasarre" "guitarra" "Puerto Rico"	1'9	"calzada" "romana" "reina" "España" "puente"	"calzada"+"romana" "Puente"+"Reina" "Puerto"+"Rico" "Manuel"+"Navarro" "ejército"+"español"	*Dibujos de los niños de la guerra *Grabado contemporáneo *Hispanoamérica *Obras maestras *Carteles	* Autor * Temático
Grupo 3	"Thesaurus Sacrarum Historiarum" "el toreo" "partituras" "guitarra" "óperas"	1'8	"Juan" "Antonio" "historia" "san" "España"	[Combinaciones de dos términos entre "Thesaurus", "Sacrarum" e "Historiarum"] "Puerto"+"Rico" "Edvard"+"Grieg" "Guerra"+"Independencia" "Don"+"Quijote"	*Obras maestras *Carteles *Grabados de Durero *Ephemera *Dibujos de los niños de la guerra	* Temático * Autor

Es por ello que hemos empleado para designar al grupo 1 de usuarios, el más numeroso y con iguales valores a los generales, la denominación de usuarios experimentados. Los grupos restantes confirman esta argumentación: el segundo grupo de usuarios, menos experimentado, debería presentar, en consecuencia, un porcentaje mayor de consultas avanzadas, como así es (con un 39,35%); al tercer grupo de usuarios, muy experimentados, le correspondería un menor empleo de consultas avanzadas, como en efecto sucede (con un 32,17%).

El empleo de filtros por parte de los distintos grupos confirma el razonamiento basado en el grado de conocimiento de los documentos buscados y la experiencia previa en el empleo del sistema de recuperación de la BDH, aunque con menor intensidad. Los valores globales dan un 40,55% de utilización de algún filtro, frente a un 59,45% de sesiones en que no se utilizó ningún filtro.

Tanto el primer grupo (la mayoría de los usuarios, con experiencia y conocimientos previos) como el segundo grupo (una minoría de usuarios,

menos experimentados) presentan un grado de utilización parejo a estos valores generales: un 41,05% de las sesiones del grupo 1 y un 40,03% de las sesiones del grupo 2 utilizaron alguno de los filtros, no percibiéndose una diferencia significativa. El segundo grupo, sin embargo, se distingue por un mayor empleo de los filtros temático y de autor que el grupo 1 y que los usuarios de la BDH en general, lo que reflejaría unas necesidades informativas menos delimitadas por parte de este grupo de usuarios.

Por su parte, el tercer grupo (de usuarios más experimentados que la media) presenta un grado significativamente menor de empleo de algún filtro (un 35,24%) y específicamente de los filtros temático y de colecciones, acorde con una concreción mayor del documento deseado y una menor necesidad de herramientas que permitan restringir la búsqueda.

Estos datos confirman los resultados obtenidos por Ferl y Millsap, quienes destacan que una proporción mucho mayor de estudiantes (usuarios menos experimentados) realizaron búsquedas

das por temas que el profesorado y el personal de la biblioteca de la Universidad de California (usuarios más experimentados) (Fertl y Millsap, 1996). Larson había identificado anteriormente los problemas del acceso por encabezamientos temáticos en los OPACs, entre los que enunciaba precisamente la falta de conocimiento por parte de los usuarios de dicha herramienta de acceso (Library of Congress Subject Headings o LCSH) (Larson, 1991).

En relación a la duración de las sesiones, muy diversos factores influyen en una estancia más o menos prolongada por parte de los usuarios, pudiéndose destacar el contenido del sitio web, la tarea que afronta el usuario (en la que se enmarca la búsqueda de información) y las características del usuario (Lalmas y otros, 2014).

El tiempo transcurrido en un cierto sitio web depende en buena medida de la clase de contenidos que oferte. Se ha constatado que los sitios web a los que se acude a consultar noticias presentan sesiones mucho más cortas que los sitios que involucran la realización de procesos o actividades durante la visita, como el comercio electrónico o las redes sociales (Benevenuto y otros, 2009). Por este motivo las sesiones detectadas en los buscadores generalistas son mucho más cortas que las sesiones en sistemas de recuperación de bibliotecas o repositorios de documentos digitales (Park y Lee, 2013).

La tarea que se desarrolle en el momento de la visita repercute igualmente en el tiempo de permanencia en el sitio web. Un usuario con una necesidad informativa específica (como la comprobación de recepción de un correo) ocupa menos tiempo que si el usuario navega por mera curiosidad en un sitio web sobre su afición favorita (Wang y otros, 2011b). De igual forma, si el sitio web enlaza con otros sitios o páginas relacionadas, añadiendo tareas alternativas o complementarias, las sesiones pueden prolongarse en buena medida (Lalmas y otros, 2014). Es el caso de la BDH, pues las acciones de visualización, lectura, copia o impresión de uno o varios de los resultados de la búsqueda se añaden al tiempo de la sesión de consulta. Cabe esperar, pues, que los usuarios cuya necesidad informativa se limite a la localización de cierta documentación tengan sesiones de duración más cortas que los usuarios que deban o deseen consultar la documentación previamente recuperada, bien por enmarcarse en tareas académicas o profesionales, ya sea por simple curiosidad.

Por este motivo caracterizamos los grupos 2 y 3 (con sesiones más prolongadas) con necesidades informativas más complejas (incluyendo la visua-

lización de documentos), mientras que las sesiones más cortas del grupo 1 se explican bien si las necesidades informativas de sus usuarios son más sencillas o específicas, limitándose a la búsqueda y localización de documentos.

Otro gran grupo de factores que afectan a la duración de las sesiones tienen que ver con las aptitudes y actitudes del usuario que acude a un sitio web. Si un usuario se topa, por ejemplo, con dificultades en la interacción con el sistema, por falta de comprensión de la estructura y organización de la información que presenta el sitio web, la sesión será más prolongada que si el usuario conoce bien cómo proceder para resolver su necesidad. De igual forma, un usuario distraído o que está involucrado simultáneamente en diversas tareas tendrá sesiones más largas que un usuario concentrado o interesado en la información disponible en el sitio web (Huang y White, 2010).

Así pues, este grupo de factores explicaría bien las sesiones más prolongadas en el grupo 2 (caracterizándolo como poco experimentados) frente a las sesiones más cortas del grupo 3 (muy experimentados, aunque involucrados en tareas más complejas).

Como puede observarse en la Tabla VI, la longitud media de las consultas efectuadas al sistema (esto es, el número de términos que incluyen) está cerca de los dos términos por consulta, valor constante en todos los grupos de usuarios de la BDH. Estos valores confirman los datos hallados en estudios anteriores centrados en bibliotecas digitales, donde se concluye que el 80% de las consultas incluyen hasta 3 términos como máximo (Agosti y otros, 2012).

Los términos más frecuentes en dichas consultas (eliminadas las palabras vacías) muestran el carácter de los fondos predominantes de la BDH, directamente relacionados con los objetivos prioritarios de la Biblioteca Nacional, de manera que los usuarios solicitan primordialmente documentación de carácter histórico relativo a España (los términos 'historia' y 'España' destacan entre los más habituales en todos los grupos).

Las parejas de términos que en mayor número aparecen simultáneamente en las consultas (eliminadas las palabras vacías) permiten precisar mejor las necesidades informativas de los usuarios. Considerando los co-términos más frecuentes en los tres grupos, predominan claramente las consultas sobre personas, lugares geográficos y obras.

Esta tipología contrasta con la reducida utilización de la búsqueda avanzada (un tercio de consultas aproximadamente), modalidad que permi-

tiría al usuario precisar mejor su necesidad informativa. En parte este hecho puede estar motivado por una falta de adecuación entre los campos que se ofrecen al usuario en la consulta avanzada y la naturaleza de la necesidad informativa. Por ejemplo, si el usuario busca información sobre una determinada persona, no necesariamente está interesado en las obras de las que sea autor. Este hecho, unido a que el usuario raramente modifica la configuración inicial de las herramientas puestas a su disposición (Jones y otros, 2000), hace que la opción por defecto ('Todos los campos' en la BDH), sea la más empleada –incluyendo los casos en que el usuario no encuentra una opción adecuada a su necesidad concreta–.

Una recomendación, pues, que permitiría mejorar la experiencia del usuario de la BDH (Tobias y Blair, 2015) consistiría en modificar la interfaz de búsqueda sencilla de manera que, conforme el usuario vaya introduciendo los términos de la consulta, el sistema despliegue sugerencias de nombres basadas en el catálogo de autoridades de persona, geográficos y de título que posee la Biblioteca Nacional, y que está actualmente a disposición de los usuarios pero en una página web independiente (<http://catalogo.bne.es/uhtbin/authoritybrowse.cgi>). De esta manera no solo facilitamos la tarea de búsqueda al usuario, sino que mitigaremos las búsquedas fallidas (Moulaison y Stanley, 2013) al evitar nombres mal escritos o ambigüedades (por ejemplo, 'Manuel Navarro', consulta muy frecuente en el grupo 2, puede referirse a un político argentino, un poeta cubano o un compositor español).

De igual forma, aunque las colecciones más consultadas son distintas en cada uno de los grupos de usuarios, también es cierto que existen tres colecciones que son solicitadas por todos los grupos: Obras maestras, Carteles y las de Grabados (especialmente los de Durero y los contemporáneos). Una segunda recomendación a este respecto sería añadir, en el menú desplegable situado justo a la derecha del área de texto en la búsqueda sencilla, al menos las tres colecciones más consultadas en todos los grupos de usuarios: Obras maestras, Carteles y las de Grabados. Con ello se pone el énfasis en los aspectos más empleados y, al tiempo, optimizamos las estrategias de búsqueda a disposición de los usuarios al añadir una posibilidad más en la modalidad de consulta sencilla.

En relación al procedimiento seguido para hallar los grupos de usuarios, debe destacarse la dimensión de los datos como un factor determinante a favor del algoritmo k-means, en detrimento de otras posibilidades matemáticas como el clustering jerárquico, cuya complejidad hace que sea un

procedimiento poco adecuado para grandes volúmenes de datos, como es nuestro caso (Steinbach y otros, 2000).

Otro gran grupo de procedimientos de agrupamiento se basan en modelos de distribución, entre los que destacan las distribuciones gaussianas, que se basan en la probabilidad de adscribir cada elemento a una determinada distribución gaussiana. A su vez, dentro de este tipo de distribución, sobresalen los modelos mixtura de gaussianas, que emplean algoritmos esperanza-maximización (Hastie y otros, 2009) con el objetivo de hallar iterativamente los parámetros de las distribuciones gaussianas que mejor se adapten a los datos (Bouveyron y otros, 2007).

Sometiendo el corpus de datos a este algoritmo entre 1 y 11 grupos, se aconseja formar 10 grupos con las sesiones. Este número tan elevado de grupos puede deberse al problema de sobreajuste que afecta a esta clase de procedimientos (Tu, 2005).

Los modelos de agrupamiento basados en densidad descubren grupos en áreas donde se localiza una mayor concentración o densidad de elementos. Los elementos dispersos que no pertenezcan a zonas de mayor aglomeración de observaciones se denominan ruido (Ester y otros, 1996).

Realizadas las pruebas correspondientes, pueden obtenerse entre 8 y 14 grupos, imponiendo un valor de 3 para el parámetro 'eps' y variando el parámetro minPts (con valor minPts=100 se obtienen 8 grupos y con valor minPts=50 se obtienen 14 grupos). Conforme disminuye el número de grupos, más elementos son considerados como ruido o puntos límite (para 14 grupos obtenemos 187 elementos no adscritos a ninguno de dichos grupos, hasta un máximo de 634 sesiones consideradas ruido o puntos límite para 8 grupos). A su vez, si disminuimos el valor del parámetro 'eps' (entre 0,40 y 1,5), obtenemos un número muy abultado de grupos (con eps= 0,43 se obtienen 69 grupos, mientras que con eps= 1,5 se obtienen 56 grupos). En consecuencia, este tipo de modelos no ofrece una solución aplicable a nuestro caso, debido al elevado número de grupos propuesto.

Por último, se han realizado pruebas con un modelo de clustering semejante en sus principios al algoritmo k-means denominado k-medoids. La diferencia entre ambos algoritmos estriba en que, mientras los grupos en k-means están representados por un punto central que no tiene por qué ser necesariamente un elemento del grupo, en k-medoids los grupos están representados por un elemento del grupo cuya disimilaridad media con todos los objetos o elementos del grupo es mínima (Velmurugan y Santhanam, 2010; Kaufman y Rousseeuw, 2005).

Los resultados obtenidos en las pruebas proporcionan 3 grupos con un número semejante de elementos al algoritmo k-means (un grupo 1 con 168.427 elementos; el grupo 2 posee 9.789 elementos; y el grupo 3 reúne 17.281 elementos), aunque con peores valores en los índices Calinski_Harabasz (1013477 frente a 1107739 del algoritmo k-means) y Silhouette (0,635 frente a 0,676 del algoritmo k-means).

Por último, comentar que uno de los inconvenientes que suele señalarse al analizar el algoritmo k-means es la tendencia a formar grupos de tamaño similar (Kaufman y Rousseeuw, 2005). Como hemos podido observar, en nuestro caso los grupos son muy dispares entre sí en cuanto al número de elementos, por lo que esta tendencia no ha tenido una influencia negativa en el caso analizado.

5. CONCLUSIONES

El algoritmo k-means es un procedimiento de agrupamiento muy conocido que se adecua bien al análisis de extensos ficheros log de sesiones de consulta, debido a su poca complejidad, su eficiencia en tiempos de ejecución y a su aplicabilidad a conjuntos de datos de grandes dimensiones, además de los altos valores que obtiene en los diversos índices intra-grupos e inter-grupos empleados para su validación.

El análisis cuantitativo de los ficheros log de la BDH correspondientes a un año ha permitido determinar tres grupos de usuarios en función de su comportamiento en tareas de búsqueda de información mediante consulta. Distintos valores de la duración de las sesiones, de la utilización de la consulta avanzada y del empleo de filtros permite caracterizar cada uno de esos grupos, cuyo resumen se puede consultar en las Tablas V y VI. A raíz de estos datos se proponen recomendaciones para mejorar la experiencia de búsqueda del usuario.

La relación inversa entre la utilización de herramientas para la formulación de la consulta y el nivel de conocimiento de la información que busca el usuario o la experiencia previa en el funcionamiento del buscador, relación señalada ya en estudios anteriores, permite calificar los grupos con las denominaciones 'experimentados', 'poco experimentados' y 'muy experimentados' en función de los niveles de utilización de la consulta avanzada y de la herramienta de filtración de resultados.

El grupo 2 (usuarios poco experimentados) emplea en mayor medida la ayuda que brinda la con-

sulta avanzada, distinguiéndose al tiempo por una mayor utilización de los filtros temático y de autor. Ello se explica por unas necesidades informativas menos perfiladas y por un menor conocimiento de la BDH en general y de su sistema de recuperación en particular. El grupo 3 (usuarios muy experimentados), por el contrario, emplea menos la consulta avanzada y los filtros temático y de colecciones, debido a una concreción mayor de la documentación buscada y de un mejor conocimiento de los fondos de la BDH y de su buscador.

De cara al futuro, sería de gran utilidad poder completar estas conclusiones con información sobre el proceso de consulta llevado a cabo por los usuarios y las diversas acciones que la componen, de manera que podamos analizar en profundidad el modo de proceder de los usuarios durante las tareas de búsqueda y recuperación de información. A este respecto sería de interés simplificar la estructura de servidores que recopilan los datos log y conservarlos para mejorar el servicio mediante su análisis periódico.

De igual forma, sería conveniente completar la información meramente cuantitativa sobre los grupos hallados con un análisis posterior, de carácter cualitativo, que permita sacar a la luz características de los grupos que el tratamiento cuantitativo no desvela. Al mismo tiempo, este segundo análisis permitiría corroborar los resultados numéricos obtenidos previamente.

6. AGRADECIMIENTOS

Quisiera agradecer la colaboración de los responsables de la Biblioteca Digital Hispánica de la Biblioteca Nacional de España, sin cuya ayuda este análisis no habría sido posible. En especial quiero mostrar mi agradecimiento a Isabel Bordes Cabrera, jefa de Área de Biblioteca Digital, por su interés y permanente ayuda al facilitarme todos los datos utilizados en este estudio.

ACKNOWLEDGEMENTS

This work was carried out with the collaboration of the manager staff of the Hispanic Digital Library of the National Library of Spain, without whose help this analysis would not have been possible. In particular I want to show my gratitude to Isabel Bordes Cabrera, head of Digital Library area, for her interest and permanent help to provide me with all data used in this study.

7. REFERENCIAS

- Adèr, H. J.; Mellenberg, G. J.; Hand, D. J. (2008). *Advising on research methods: a consultant's companion*. Johannes van Kessel Publishing; Huizen, the Netherlands.
- Agosti, M.; Crivellari, F.; Di Nunzio, G. M. (2012). Web log analysis: a review of a decade of studies about information acquisition, inspection and interpretation of user interaction. *Data Mining and Knowledge Discovery*, vol. 24(3), 663-696. <https://doi.org/10.1007/s10618-011-0228-8>
- Ahmad, P.; Brogan, M.; Johnstone, M. N. (2014). The e-book power user in academic and research libraries: Deep log analysis and user customization. *Australian Academic & Research Libraries*, vol. 45(1), 35-47. <https://doi.org/10.1080/00048623.2014.885374>
- Amorim, R.; Hennig, C. (2015). Recovering the number of clusters in data sets with noise features using feature rescaling factors. *Information Sciences*, vol. 324, 126-145. <https://doi.org/10.1016/j.ins.2015.06.039>
- Amorim, R.; Mirkin, B. (2012). Minkowski metric, feature weighting and anomalous cluster initializing in k-means clustering. *Pattern Recognition*, vol. 45, 1061-1075. <https://doi.org/10.1016/j.patcog.2011.08.012>
- Arnason, H.; Reimer, L. (2012). Analyzing public library service interactions to improve public library customer service and technology systems. *Evidence Based Library and Information Practice*, vol. 7(1), 22-40. <https://doi.org/10.18438/B8NP6T>
- Asunka, S.; Chae, H. S.; Natriello, G. (2011). Towards an understanding of the use of an institutional repository with integrated social networking tools: A case study of PocketKnowledge. *Library & Information Science Research*, vol. 33(1): 80-88. <https://doi.org/10.1016/j.lisr.2010.04.006>
- Benevenuto, F.; Rodrigues, T.; Cha, M.; Almeida, V. (2009). Characterizing user behavior in online social networks. *Proceedings of the 9th ACM SIG-COMM Conference on Internet Measurement Conference*, pp. 49-62. ACM; New York. <https://doi.org/10.1145/1644893.1644900>
- Berndt-Morris, E.; Minnis, S. M. (2014). The chat is coming from inside the house: An analysis of perceived chat behavior and reality. *Journal of Library & Information Services in Distance Learning*, vol. 8(3-4), 168-180. <https://doi.org/10.1080/1533290X.2014.945833>
- Borra, E.; Weber, I. (2012). Political insights: Exploring partisanship in web search queries. *First Monday*, vol. 17(7). <https://doi.org/10.5210/fm.v17i7.4070>
- Borrego, A.; Fry, J. (2012). Measuring researchers' use of scholarly information through social bookmarking data: A case study of BibSonomy. *Journal of Information Science*, vol. 38(3), 297-308. <https://doi.org/10.1177/0165551512438353>
- Bouveyron, C.; Girard, S.; Schmid, C. (2007). High-dimensional data clustering. *Computational Statistics and Data Analysis*, vol. 52(1), 502-519. <https://doi.org/10.1016/j.csda.2007.02.009>
- Brett, K.; German, E.; Young, F. (2015). Tabs and tabulations: Results of a transaction log analysis of a tabbed-search interface. *Journal of Web Librarianship*, vol. 9(1), 22-41. <https://doi.org/10.1080/19322909.2015.1004502>
- Calinski, T.; Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, vol. 3(1), 1-27. <https://doi.org/10.1080/03610927408827101>
- Celebi, M. E.; Kingravi, H. A.; Vela, P. A. (2013). A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Systems with Applications*, vol. 40(1), 200-210. <https://doi.org/10.1016/j.eswa.2012.07.021>
- Chapman, J. L. (1981). A state-transition analysis of online information seeking behavior. *Journal of the American Society for Information Science*, vol. 32(5), 325-333. <https://doi.org/10.1002/asi.4630320503>
- Chen, C. C.; Tsai, Y. (2012). A novel business cycle surveillance system using the query logs of search engines. *Knowledge-Based Systems*, vol. 30, 104-114. <https://doi.org/10.1016/j.knsys.2011.12.012>
- Clifton, B. (2012). *Advanced web metrics with Google Analytics*. John Wiley & Sons; Indianapolis, Indiana.
- Dempsey, M.; Valenti, A. M. (2016). Student use of keywords and limiters in web-scale discovery searching. *Journal of Academic Librarianship*, vol. 42(3), 200-206. <https://doi.org/10.1016/j.acalib.2016.03.002>
- Dick, S.; Yazdanbaksh, O.; Tang, X.; Huynh, T.; Miller, J. (2014). An empirical investigation of web session workloads: Can self-similarity be explained by deterministic chaos? *Information Processing and Management*, vol. 50(1), 41-53.
- Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pp. 226-231. AAAI Press; Menlo Park, California.
- Ferl, T. E.; Millsap, L. (1996). The knuckle-cracker's dilemma: a transaction log study of OPAC subject searching. *Information Technology and Libraries*, vol. 15(2), 81-98.
- Fisher, K. E.; Erdelez, S.; McKechnie, L. (editores) (2005). *Theories of information behavior*. Information Today; Medford, NJ, EE.UU.
- González-Teruel, A.; Barrios Cerrejón, M. (2012). *Métodos y técnicas para la investigación del comportamiento informacional: fundamentos y nuevos desarrollos*. Editorial Trea; Gijón.
- Guerbas, A.; Addam, O.; Zaarour, O.; Nagi, M.; Elhajj, A.; Ridley, M.; Alhajj, R. (2013). Effective web log mining and online navigational pattern prediction. *Knowledge-Based Systems*, vol. 49, 50-62. <https://doi.org/10.1016/j.knsys.2013.04.014>

- Gul, S.; Nabi, S.; Mushtaq, S.; Shah, T. A.; Ahmad, S. (2013). Political unrest and educational electronic resource usage in a conflict zone, Kashmir (indian administered Kashmir): Log analysis as politico analytical tool. *Information World*, vol. 14(2): 388-399.
- Halkidi, M.; Batistakis, Y.; Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, vol. 17, 107-145. <https://doi.org/10.1023/A:1012801612483>
- Hancock-Beaulieu, M. (1989). Online catalogues: a case for the user. En: Hildreth, C. R. (editor) *The online catalogue: developments and directions*. The Library Association; London.
- Hastie, T.; Tibshirani, R.; Friedman, J. (2009). The EM algorithm. En: Hastie, T.; Tibshirani, R.; Friedman, J. (autores) *The elements of statistical learning: data mining, inference, and prediction*. Springer; New York.
- Hershkovitz, A.; Hardof-Jaffe, S.; Nachmias, R. (2014). Content consumption and hierarchical structures of web-supported courses. *Journal of Interactive Learning Research*, vol. 25(3), 353-371.
- Huang, J.; White, R. W. (2010). Parallel browsing behavior on the web. *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia (HT '10)*, pp. 13-18. ACM; New York. <https://doi.org/10.1145/1810617.1810622>
- Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, vol. 2(3), 283-304. <https://doi.org/10.1023/A:1009769707641>
- Hunt, S.; Cimino, J. J.; Koziol, D. E. (2013). A comparison of clinicians' access to online knowledge resources using two types of information retrieval applications in an academic hospital setting. *Journal of the Medical Library Association*, vol. 101(1), 26-31. <https://doi.org/10.3163/1536-5050.101.1.005>
- Iyer, L. S.; Raman, R. M. (2011). Intelligent analytics: Integrating business intelligence and web analytics. *International Journal of Business Intelligence Research*, vol. 2(1), 31-45. <https://doi.org/10.4018/jbir.2011010103>
- Jansen, B. J. (2006). Search log analysis: what it is, what's been done, how to do it. *Library & Information Science Research*, vol. 28, 407-432. <https://doi.org/10.1016/j.lisr.2006.06.005>
- Jansen, B. J.; Pooch, U. (2001). A review of web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technology*, vol. 52(3), 235-246. [https://doi.org/10.1002/1097-4571\(2000\)9999:9999<::AID-ASI1607>3.0.CO;2-F](https://doi.org/10.1002/1097-4571(2000)9999:9999<::AID-ASI1607>3.0.CO;2-F)
- Jansen, B. J.; Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management*, vol. 42(1), 248-263. <https://doi.org/10.1016/j.ipm.2004.10.007>
- Jones, S.; Cunningham, S. J.; McNab, R.; Boddie, S. (2000). A transaction log analysis of a digital library. *International Journal on Digital Libraries*, vol. 3(2), 152-169. <https://doi.org/10.1007/s007999900022>
- Kahlon, M.; Yuan, L.; Daigre, J.; Meeks, E.; Nelson, K.; Piontkowski, C.; Reuter, K.; Sak, R.; Turner, B.; Weber, G. M.; Chatterjee, A. (2014). The use and significance of a research networking system. *Journal of Medical Internet Research*, vol. 16(2). <https://doi.org/10.2196/jmir.3137>
- Kapoor, K. (2010). Print and electronic resources: Usage statistics at Guru Gobind Singh Indraprastha University library. *Program: Electronic Library and Information Systems*, vol. 44(1), 59-68. <https://doi.org/10.1108/00330331011019690>
- Kaufman, L.; Rousseeuw, P. J. (2005). *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons; Hoboken, New Jersey.
- Lalmas, M.; O'Brien, H.; Yom-Tov, E. (2014). Measuring User Engagement. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 6(4), 1-132. <https://doi.org/10.2200/S00605ED1V01Y201410ICR038>
- Lambert, F. (2013). Seeking electronic information from government resources: A comparative analysis of two communities' web searching of municipal government websites. *Government Information Quarterly*, vol. 30(1), 99-109. <https://doi.org/10.1016/j.giq.2012.07.007>
- Larson, R. R. (1991). Classification clustering, probabilistic information retrieval, and the online catalog. *The Library Quarterly*, vol. 61(2), 133-173. <https://doi.org/10.1086/602331>
- Lai, Y.; Zeng, J. (2013). A cross-language personalized recommendation model in digital libraries. *The Electronic Library*, vol. 31(3), 264-277. <https://doi.org/10.1108/EL-08-2011-0126>
- Leeder, C.; Lonn, S. (2014). Faculty usage of library tools in a learning management system. *College & Research Libraries*, vol. 75(5), 641-663. <https://doi.org/10.5860/crl.75.5.641>
- Liu, Y.; Li, Z.; Xiong, H.; Gao, X.; Wu, J. (2010). Understanding of internal clustering validation measures. *Proceedings of the 10th IEEE International Conference on Data Mining*, pp. 911-916. IEEE Computer Society; Los Alamitos, California.
- Ma, H. (2013). Tech services on the web: Google Analytics. *Technical Services Quarterly*, vol. 30(1), 119-200. <https://doi.org/10.1080/07317131.2013.735978>
- Maabreh, M. A.; Al-Kabi, M.; Alsmadi, I. M. (2012). Query classification and study of university students' search trends. *Program: Electronic Library and Information Systems*, vol. 46(2), 220-241. <https://doi.org/10.1108/00330331211221855>
- MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical*

- Statistics and Probability*, pp. 281-297. University of California Press.
- Mahoui, M.; Jo Cunningham, S. (2000). A Comparative Transaction Log Analysis of Two Computing Collections. *Research and Advanced Technology for Digital Libraries: Proceedings of the 4th European Conference, ECDL 2000*, pp. 418-423. Springer; Berlin, Heidelberg. https://doi.org/10.1007/3-540-45268-0_53
- Malliari, A.; Moreli-Cacouris, M.; Kapsalis, K. (2010). Usage patterns in a greek academic library catalogue: A follow-up study. *Performance Measurement and Metrics*, vol. 11(1), 47-55. <https://doi.org/10.1108/14678041011026865>
- Markey, K. (2007). Twenty-five years of end-user searching, Part 2: Future research directions. *Journal of the American Society for Information Science and Technology*, vol. 58(8), 1123-1130. <https://doi.org/10.1002/asi.20601>
- Maulik, U.; Bandyopadhyay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24(12), 1650-1654. <https://doi.org/10.1109/TPAMI.2002.1114856>
- Mbabu, L. G.; Bertram, A.; Varnum, K. (2013). Patterns of undergraduates' use of scholarly databases in a large research university. *The Journal of Academic Librarianship*, vol. 39(2), 189-193. <https://doi.org/10.1016/j.acalib.2012.10.004>
- Moulaison, H. L.; Stanley, S. N. (2013). Beyond failure: Potentially mitigating failed author searches in the online library catalog through the use of linked data. *Journal of Web Librarianship*, vol. 7(1), 37-57. <https://doi.org/10.1080/19322909.2013.738562>
- Munson, D. M.; Otto, J. L. (2013). Have link resolvers helped or hurt? The relationship between ILL and OpenURL at a non-SFX library. *OCLC Systems & Services: International Digital Library Perspectives*, vol. 29(2), 78-86. <https://doi.org/10.1108/10650751311319287>
- Ortega Priego, J. L. (2004). Análisis del consumo de información de una revista electrónica: análisis de ficheros log de Cybermetrics. *Revista Española de Documentación Científica*, vol. 27(4), 455-468.
- Ortega Priego, J. L. (2005). Análisis de sesiones de la web del CINDOC: una aproximación a la minería de uso web. *El Profesional de la Información*, vol. 14(3), 190-198. <http://www.elprofesionaldelainformacion.com/contenidos/2005/mayo/4.pdf>
- Ozen, Z.; Bakiolu, F.; Beden, S. (2014). The examination of user habits through the Google Analytic data of academic education platforms. *International Journal of E-Adoption*, vol. 6(2), 31-45. <https://doi.org/10.4018/ijea.2014070103>
- Park, M.; Lee, T. S. (2016). A longitudinal study of information needs and search behavior in science and technology: a query analysis. *The Electronic Library*, vol. 34(1), 83-98. <https://doi.org/10.1108/EL-04-2014-0058>
- Park, M.; Lee, T. S. (2013). Understanding science and technology information users through transaction log analysis. *Library Hi Tech*, vol. 31(1), 123-140. <https://doi.org/10.1108/07378831311303976>
- Peeples, M. A. (2011). R script for k-means cluster analysis. <http://www.mattpeeples.net/kmeans.html> [Consulta: 20/08/2016]
- Peters, T. A. (1993). The history and development of transaction log analysis. *Library Hi Tech*, vol. 11(2), 41-66. <https://doi.org/10.1108/eb047884>
- Priya, R. V.; Vadivel, A. (2012). User behaviour pattern mining from weblog. *International Journal of Data Warehousing and Mining*, vol. 8(2), 1-22. <https://doi.org/10.4018/jdwm.2012040101>
- Rechavi, A.; Rafaeli, S. (2014). Active players in a network tell the story: Parsimony in modeling huge networks. *First Monday*, vol. 19(8). <https://doi.org/10.5210/fm.v19i8.5217>
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, vol. 20, 53-65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Rozaklis, L.; MacDonald, C. M. (2011). A typology of collaborative communication in a digital reference environment. *Reference Librarian*, vol. 52(4), 308-319. <https://doi.org/10.1080/02763877.2011.586907>
- Schalkoff, R. J. (2001). Pattern Recognition. En: *Wiley Encyclopedia of Electrical and Electronics Engineering*. John Wiley & Sons, Inc.; Indianapolis, Indiana.
- Shieh, J. (2012). From website log to findability. *The Electronic Library*, vol. 30(5), 707-720. <https://doi.org/10.1108/02640471211275747>
- Shiri, A. (2011). Revealing interdisciplinarity in nanoscience and technology queries: A transaction log analysis approach. *Knowledge Organization*, vol. 38(2), 135-153.
- Spink, A.; Jansen, B. J. (2004). *Web search: Public searching of the Web*. Kluwer; New York.
- Spiteri, L. F.; Tarulli, L. (2012). Social discovery systems in public libraries: If we build them, will they come? *Library Trends*, vol. 61(1), 132-147. <https://doi.org/10.1353/lib.2012.0019>
- Steinbach, M.; Karypis, G.; Kumar, V. (2000). A comparison of document clustering techniques. *KDD-2000 workshop on text mining*, pp. 525-526. Boston.
- Strohmaier, M.; Kroll, M. (2012). Acquiring knowledge about human goals from search query logs. *Information Processing and Management*, vol. 48(1), 63-82. <https://doi.org/10.1016/j.ipm.2011.03.010>
- Stuit, M.; Wortmann, H. (2012). Discovery and analysis of e-mail-driven business processes. *Information Systems*, vol. 37(2): 142-168. <https://doi.org/10.1016/j.is.2011.09.008>

- Tobias, C.; Blair, A. (2015). Listen to what you cannot hear, observe what you cannot see: An introduction to evidence-based methods for evaluating and enhancing the user experience in distance library services. *Journal of Library & Information Science in Distance Learning*, vol. 9(1-2), 148-156. <https://doi.org/10.1080/1533290X.2014.946354>
- Tu, Z. (2005). Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. *Tenth IEEE International Conference on Computer Vision*, vol. 2, pp. 1589-1596. IEEE; Beijing.
- Van Gemert-Pijnen, J.; Kelders, S. M.; Bohlmeijer, E. T. (2014). Understanding the usage of content in a mental health intervention for depression: An analysis of log data. *Journal of Medical Internet Research*, vol. 16(1). <https://doi.org/10.2196/jmir.2991>
- Verma, M.; Srivastava, M.; Chack, N.; Kumar, A.; Gupta, N. (2012). A comparative study of various clustering algorithms in data mining. *International Journal of Engineering Research and Applications*, vol. 2(3), 1379-1384.
- Velmurugan, T.; Santhanam, T. (2010). Computational complexity between k-means and k-medoids clustering algorithms for normal and uniform distributions of data points. *Journal of Computer Science*, vol. 6(3), 363-368. <https://doi.org/10.3844/jcssp.2010.363.368>
- Villén-Rueda, L.; Senso, J. A.; Moya-Anegón, F. de (2007). The use of OPAC in a large academic library: a transactional log analysis study of subject searching. *The Journal of Academic Librarianship*, vol. 33(3), 327-337. <https://doi.org/10.1016/j.acalib.2007.01.018>
- Vogt, W. P. (editor) (2011). *Sage quantitative research methods*. SAGE; Los Angeles. <https://doi.org/10.4135/9780857028228>
- Waller, V. (2010). Accessing the collection of a large public library: An Analysis of OPAC use. *LIBRES: Library and Information Science Research Electronic Journal*, vol. 20(1).
- Wang, C.; Ke, H.; Lu, W. (2012). Design and performance evaluation of mobile web services in libraries: A case study of the Oriental Institute of Technology library. *The Electronic Library*, vol. 30(1), 33-50. <https://doi.org/10.1108/02640471211204051>
- Wang, J.; Huffaker, D. A.; Treem, J. W.; Fullerton, L.; Ahmad, M. A.; Williams, D.; Poole, M. S.; Contractor, N. (2011a). Focused on the prize: Characteristics of experts in massive multiplayer online games. *First Monday*, vol. 16(8). <https://doi.org/10.5210/fm.v16i8.3672>
- Wang, P.; Berry, M. W.; Yang, Y. (2003). Mining longitudinal web queries: trends and patterns. *Journal of the American Society for Information Science and Technology*, vol. 54(8), 743-758. <https://doi.org/10.1002/asi.10262>
- Wang, S.; Zhang, J.; Yang, F.; Ye, J. (2014). Research on cluster analysis method of E-government public hotspot information based on web log analysis. *CIT - Journal of Computing and Information Technology*, vol. 22, 11-19. <https://doi.org/10.2498/cit.1002281>
- Wang, X.; Shen, D.; Chen, H.; Wedman, L. (2011b). Applying web analytics in a K-12 resource inventory. *The Electronic Library*, vol. 29(1), 20-35. <https://doi.org/10.1108/02640471111111415>
- Yom-Tov, E.; White, R. W.; Horvitz, E. (2014). Seeking insights about cycling mood disorders via anonymized search logs. *Journal of Medical Internet Research*, vol. 16(2). <https://doi.org/10.2196/jmir.2664>
- Zhang, J.; An, L. (2010). Visual component plane analysis for the medical based on transaction log. *The Canadian Journal of Information and Library Science*, vol. 34(1), 83-111. <https://doi.org/10.1353/ils.0.0006>
- Zhang, J.; Zhao, Y. (2013). A user term visualization analysis based on a social question and answer log. *Information Processing and Management*, vol. 49(5), 1019-1048. <https://doi.org/10.1016/j.ipm.2013.04.003>
- Zhu, D.; Guralnik, D.; Wang, X.; Li, X.; Moran, B. (2015). Statistical estimation for Single Linkage Hierarchical Clustering. *Proceedings of the IEEE 5th Annual International Conference on Cyber Technology in Automation, Control and Intelligent Systems (CYBER 2015)*, pp. 745-750. IEEE Computer Society; Los Alamitos, California. <https://doi.org/10.1109/CYBER.2015.7288035>

ESTUDIOS / RESEARCH STUDIES

Características generales de las revistas científicas peruanas

Julio Santillán-Aldana*, Mónica Arakaki**, Aurora de la Vega**, Mónica Calderón-Carranza**, Josmel Pacheco-Mendoza***

* Universidade de Brasília (UnB)

Correo-e: santillan@aluno.unb.br | ORCID iD: <http://orcid.org/0000-0002-1906-2370>

** Pontificia Universidad Católica del Perú (PUCP)

Correo-e: monica.arakaki@pucp.edu.pe | ORCID iD: <http://orcid.org/0000-0002-6543-7962>

**Pontificia Universidad Católica del Perú (PUCP)

Correo-e: avega@pucp.edu.pe | ORCID iD: <http://orcid.org/0000-0002-1242-2564>

** Pontificia Universidad Católica del Perú (PUCP)

Correo-e: mcalder@pucp.edu.pe | ORCID iD: <http://orcid.org/0000-0003-3127-3311>

*** Universidad San Ignacio de Loyola (USIL)

Correo-e: jpacheco@usil.edu.pe | ORCID iD: <http://orcid.org/0000-0002-2251-8092>

Recibido: 04-09-2016; 2ª versión: 02-01-2017; Aceptado: 09-01-2017

Cómo citar este artículo/Citation: Santillán-Aldana, J., Arakaki, M., De la Vega, A., Calderón-Carranza, M., Pacheco-Mendoza, J. (2017). Características generales de las revistas científicas peruanas. *Revista Española de Documentación Científica*, 40(3): e182. doi: <http://dx.doi.org/10.3989/redc.2017.3.1419>

Resumen: El texto presenta, por primera vez, un panorama situacional de las revistas científicas peruanas vigentes. A partir de los registros de LATINDEX y luego de un proceso de depuración, se identificaron 138 publicaciones y se caracterizaron sobre la base de cinco variables: procedencia geográfica, organización editora, periodicidad de publicación, formato de distribución y área temática. Se encontró que, en su mayoría, las revistas peruanas son editadas en Lima (66.7%) por una universidad (73.9%), tienen una periodicidad semestral (56.5%), circulan tanto en formato impreso como electrónico (67.4%) y publican principalmente artículos sobre ciencias (agrícolas, de la ingeniería, exactas, naturales y médicas) (35.5%). El estudio señala, además, ciertas situaciones desfavorables y oportunidades de mejora en el sector editorial científico peruano.

Palabras clave: Revistas científicas; comunicación científica; Perú.

General characteristics of Peruvian scientific journals.

Abstract: This paper presents for the first time an overview of the state of current Peruvian journals. After retrieving and filtering records from LATINDEX, 138 journals were identified and characterized by the following variables: geographical origin, publisher, frequency, format (print, electronic, or both), and subject. The results show that Peruvian journals are mostly published in Lima (66.7%) by a university (73.9%), come out every six months (56.5%) both in print and electronically (67.4%), and publish articles mainly about natural sciences (35.5%). The article points out some problematic issues as well as certain measures for improving the Peruvian journal publishing market.

Keywords: Scientific journals; scholarly communication; Peru.

Copyright: © 2017 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

Aunque la producción de los investigadores peruanos ha sido analizada en el contexto latinoamericano (Santa y Herrero Solana, 2010; Hernández Asensio, 2014), se ha escrito relativamente poco sobre el otro indicador del grado de institucionalización de la ciencia en el Perú: las revistas científicas. En muchas disciplinas, ellas constituyen el principal vehículo de difusión del conocimiento (Garvey y Griffith, 1972; Fyfe y otros, 2015).

Algunos estudios han dado cuenta de la inclusión y visibilidad de las publicaciones peruanas en bases de datos internacionales (Huamaní y Pacheco-Romero, 2009; Miguel, 2011). Sin embargo, se desconoce su estado en términos de características editoriales.

Este artículo presenta una primera aproximación a las revistas científicas que se publican en el Perú, sobre la base de cinco variables: procedencia geográfica, organización editora, periodicidad de publicación, formato de distribución y área temática. Se espera que los resultados permitan a los actores competentes imprimir dinamismo y elevar la calidad del sector editorial científico del país.

2. ANTECEDENTES

De acuerdo con Guibovich (2003), entre los primeros antecedentes de publicaciones periódicas científicas en el Perú se encuentra *El Conocimiento de los Tiempos*, un conjunto de pronósticos astronómicos y matemáticos publicados desde 1680 por el cosmógrafo mayor del Virreinato del Perú, importante autoridad científica durante la época de la colonia.

Las publicaciones científicas en el Perú inician un auge en el siglo XVIII. En 1791, aparece *El Mercurio Peruano*, el bisemanario de la Sociedad Académica de Amantes del País, considerada la primera publicación periódica de enfoque científico desarrollada en territorio peruano y con amplia difusión en las principales ciudades de Hispanoamérica (Milla, 1986; Tauro del Pino, 1987). Según Pamo Reyna (2005), en los cuatro años de vigencia de esta publicación, el 25% de sus contenidos estuvo referido a las ciencias. Como referencia, *El Mercurio Volante*, primera revista médica del continente, había hecho su aparición en México en 1772 (Bartolache, 1979; López Espinosa, 2000).

Para 1827, en los inicios del periodo republicano, Abel Victorino Brandin publica *Anales Medicas*, semanario de medicina, cirugía, boticaria e historia natural, considerada la primera revista médica peruana (Valdizán, 1929; Pamo Reyna, 1997). Por su parte, Cueto (1989) da cuenta de otras publicaciones destacadas como *El Siglo*, editado por la So-

ciudad de Amantes del Saber entre 1874 y 1879; los *Anales de Construcciones Civiles y de Minas*, editados desde 1880 por la Escuela de Ingeniería; y el *Boletín de la Sociedad Geográfica de Lima*, que se publica desde 1891 hasta la actualidad.

También existieron asociaciones y publicaciones vinculadas a hospitales o a grupos particulares de médicos. Entre estas destacó *La Gaceta de los Hospitales*, que apareció en 1903 y cuyo último número data de 1913. Otras revistas del mismo estilo fueron: *Farmacia Peruana*, publicada irregularmente entre 1893 y 1916 por la Sociedad Farmacéutica de Lima; *La Reforma Médica*, publicada por los doctores Carlos Enrique Paz Soldán y Baltazar Caravedo entre 1915 y 1930; *La Crónica Dental*, publicación de la Asociación de Estudiantes de Odontología, fundada en 1917; y la *Revista de Psiquiatría y Disciplinas Conexas*, creada en 1918 por el doctor Honorio Delgado. Algunas de estas publicaciones continuaron hasta los años veinte del siglo XX (Cueto, 1989).

Destacan, también, los *Anales de la Facultad de Medicina*, revista fundada en 1918, bajo la dirección del doctor Ernesto Odriozola y que tuvo como secretario de edición al doctor Hermilio Valdizán. Fue, en su momento, la publicación médica peruana más importante. En ella se difundieron estudios sobre la biología andina, de gran repercusión en el desarrollo científico del país, como lo señala Guerra García (1979). Años más tarde, en 1938, aparece la *Revista de Neuro Psiquiatría*, fundada por los doctores Honorio Delgado y Oscar Trelles. Tanto los *Anales de la Facultad de Medicina* como la *Revista de Neuro Psiquiatría* continúan publicándose en la actualidad; la primera, por la Universidad Nacional Mayor de San Marcos y la segunda, por la Universidad Peruana Cayetano Heredia (Pacheco, 2008; Alarcón y otros, 2015).

Salvo para el caso de medicina, que es el ámbito científico más desarrollado y de mayor tradición en el Perú, no existe un recuento detallado de la evolución de las revistas científicas en el país. Un caso de especial atención, que advierte sobre la necesidad de la publicación oportuna de hallazgos en revistas científicas es el que se reporta Cueto (1989) sobre el científico Santiago Antúnez de Mayolo:

Uno de los cuestionamientos principales a Antúnez de Mayolo es al reclamo que hizo por tener la prioridad en el descubrimiento del elemento neutrón en la composición del átomo. El descubrimiento del neutrón se atribuye al británico James Chadwick que ganó el premio Nobel por su hallazgo en 1936. Según Antúnez, él había propuesto con anterioridad la existencia del neutrón en una comunicación presentada al Tercer Congreso Científico Panamericano celebrado en

Lima a fines de 1924 [...]. Sin embargo, este trabajo nunca fue publicado en una revista científica y solo fue conocido, años más tarde, a través del diario El Tiempo [...]. Según los científicos contemporáneos, Chadwick, a diferencia de Antúnez de Mayolo, sustentó su descubrimiento con evidencia experimental y publicó sus trabajos en revistas internacionales especializadas.

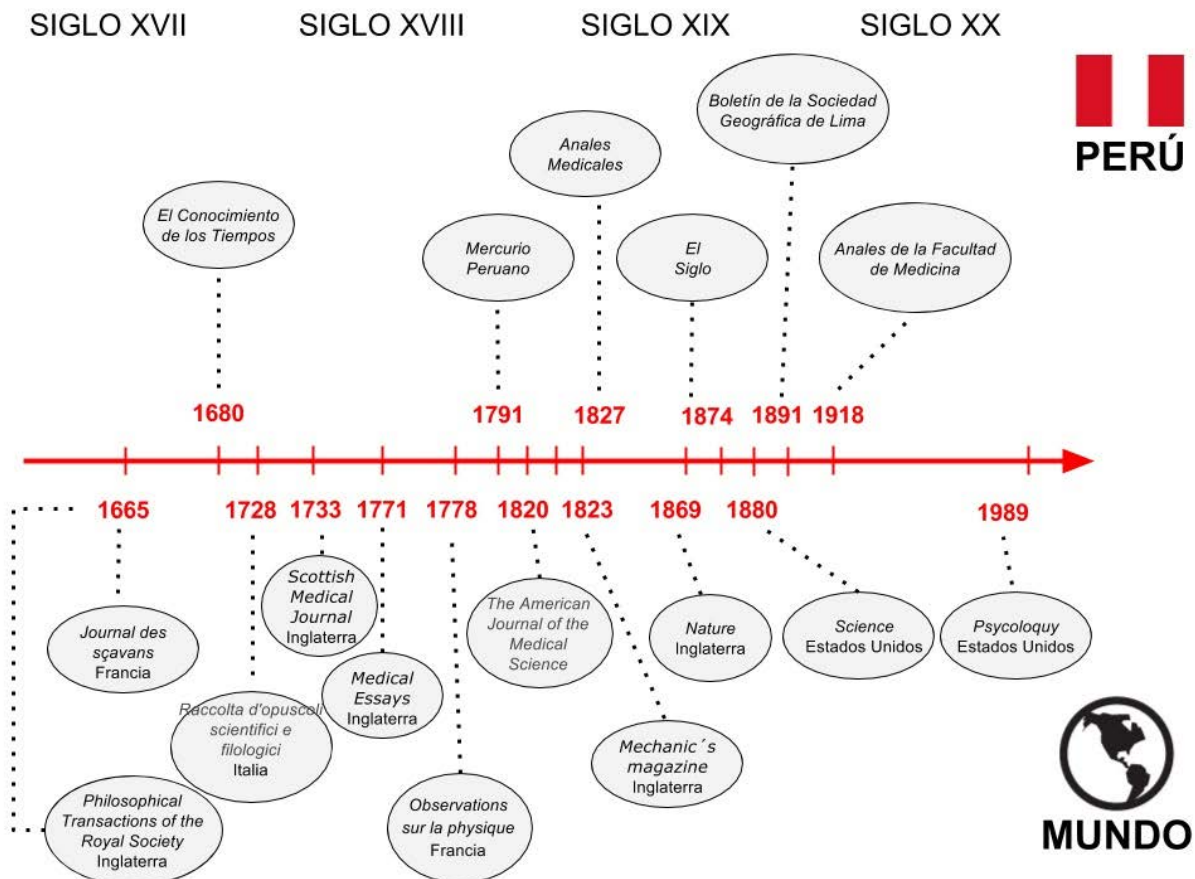
A pesar de su carácter emblemático, las revistas de medicina no han estado exentas de dificultades. Pamo Reyna (2005) describe como limitaciones el hecho de que la mayoría se publica en Lima, con retraso, con un tiempo de vida menor de cinco años, o sin ISSN; o que los editores no estén debidamente formados o cambien el formato de la publicación con cierta frecuencia.

Es solo a partir de mediados de la década pasada que las revistas científicas peruanas comienzan a incorporarse masivamente a los índices internacionales, por lo que aún se nota un retraso considerable

respecto a materias distintas a la medicina (Hernández Asensio, 2014). Por ejemplo, el análisis en SCImago Journal & Country Rank revela que, inicialmente, las revistas indexadas pertenecían al entorno de la medicina y ciencias afines (Hernández Asensio, 2014). La predominancia de la medicina, en el caso peruano, ya había sido evidenciada por Huamaní y Mayta-Tristán (2010). Según estos autores, la producción científica en medicina, representada en la base de datos multidisciplinar Web of Science, aunque todavía escasa, estaba en crecimiento, pero concentrada en Lima y en pocas instituciones. Las más productivas colaboraban más intensamente con instituciones extranjeras que con nacionales, a partir de redes de colaboración.

A manera de resumen, la Figura 1 presenta una línea del tiempo de las revistas científicas peruanas más significativas y sus homólogas internacionales entre los siglos XVII y XX.

Figura 1. Línea del tiempo de las revistas científicas peruanas



Fuente: Bishop (1984), Cueto (1989), Houghton (1975)

3. METODOLOGÍA

El presente estudio, de carácter descriptivo, utilizó como fuente de datos el Directorio Latindex (Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal), por considerarse que concentra el mayor número de registros sobre revistas científicas peruanas. Los datos fueron recopilados entre mayo y junio de 2015, y procesados y normalizados en los cuatro meses siguientes.

Se utilizó la opción de *búsqueda avanzada* de Latindex para determinar el conjunto preliminar de estudio. En este se incluyeron todas las publicaciones peruanas clasificadas como *revista de investigación científica* en situación vigente.

Después de fusionar aquellos resultados que, compartiendo el mismo título, Latindex diferenciaba como revista *impresa* y *en línea*, se obtuvo un grupo de 253 títulos únicos. Bajo la premisa de que estos registros podrían estar desactualizados, se emplearon fuentes en línea para verificar y, en caso que fuera necesario, modificar el estado de vigencia reportado por Latindex. Para ello, se definieron las siguientes categorías para las revistas:

- Vigente, si el último número había sido publicado en 2013 o 2014.
- Descontinuada, si el último número tenía un año de publicación anterior a 2013.
- Indeterminada, si no había podido recabarse información sobre su vigencia.

Los resultados, luego de verificar la situación de vigencia, se muestran en la Tabla I. Se aprecia que poco más de la mitad de las revistas —138 de las 253 publicaciones (54.54%)—, se encontraron vigentes. La desactualización del directorio Latindex en términos de vigencia había sido también advertida por Liberatore y otros (2013) en un estudio sobre revistas argentinas en ciencias sociales y humanidades.

Tabla I. Situación de vigencia de las revistas peruanas

Estado	Nº de revistas	Porcentaje
Vigente	138	54.5%
Descontinuada	68	26.9%
Indeterminada	47	18.6%
TOTAL	253	100%

El presente estudio se centra en este conjunto final de 138 revistas y las analiza considerando las siguientes variables: procedencia geográfica, organización editora, periodicidad de publicación, formato de distribución (impreso, electrónico) y área temática.

Los datos se recopilaron, en su mayor parte, consultando los sitios web de las publicaciones. En los casos donde la información era insuficiente o incompleta, se contactó a los editores por correo electrónico o por vía telefónica.

4. RESULTADOS Y DISCUSIÓN

El hallazgo de 138 revistas peruanas vigentes contrasta, tanto con lo registrado por SciELO Perú, como por los directorios DOAJ o Ulrich, según valores obtenidos en julio de 2016 (Tabla II).

Tabla II. Revistas peruanas registradas en directorios

Servicio	Nº revistas peruanas
Ulrich's Periodicals Directory	202
DOAJ (Directory of Open Access Journals)	33
SciELO Perú	23

Esta diferencia en cifras se debe a los criterios de admisión de cada sistema. Mientras que SciELO Perú "abarca una colección seleccionada de revistas científicas peruanas" (SciELO Perú, 2016), DOAJ comprende "revistas científicas de alta calidad, de acceso abierto y revisadas por pares" (DOAJ, 2016) y Ulrich contiene revistas tanto de investigación como de divulgación. De forma similar a Ulrich, Latindex tiene una política de indización más exhaustiva, al registrar "revistas de investigación científica, técnico-profesionales y de divulgación científica y cultural" (Latindex, 2016).

Como se mencionó anteriormente, cuando no fue posible determinar el estado real de vigencia de las revistas utilizando recursos en línea, se estableció contacto con los editores por teléfono o correo electrónico. Varias llamadas se hicieron a universidades de las zonas centro y sur del Perú. En algunos casos, se encontraron revistas vigentes, publicadas en formato impreso, sin referencia alguna en línea. En otros casos, se identificaron publicaciones con retraso o descontinuadas. Las causas de esta última situación, indicadas por los editores contactados, fueron las siguientes:

- La institución no cuenta con recursos económicos suficientes para subvencionar la publicación de las revistas.
- La institución cuenta con recursos económicos, pero no está en sus políticas subvencionar la publicación de las revistas.
- La institución cuenta con recursos económicos y subvenciona la publicación de revistas, pero solo de algunas.

4.1. Procedencia geográfica

Esta variable permite identificar puntos de concentración (clústeres) para la investigación, así como las vinculaciones con las organizaciones responsables de las publicaciones.

La Tabla III muestra cómo es que las 138 revistas están repartidas en el territorio peruano, considerando la circunscripción sede de la organización editora. Las revistas son publicadas en 16 de las 25 entidades territoriales de mayor nivel (departamento o provincia constitucional). No se identificaron revistas científicas vigentes en nueve departamentos: Amazonas, Apurímac, Huancavelica, Huánuco, Ica, Loreto, Madre de Dios, Moquegua y Pasco.

Tabla III. Revistas por procedencia geográfica

Departamento	N° Revistas	Porcentaje
Lima	92	66.7%
La Libertad	9	6.5%
Lambayeque	7	5.1%
Tacna	4	2.9%
Junín	4	2.9%
Arequipa	4	2.9%
Puno	3	2.2%
Ucayali	2	1.4%
Piura	2	1.4%
Cusco	2	1.4%
Callao	2	1.4%
Cajamarca	2	1.4%
Ancash	2	1.4%
Tumbes	1	0.7%
San Martín	1	0.7%
Ayacucho	1	0.7%
TOTAL	138	100%

Se aprecia que Lima es el departamento que concentra el mayor número de revistas con 66.7% del total (92). Le sigue el departamento de La Libertad, con apenas un 6.5% (9). El resto de departamentos registra de una a siete revistas científicas.

Se tiene que dos de cada tres publicaciones se editan en Lima. Este hallazgo respalda lo expresado por Pamo Reyna (2005) y, Huamaní y Mayta-Tristán (2010), quienes habían denunciado el centralismo en la producción de las revistas.

Una situación similar estaría experimentándose en, al menos, otros dos países de la región, según reportes. En Chile, el 59.4% de las revistas académicas se editan en la Región Metropolitana (Lara y Martínez, 2014) y en México, este valor asciende a 74% en la región centro para el conjunto de publicaciones de investigación, de divulgación científica y de naturaleza técnico-profesional (Alonso Gamboa y Reyna Espinosa, 2015).

4.2. Organización editora

En el contexto del presente estudio, identificar la organización editora significa determinar las instituciones que hacen sostenibles las publicaciones. Este apoyo va desde avalar nominalmente la revista, hasta financiarla o facilitar recursos al proceso editorial.

La Tabla IV muestra la cantidad de revistas según la tipología definida para clasificar las organizaciones responsables de su publicación: asociación profesional, instituto de investigación, sociedad científica, universidad y otros.

Tabla IV. Revistas por tipo de organización editora

Tipo de organización	N° Revistas	Porcentaje
Universidad	102	73.9%
Asociación profesional	8	5.8%
Instituto de investigación	8	5.8%
Sociedad científica	8	5.8%
Otros (institutos de capacitación, ONG, asociaciones independientes, etc.)	12	8.7%
TOTAL	138	100%

En el grupo mayoritario de 102 publicaciones, pudo constatarse que, prácticamente, dos tercios (67.6%) son editadas por universidades privadas; el tercio restante (32.4%) pertenece al sector público. Véase la Tabla V. Este hallazgo, es decir, la prevalencia de universidades privadas como instituciones editoras, coincide con el reportado por Morales Morante (2016) para revistas peruanas en ciencias sociales. El hecho de que las universidades privadas cuenten, en la mayoría de los casos, con fondos editoriales, parece contribuir a la sostenibilidad de las revistas.

Si se considera la totalidad de las publicaciones vigentes (138), la balanza vuelve a inclinarse en una proporción muy similar (2:1) hacia el sector privado: 68.8% de las revistas peruanas vigentes son editadas por organizaciones privadas y 31.2% por entidades públicas. Véase la Tabla VI.

Tabla V. Revistas editadas por universidades

Tipo de universidad	N° Revistas	Porcentaje
Privada	69	67.6%
Pública	33	32.4%
TOTAL	102	100%

Tabla VI. Revistas por sector al que pertenece la organización editora

Tipo de organización editora	N° Revistas	Porcentaje
Privada	95	68.8%
Pública	43	31.2%
TOTAL	138	100%

Se observa que, en el Perú, las organizaciones editoras corresponden, en su mayoría (73.9%), a universidades, dos tercios de las cuales son de carácter privado. En Chile, el 65.7% de las publicaciones son editadas por universidades (Lara y Martínez, 2014). En México, poco más de la mitad de las publicaciones (53%) están respaldadas por una institución de educación superior y el resto, de forma más o menos equitativa, por sociedades académicas o profesionales, privadas, gubernamentales y de investigación (Alonso Gamboa y Reyna Espinosa, 2015). Situaciones similares se reportan en Uruguay (Valenzuela y Machado, 2015) y Argentina (Flores y Casado, 2015), donde las entidades editoras que más contribuyen en cantidad de publicaciones son las universidades. Téngase en cuenta que, a diferencia de los dos primeros casos, los estudios mexicano, uruguayo y argentino incluyen, además de las publicaciones estrictamente de investigación, las de divulgación científica y las de corte técnico-profesional.

4.3. Periodicidad de publicación

La periodicidad de la revista y la puntualidad de publicación reflejan, en la práctica, su capacidad para sustentar un flujo constante de artículos y la eficiencia de su gestión editorial (Gonçalves y otros, 2006).

Según se aprecia en la Tabla VII, la periodicidad más común en el Perú es la semestral (56.5%), seguida por la anual (23.2%). En Chile, si bien es cierto la frecuencia de publicación más común también es la semestral, con un 59%, el segundo lugar lo ocupa la cuatrimestral, con un 16.3% (Lara y Martínez, 2014). Por su parte, en México, los dos primeros lugares corresponden a la periodicidad semestral, con un 26% y trimestral, con un 22% (Alonso Gamboa y Reyna Espinosa, 2015).

Tabla VII. Revistas por periodicidad de publicación

Periodicidad	N° Revistas	Porcentaje
Semestral	78	56.5%
Anual	32	23.2%
Trimestral	17	12.3%
Cuatrimestral	5	3.6%
N/D	2	1.4%
Bienal	2	1.4%
Irregular	1	0.7%
Mensual	1	0.7%
TOTAL	138	100%

Se tiene así que, en el Perú, casi el 80% de las revistas aparecen con una frecuencia anual o semestral, a diferencia de otros países de la región, que exhiben ritmos más dinámicos de producción editorial. La prevalencia de publicaciones semestrales y anuales sugiere, en el caso peruano, una capacidad moderada-baja en el procesamiento de artículos. Los periodos más largos de edición —que van en detrimento de la comunicación oportuna de los resultados de investigación, sobre todo en las ciencias— podrían explicarse, en parte, por las situaciones de escasez de recursos institucionales, mencionadas al inicio de este acápite.

4.4 Formato de distribución

Según López Ornelas y Cordero Arroyo (2005) y Laakso y otros (2011), una gran parte de las revistas científicas en el mundo se encuentra en proceso de transición de la versión impresa a la electrónica y, al parecer, las publicaciones peruanas no serían ajenas a este fenómeno.

Los hallazgos en este rubro se resumen en la Tabla VIII.

Tabla VIII. Revistas por formato de distribución

Distribución	N° Revistas	Porcentaje
Tanto impreso como electrónico	93	67.4%
Solo electrónico	27	19.6%
Solo impreso	18	13%
TOTAL	138	100%

En el grupo de revistas con volúmenes electrónicos (ya sea como complemento a la versión impresa o de forma exclusiva), se encontró que 44 se valen de una plataforma de gestión. Puede especularse que la cada vez mayor disponibilidad de plataformas en línea, gratuitas en muchos casos, ha facilitado la transición a medios digitales. Cabe destacar que *Anales de la Facultad de Medicina*, una de las publicaciones peruanas más antiguas, editada desde 1918 por la Universidad Nacional Mayor de San Marcos, migró a su actual versión electrónica en 2005.

La situación en el Perú tiene cierta similitud con la de Chile, donde el 76.1% de las publicaciones se distribuye tanto en formato impreso como electrónico, el 19.5% circula exclusivamente por medios electrónicos y el 4.4% solo por medios impresos (Lara y Martínez, 2014). México, en cambio, presenta una prevalencia del formato impreso con un 72%, seguida por la opción híbrida con un 22% y la versión electrónica con un 6% (Alonso Gamboa y Reyna Espinosa, 2015).

4.5 Área temática

Las publicaciones se clasificaron, además, por la temática cubierta. Para ello, se utilizó un esquema adaptado de Latindex, que organiza el conocimiento en cuatro grandes áreas:

- Artes y humanidades.
- Ciencias (agrícolas, ciencias de la ingeniería, ciencias exactas y naturales, ciencias médicas).
- Ciencias sociales.
- Multidisciplinarias.

De acuerdo con este esquema, disciplinas como derecho, educación, historia, arqueología, antropología, psicología o ciencias de la información se encuentran en el tercer grupo (ciencias sociales); mientras que filosofía, lingüística, literatura, filología y diseño se enmarcan en el primer grupo (artes y humanidades).

Las 138 publicaciones quedaron clasificadas como se muestra en la Tabla IX.

Tabla IX. Revistas por área temática general

Área temática general	N° Revistas	Porcentaje
Ciencias	49	35.5%
Multidisciplinarias	42	30.4%
Ciencias sociales	40	29%
Artes y humanidades	7	5.1%
TOTAL	138	100%

Se aprecia que el área temática con mayor número de ítems es el de las ciencias, aunque no de una manera categórica. La clasificación de las 49 publicaciones de este grupo arrojó los resultados que se muestran en la Tabla X.

Tabla X. Revistas sobre ciencias

Subárea	N° Revistas	Porcentaje
Ciencias médicas	26	53.1%
Ciencias exactas y naturales	14	28.6%
Ciencias de la ingeniería	6	12.2%
Ciencias agrícolas	3	6.1%
TOTAL	49	100%

Cabe mencionar que entre las 26 revistas de ciencias médicas se encuentra una de las publicaciones científicas más antiguas del Perú: *Anales de la Facultad de Medicina*, editada por la Universidad Nacional Mayor de San Marcos desde 1918.

La tabla XI muestra el conteo en el grupo de las 40 publicaciones sobre ciencias sociales.

Tabla XI. Revistas sobre ciencias sociales

Subárea	N° Revistas	Porcentaje
Psicología	10	25%
Derecho	7	17.5%
Educación	4	10%
Economía	4	10%
Otros	15	37.5%
TOTAL	40	100%

Estos resultados coinciden ligeramente con los reportados por Morales Morante (2016) en su estudio sobre 25 publicaciones peruanas en ciencias sociales. Su investigación identificó como disciplinas más representadas a la psicología, la administración de empresas unida a la economía, y el derecho (en ese orden).

Entre los 15 ítems del rubro *otros* se encontraron revistas sobre arqueología (3), sociología (2), historia (2), ciencia política y administración pública (2), así como tecnología (1), ciencias de la información (1), turismo (1), contabilidad (1), administración (1) y ciencias de la comunicación (1).

Finalmente, las siete revistas sobre artes y humanidades se dedican a las temáticas que se muestran en la Tabla XII.

Tabla XII. Revistas sobre artes y humanidades

Subárea	Nº Revistas	Porcentaje
Filosofía	2	28.6%
Lingüística	2	28.6%
Literatura	1	14.3%
Arte	1	14.3%
Humanidades	1	14.3%
TOTAL	7	100%

La ligera prevalencia de las publicaciones sobre ciencias y, en especial, de las ciencias médicas, parece ser una situación muy particular del Perú. En el contexto latinoamericano, otras investigaciones habían dado cuenta del predominio de las revistas de ciencias sociales. En México, el 42% de las publicaciones desarrollan temáticas vinculadas a las ciencias sociales y en segundo lugar, el 16% corresponde a las ciencias médicas (Alonso Gamboa y Reyna Espinosa, 2015). En Uruguay, las ciencias sociales vuelven a ocupar la primera posición (47%), seguidas también por las ciencias médicas (16%) (Valenzuela y Machado, 2015). Estos dos campos del conocimiento son, de la misma forma, los más cubiertos en Argentina (Flores y Casado, 2015). En Chile, el porcentaje mayoritario (28.8%) lo exhiben las publicaciones sobre ciencias sociales, historia, filosofía y humanidades; seguidas de las revistas sobre medicina y ciencias de la salud (16.6%) (Lara y Martínez, 2014). Nótese, sin embargo, que en el caso chileno se están consolidando las publicaciones sobre ciencias sociales junto con las de artes y humanidades, de acuerdo con la clasificación temática del presente estudio.

5. REFLEXIONES FINALES

En resumen, el presente estudio identificó 138 revistas científicas peruanas. Se encontró que, en su mayoría, son editadas en Lima (66.7%) por una universidad (73.9%), tienen una periodicidad semestral (56.5%), circulan tanto en formato impreso como electrónico (67.4%) y publican, principalmente, artículos sobre ciencias agrícolas, de la ingeniería, exactas, naturales y médicas (35.5%).

Si bien son notables los esfuerzos editoriales por comunicar la ciencia en el país, resulta preocupante que un número significativo de revistas se haya visto en la necesidad de retrasar o, en el peor de los casos, interrumpir (temporal o definitivamente) la publicación de artículos, debido a la escasez o falta de recursos institucionales. Cabe recordar que, del grupo inicial de 253 revistas identificadas, 115 (es decir, el 45%) se encontraron descontinua-

das o en situación indeterminada. Morales Morante (2016) también daba cuenta de esta irregularidad en revistas peruanas del ámbito de las ciencias sociales. En ese sentido, es crucial tomar acciones que contribuyan a fomentar y dar sostenibilidad a los aparatos editoriales científicos, desde las instituciones y el gobierno, y con particular énfasis, en las regiones del interior del país.

Por ello, resulta alentador —aunque todavía insuficiente—, la puesta en marcha de algunas medidas gubernamentales. Destaca el fondo concursal *Publicación de Revistas Científicas Peruanas Indizadas* (CONCYTEC, 2015), cuya finalidad es “fortalecer y elevar la calidad editorial de las revistas científicas y tecnológicas peruanas indizadas, para que logren su inserción en bases de datos bibliográficas internacionales”. Este fondo concursal en el Perú es similar a los lanzados en otros países de la región, como Brasil (CAPES, 2016), México (CONACYT, 2016) y Chile (CONICYT, 2015).

En términos más amplios, el incremento del gasto en investigación y desarrollo es un punto pendiente en la agenda del gobierno peruano (Robles Alfaro y otros, 2016). La creación obligatoria de Vicerrectorados de Investigación en todas las universidades del país, según estipula la Ley No. 30220, promulgada en 2014, es también una oportunidad para impulsar la producción científica y la publicación de trabajos desde la academia. En el ámbito profesional, cabe unir esfuerzos alrededor de agrupaciones como la Asociación Peruana de Editores Científicos - APECI (Gálvez Calla, 2006), uno de cuyos fines es fomentar la publicación de artículos y revistas científicas peruanas con estándares internacionales de calidad. Otras acciones que los editores científicos peruanos pueden poner en práctica han sido señaladas por Dorta Contreras (2016) en el caso cubano.

Por otro lado, se ha reportado que una parte significativa de las investigaciones realizadas en Latinoamérica se publica fuera de la región (Miguel, 2011; Fischman y Alperin, 2015). Esta preferencia por las revistas extranjeras sobre las locales debe ser considerada por los gestores de políticas de incentivo a la investigación. No solo hay que investigar más en el Perú, también hay que promover que los resultados de esas investigaciones se publiquen localmente, si se busca imprimir más dinamismo al sector editorial. Sin embargo, ningún incentivo tendría efecto si los investigadores no encuentran espacios de publicación de calidad. Ese es, tal vez, el aspecto en el que hay que trabajar más, además de la cuestión de sostenibilidad que se señalaba casi al principio de este apartado. Tal como Alonso Gamboa y Reyna Espinosa (2015) bien señalan, “el trinomio pervivencia + calidad + prestigio en las publicaciones académicas les [permitirían] no solo su inclusión en las bases

de datos más selectivas, sino abonar en una mayor visibilidad, internacionalización y reconocimiento”.

El panorama encontrado no es halagüeño. Sin embargo, sin una radiografía sería imposible trazar estrategias de mejora. Apostar por el desarrollo del sector editorial científico implica trabajar desde diferentes frentes y en condiciones, en muchos casos, desfavorables. En este escenario, no bastan esfuerzos individuales o aislados.

6. AGRADECIMIENTOS

Los autores desean expresar su agradecimiento a los editores que gentilmente proporcionaron datos sobre sus respectivas publicaciones: *Acta Médica Orreguiana Hampi Runa* (Trujillo), *Cajamarca* (Cajamarca), *Conocimiento Amazónico* (Iquitos), *Revista de Derecho* (Piura), *Revista de Investigaciones Altoandinas* (Puno), *Revista de Psicología* (Arequipa), *Revista Forestal de Ucayali* (Pucallpa), *Revista Universitaria* (Cusco), *Sistemas Agroecológicos y Modelos Biomatemáticos* (Tarapoto), *Bo-*

letín de la Sociedad Geográfica de Lima, *Revista Peruana de Psicología y Trabajo Social* (Lima), *Revista Odontología Pediátrica* (Lima), así como del *Boletín, Informe y Anuario Científico Tecnológico* del Instituto del Mar del Perú (Lima).

7. ACKNOWLEDGEMENTS

For kindly providing information about their journals, the authors are grateful to the editors of *Acta Médica Orreguiana Hampi Runa* (Trujillo), *Cajamarca* (Cajamarca), *Conocimiento Amazónico* (Iquitos), *Revista de Derecho* (Piura), *Revista de Investigaciones Altoandinas* (Puno), *Revista de Psicología* (Arequipa), *Revista Forestal de Ucayali* (Pucallpa), *Revista Universitaria* (Cusco), *Sistemas Agroecológicos y Modelos Biomatemáticos* (Tarapoto), *Boletín de la Sociedad Geográfica de Lima*, *Revista Peruana de Psicología y Trabajo Social* (Lima), *Revista Odontología Pediátrica* (Lima), along with *Boletín, Informe* and *Anuario Científico Tecnológico* of the Instituto del Mar del Perú (Lima).

REFERENCIAS

- Alarcón, R. D.; Querol, M.; Mori, G.; Jerí, F. R.; Ponce, M.; Sarabia, S.; Custodio, N.; Guillén, D.; Vega-Dienstmaier, J.; Cruzado, L. (2015). Tiempo, legados y promesas: coloquio intergeneracional en torno a las Bodas de Diamante de la Revista de Neuro-Psiquiatría. *Revista de Neuro-Psiquiatría*, vol. 78 (3), 179-188. <https://doi.org/10.20453/rnp.v78i3.2578>
- Alonso Gamboa, J. O.; Reyna Espinosa, F. R. (2015). Revistas académicas mexicanas. Panorama y perspectiva. *CIENCIA ergo-sum*, vol. 22 (3), 181-191. <http://cienciaergosum.uaemex.mx/index.php/ergosum/article/view/1876/3349> [Consulta: 21/12/2016]
- Bartolache, J. I. (1979). *Mercurio Volante: 1772-1773*; Universidad Nacional Autónoma de México; Ciudad de México.
- Bishop, C. (1984). *How to edit a scientific journal*; ISI Press; Philadelphia.
- CAPES (2016). *Editores e Publicação de Periódicos Científicos Brasileiros*. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; Brasília. <http://www.capes.gov.br/bolsas/programas-especiais/capes-editoracao> [Consulta: 21/12/2016]
- CONACYT (2016). *Sistema de Clasificación de Revistas Mexicanas de Ciencia y Tecnología*. Consejo Nacional de Ciencia y Tecnología; Ciudad de México. <http://www.conacyt.gob.mx/index.php/comunicacion/sistema-de-clasificacion-de-revistas-mexicanas-de-ciencia-y-tecnologia> [Consulta: 21/12/2016]
- CONCYTEC (2015). *Publicación de Revistas Científicas Peruanas Indizadas*. Fondo Nacional de Desarrollo de Ciencia, Tecnología e Innovación Tecnológica (CIENCIACTIVA); Lima.
- CONICYT (2015). *Concurso Fondo de Publicación de Revistas Científicas Chilenas 2015*. Comisión Nacional de Investigación Científica y Tecnológica; Santiago. <http://www.conicyt.cl/informacioncientifica/2015/07/07/concurso-fondo-de-publicacion-de-revistas-cientificas-chilenas-2015/> [Consulta: 21/12/2016]
- Cueto, M. (1989). *Excelencia científica en la periferia. Actividades científicas e investigación biomédica en el Perú (1890-1950)*; GRADE, CONCYTEC; Lima.
- DOAJ (2016). <https://doaj.org/> [Consulta:21/12/2016]
- Dorta Contreras, A. J. (2016). Desarrollar la cultura de autor: educar y estimular la producción científica en revistas cubanas. *Medisur*, vol. 14 (3). http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1727-897X2016000300006&lng=es&nrm=iso [Consulta: 28/12/2017]
- Fischman, G. E.; Alperin, J. P. (2015). Sobre luces y sombras: las revistas científicas hechas en Latinoamérica. En: Alperin, J. P.; Fischman, G. (editores). *Hecho en Latinoamérica: acceso abierto, revistas académicas e innovaciones regionales*. CLACSO; Ciudad Autónoma de Buenos Aires. <http://biblioteca.clacso.edu.ar/clacso/se/20150722110704/HechoEnLatinoamerica.pdf> [Consulta: 21/12/2016]
- Flores, A. M.; Casado, A. (2015). El sistema Latindex en Argentina. *Ciência da Informação*, vol. 44 (2). <http://revista.ibict.br/ciinf/article/view/1860/2367> [Consulta: 29/12/2016]
- Fyfe, A.; McDougall-Waters, J.; Moxham, N. (2015). 350 years of scientific periodicals. *Notes and Records*, vol. 69 (3). <https://doi.org/10.1098/rsnr.2015.0036>
- Gálvez Calla, L. H. (2006). Encuentro de editores científicos. *Odontología Sanmarquina*. vol. 9 (2). [Rev. Esp. Doc. Cient., 40\(3\), julio-septiembre 2017, e182. ISSN-L: 0210-0614. doi: <http://dx.doi.org/10.3989/redc.2017.3.1419>](http://re-</p>
</div>
<div data-bbox=)

- vistasinvestigacion.unmsm.edu.pe/index.php/odont/article/view/5342/4568 [Consulta: 21/12/2016]
- Garvey, W. D.; Griffith, B. C. (1972). Communication and information processing within scientific disciplines: Empirical findings for psychology. *Information Storage and Retrieval*, vol. 8, 123-126. [https://doi.org/10.1016/0020-0271\(72\)90041-1](https://doi.org/10.1016/0020-0271(72)90041-1)
- Gonçalves, A.; Ramos, L. M.; Castro, R. (2006). Revistas científicas: características, funções e critérios de qualidade. En: Población, D.; Witter, G.; Silva, J. F. (organizadores) *Comunicação & produção científica: contexto, indicadores e avaliação*. Angellara; São Paulo.
- Guerra García, R. (1979). La investigación científica en el Perú en los años veinte. *Revista de la Universidad Católica*, vol. 5, 25-42.
- Guibovich, P. (2003). *Censura, libros e Inquisición en el Perú colonial, 1570-1754*; CSIC, Universidad de Sevilla, Diputación de Sevilla; Sevilla.
- Hernández Asensio, R. (2014). ¿Quién escribe más y sobre qué? Cambios recientes en la geopolítica de la producción científica en América Latina y el Caribe; Instituto de Estudios Peruanos; Grupo Faro; Lima. <http://repositorio.iep.org.pe/handle/IEP/20> [Consulta: 21/12/2016]
- Houghton, B. (1975). *Scientific periodicals, their historical development, characteristics and development*; Clive Bingley; London.
- Huamaní, C.; Mayta-Tristán, P. (2010). Producción científica peruana en medicina y redes de colaboración, análisis del Science Citation Index 2000-2009. *Revista Peruana de Medicina Experimental y Salud Pública*, vol. 27 (3), 315-325. <https://doi.org/10.1590/S1726-46342010000300003>
- Huamaní, C.; Pacheco-Romero, J. (2009). Visibilidad y producción de las revistas biomédicas peruanas. *Revista de Gastroenterología del Perú*, vol. 29 (2), 132-139. <http://www.scielo.org.pe/pdf/rgp/v29n2/a05v29n2.pdf> [Consulta: 21/12/2016]
- Laakso, M.; Welling, P.; Bukvova H.; Nyman, L.; Björk, B.; Hedlund, T. (2011). The Development of Open Access Journal Publishing from 1993 to 2009. *PLoS ONE*, vol. 6 (6), e20961. <https://doi.org/10.1371/journal.pone.0020961>
- Lara, J. C.; Martínez, M. (2014). *Revistas académicas chilenas: políticas editoriales y derechos de autor*; ONG Derechos Digitales; Santiago. <https://www.derechos-digitales.org/wp-content/uploads/Revistas-Academicas-Chilenas-2014-.pdf> [Consulta: 21/12/2016]
- Latindex (2016). ¿Qué es Latindex? <http://www.latindex.org/latindex/descripcion> [Consulta: 26/12/2016]
- Ley No. 30220, Ley Universitaria (2014). Diario Oficial El Peruano, Lima, 9 de julio de 2014.
- Liberatore, G.; Vuotto, A.; Fernández, G. (2013). Una cartografía de las revistas científicas argentinas en ciencias sociales y humanidades en Argentina: diagnóstico, evaluación y marcos de referencia. *Perspectivas em Gestão & Conhecimento*, vol 3 (2), 259-270. <http://periodicos.ufpb.br/ojs2/index.php/pgc/article/view/17486/10009> [Consulta: 28/12/2016]
- López Espinosa, J. A. (2000). La primera revista médica de América. *ACIMED*, vol. 8 (2), 133-139. http://www.bvs.sld.cu/revistas/aci/vol8_2_00/aci05200.htm [Consulta: 21/12/2016]
- López Ornelas, M.; Cordero Arroyo, G. (2005). Un intento por definir las características generales de las revistas académicas electrónicas. *Revista Razón y Palabra*, vol. 10 (43), 1-33. <http://www.razonypalabra.org.mx/libros/libros/caracrevelec.pdf> [Consulta: 21/12/2016]
- Miguel, S. (2011). Revistas y producción científica de América Latina y el Caribe: su visibilidad en SciELO, RedALyC y SCOPUS. *Revista Interamericana de Bibliotecología*, vol. 34 (2), 187-199. <http://eprints.rclis.org/16771/1/v34n2a6.pdf> [Consulta: 21/12/2016]
- Milla, C. (1986). *Diccionario Histórico y Biográfico del Perú*; Milla Batres; Lima.
- Morales Morante, L. F. (2016). Producción e impacto de las revistas peruanas del ámbito de las Ciencias Sociales en el catálogo Latindex. *Investigación Bibliotecológica*, vol. 30 (69), 179-204. <https://doi.org/10.1016/j.ibbai.2016.04.017>
- Pacheco, J. (2008). Anales de la Facultad de Medicina. Un recuento, 1919 a 2008, con ocasión del 90º aniversario de la Revista. *Anales de la Facultad de Medicina*, vol. 69 (4), 278-286. <https://doi.org/10.15381/anales.v69i4.1122>
- Pamo Reyna, O. G. (1997). Anales Medicales, la primera revista médica peruana. *Boletín de la Sociedad Peruana de Medicina Interna*, vol. 10 (3). http://sisbib.unmsm.edu.pe/bvrevistas/spmi/v10n3/anales_med.htm [Consulta: 21/12/2016]
- Pamo Reyna, O. G. (2005). Estado actual de las publicaciones periódicas científicas médicas del Perú. *Revista Médica Herediana*, vol. 16 (1), 65-73. <https://doi.org/10.20453/rmh.v16i1.865>
- Robles Alfaro, R.; Vela Alfaro, F.; Huapaya Huertas, O.; Chacón Torrico, H. (2016). Relación entre gasto en investigación y desarrollo con la producción científica en el Perú. *Anales de la Facultad de Medicina*, vol. 76 (4), 469-470. <https://doi.org/10.15381/anales.v76i4.11423>
- Santa, S.; Herrero Solana, V. (2010). Producción científica de América Latina y el Caribe: una aproximación a través de los datos de Scopus, 1996-2007. *Revista Interamericana de Bibliotecología*, vol. 33 (2), 379-400. <http://aprendeenlinea.udea.edu.co/revistas/index.php/RIB/article/view/7648/7075> [Consulta: 21/12/2016]
- SciELO Perú (2016). <http://www.scielo.org.pe/> [Consulta: 21/12/2016]
- Tauro del Pino, A. (1987). *Enciclopedia Ilustrada del Perú. Síntesis del conocimiento integral del Perú, desde sus orígenes hasta la actualidad*. PEISA; Lima.

Valdizán, H. (1929). *Crónicas médicas*. Talleres Gráficos del Asilo Colonia Víctor Larco Herrera; Lima.

Valenzuela, C.; Machado, L. (2015). Impacto del trabajo de Latindex en las revistas científicas uruguayas. *Ciência da Informação*, vol. 44 (2). <http://revista.ibict.br/ciinf/article/view/1799/2376> [Consulta: 28/12/2016]

ESTUDIOS / RESEARCH STUDIES

The influence of online posting dates on the bibliometric indicators of scientific articles

Mercedes Echeverría*, David Stuart**, José-Antonio Cordón-García***

*Universidad Autónoma de Madrid, España
e-mail: mercedes.echeverria@uam.es | ORCID ID: <http://orcid.org/0000-0003-4453-5122>

**University of Wolverhampton, Reino Unido
e-mail: dp_stuart@hotmail.com | ORCID ID: <http://orcid.org/0000-0002-9593-0520>

***Universidad de Salamanca, España
e-mail: jcordon@usal.es | ORCID ID: <http://orcid.org/0000-0002-8569-9417>

Received: 09-09-2016; 2nd version: 03-01-2017; Accepted: 16-01-2017

Citation/Cómo citar este artículo: Echeverría, M., Stuart, D., Cordon-García, J. A. (2017). The influence of the online posting date of scientific articles on bibliometric indicators. *Revista Española de Documentación Científica*, 40(3): e183. doi: <http://dx.doi.org/10.3989/redc.2017.3.1422>

Abstract: This article analyses the difference in timing between the online availability of articles and their corresponding print publication and how it affects two bibliometric indicators: Journal Impact Factor (JIF) and Immediacy Index. This research examined 18,526 articles, the complete collection of articles and reviews published by a set of 61 journals on Urology and Nephrology in 2013 and 2014. The findings suggest that Advance Online Publication (AOP) accelerates the citation of articles and affects the JIF and Immediacy Index values. Regarding the JIF values, the comparison between journals with or without AOP showed statistically significant differences ($P=0.001$, Mann-Whitney U test). The Spearman's correlation between the JIF and the median online-to-print publication delay was not statistically significant. As to the Immediacy Index, a significant Spearman's correlation ($r_s=0.280$, $P=0.029$) was found regarding the median online-to-print publication delays for journals published in 2014, although no statistically significant correlation was found for those published in 2013. Most journals examined ($n=52$ out of 61) published their articles in AOP. The analysis also showed different publisher practices: eight journals did not include the online posting dates in the full-text and nine journals published articles showing two different online posting dates--the date provided on the journal website and another provided by Elsevier's Science Direct. These practices suggest the need for transparency and standardization of the AOP dates of scientific articles for calculating bibliometric indicators for journals.

Keywords: Advance online publication; immediacy index; journal impact factor; online posting date; publication delays, urology and nephrology journals.

Influencia de la fecha de publicación online de los artículos científicos en los indicadores bibliométricos

Resumen: Este artículo analiza la distancia temporal entre la disponibilidad online de los artículos y su publicación en papel y cómo este lapso de tiempo afecta a los indicadores bibliométricos Factor de Impacto (JIF) e Índice de Inmediatez. Esta investigación examinó 18.526 artículos, la colección completa de artículos y revisiones publicadas por un conjunto de 61 revistas de Urología y Nefrología en 2013 y 2014. Los resultados hallados sugieren que la Publicación Electrónica Anticipada (AOP) acelera la citación de artículos, afectando a los valores de JIF e Índice de Inmediatez. Respecto a los valores de JIF, la comparación entre revistas con AOP y sin esta característica mostró diferencias estadísticamente significativas ($P=0,001$, U test Mann-Whitney). La correlación de Spearman entre JIF y la mediana del intervalo de tiempo publicación online-papel no resultó ser estadísticamente significativa. En cuanto al Índice de Inmediatez, se halló una correlación de Spearman significativa respecto a la mediana del intervalo de publicación online-papel ($r_s=0,280$, $P=0,029$) para las revistas publicadas en 2014, pero sin correlación significativa para las publicadas en 2013. La mayoría de las revistas examinadas ($n= 52$ de 61) publicaron sus artículos en AOP. El análisis mostró también diferentes prácticas de los editores, ocho revistas no indicaron en el texto completo las fechas de publicación online, y nueve revistas publicaron los mismos artículos con dos fechas diferentes de publicación online, la fecha proporcionada por la web de la revista y la fecha suministrada por Science Direct de Elsevier. Estas prácticas sugieren la necesidad de transparencia y estandarización de las fechas de publicación online de artículos científicos para el cálculo de los indicadores bibliométricos de las revistas.

Palabras clave: Publicación electrónica anticipada; índice de inmediatez; factor de impacto; fecha de publicación online; tiempos de demora en la publicación; revistas de urología y nefrología.

Copyright: © 2017 CSIC. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY) Spain 3.0.

1. INTRODUCTION

The current system for measuring the impact of scholarly journals is based on the quantitative analysis of citations received by scientific publications. Time is an important factor in the calculation of such indicators, and the most popular indicators take into consideration the time a journal has had to accumulate citations, measuring both the citations a journal has received and the time it has taken to accumulate them. However, online publishing is changing the nature of the publication date, with inevitable consequences for the calculating of bibliometric indicators and the ranking of journals.

The most influential indicator to evaluate journals is the Journal Impact Factor (JIF), devised by Garfield and Sher in the early 1960s (Garfield, 2006). A journal's impact factor provides information on the relative importance of a journal in terms of influence, prestige and visibility within its subject category. The JIF is calculated from three elements, citations received, number of substantial items published and the measurement of a two-year window after publication. Despite its apparent accuracy, reproducibility and widespread use, the calculation of these variables as proxy for measuring a journal's influence has been widely discussed (Bordons et al., 2002; Bornmann and Marx, 2016; Glänzel and Moed, 2002; Lozano et al., 2012). The citation window has been criticised for a number of reasons: the shortness of the two-year period for measuring disciplines with different citation latencies, it discriminates against papers whose citation patterns present delayed rises or delayed recognitions (Glänzel and Moed, 2002; Wang, 2013), and the time aggregation for online journal publication escapes from traditional metrics of print journal publication (Haustein et al., 2015; Stuart, 2015; Tort et al., 2012).

The Immediacy Index is another citation metric used for journal evaluation, designed to indicate how quickly articles in a journal are cited. This indicator is calculated as the average number of citations received by articles of a journal in the publication year divided by the number of articles published in that year (Thomson Reuters, 2012). In comparison to JIF, which is based on citations received by articles published in the two years before the year in question, Immediacy Index is more restrictive in its calculation, considering only citations made in the actual publication year.

This work reports the findings of an investigation into the effect of *online-to-print* delays on the JIF and Immediacy Index of journals categorised as *Urology and Nephrology* in the Journal Citation

Reports (JCR). The delay is calculated, from the initial date of online posting provided by the publishers in the full-text of the articles to the final official print date, indexed by Thomson Reuters' Web of Science (WoS). Subsequently, because JIFs form the basis of JCR rankings, we compare the *online-to-print* publication delays across the four quartiles of journals in JCR. Furthermore, the correlation between *online-to-print* delays and the number of articles published by a journal per year is studied, as well as the differences in publication delays between journals published by commercial publishers and scientific societies.

2. BACKGROUND

The traditional publication of print journals usually involved long delays from a manuscript's reception to publication. This was due to the bundling of articles into issues, as well as the backlog of manuscripts waiting for placement in a bound print volume with a limited space per year. In the late 1990s the proposal of many publishers to reduce these delays and give visibility to scientific output was to host electronic versions of their paper-based journals on their websites (Dong et al., 2006). Currently, most of major scientific publishers include posting dates in the full-text of peer-reviewed articles, once they are completely edited and formatted, and before the release print issue. For immediate online publication editors use services so-called: '*Advance online publication*' by Nature Group Publishing, '*Advance access*' by Oxford Journals, '*Articles in press*' by Elsevier, '*Early view*' by John Wiley & Sons and '*Online first*' by Springer.

AOP enables the dissemination of knowledge more quickly, reduces the publication print delays and increases the visibility and readership of articles. It might also have influence on bibliometric indicators by enlarging citation time before the print publication date, which is the official start of the citation window used by WoS for calculating the JIF and Immediacy Index.

The increasing *online-to-print* publication delays in recent years, and its influence on bibliometric indicators have been pointed out in a number of recent studies. Tort et al. (2012) showed that the delays between online availability of articles and print publication lead to earlier citations, and thus can artificially raise a journal's impact factor, calculated by WoS on 2-year citation window. Chen et al. (2013) analysed the publication times in 51 *Ophthalmology* journals finding that journals with AOP had a higher JIF compared to those without this feature, although this difference was not statistically significant. Heneberg (2013) reported

that online first publication severely affects the calculation of the JIF and Immediacy Index and disadvantages online-only and print-only journals when evaluating them according to these bibliometric indicators.

Others studies have considered the issue of publication delays, Alves-Silva et al. (2016) found that 'online posting', defined as the period spent between the submission of a manuscript and online publication, was significantly and negatively related to JIF in a set of *Ecology* journals. Amat (2008) defined 'publication delay' as the chronological distance between the stated date of reception of a manuscript by a given journal and its appearance in any print issue of that journal, and determined that the effect of online posting on the publisher's site of the whole publication process reduced the publication delay by 29% in a group of *Food Research* journals. In contrast to the assumption that OA journals have shorter publication delays, Dong et al. (2006) showed that the 'publication speed' of journals, defined as the interval of time from the reception of the manuscripts to online publication, was faster for manuscripts published in AOP by Nature Group than those published in OA by Biomed Central. Although OA articles had a greater impact than articles that were not accessible freely online, they estimated that the impact of AOP was still uncertain.

3. RESEARCH QUESTIONS

There is a limited amount of research so far into *online-to-print* delays, carried out on different publication times, research practices and different fields, and thus there is a clear need for further studies in more areas. This study contributes to current understanding of the influence of *online-to-print* delays on the JIF and Immediacy Index, through an analysis of a set of 61 journals of *Urology and Nephrology*, indexed in JCR Thomson Reuters. More specifically, the research questions in this paper are:

- What is the correlation between *online-to-print* delays with JIF and Immediacy Index in *Urology and Nephrology* journals?
- Are there any differences between the distribution of journals in quartiles (JCR) and *online-to-print* delays?
- Is there a relationship between *online-to-print* publication delays and the number of articles published per year by a journal?
- Is there any difference in publication delays between journals published by large commercial publishers and scientific societies?

4. METHODOLOGY

To gather research data (see online additional material [2013](#) and [2014](#) for details) a set of 61 journals out of 77 indexed by JCR (2015), whose subject category was *Urology and Nephrology* were analysed. In order to construct the dataset of journals (see Appendix Table I, and online additional material [2013](#) and [2014](#)), the following methodological criteria were established:

4.1. Criteria for inclusion

- All journals indexed by JCR in 2015 under the category *Urology and Nephrology*.
- All original articles and reviews, whose print publication date was included between January 2013 and December 2014.

4.2. Criteria for exclusion

- Electronic journals without print publication version.
- Journals whose articles do not provide information about the online posting dates, JIF or Immediacy Index. Furthermore, those journals that could not be accessed full-text and journals with a low number of articles published were also excluded.
- Book reviews, critiques, comments, and letters from or to editors.
- Journal supplements were not included (following Björk and Solomon, 2013; Chen et al., 2013), as well as monographic series.

All papers were examined in electronic form. When an article was posted with different online dates, the date provided on the journal's website and the date provided by *Elsevier's Science Direct*, the date selected was the first date in which the article was available online. For each article the online posting date, provided by publishers in the full-text online, and the official print date provided by WoS, were recorded (see online additional material [2013](#) and [2014](#)). When a journal did not specify information about the online posting date it was necessary to contact publishers. This only applied to two journals (*Archivos Españoles de Urología* and *Nephrology Nursing Journal*) and they confirmed that electronic and print publication dates were the same (see Appendix, Table I).

4.3. Journals and papers examined

A total of 18,526 articles and reviews published in 2013 and 2014 were examined. This number of articles and reviews corresponded faithfully to

61 out of 77 journals used by JCR to calculate the JIF score of 2015 and Immediacy Index of 2013 and 2014 of the subject category *Urology and Nephrology* Journals. The analysis excluded 16 journals, due to criteria detailed in the list of journals (see Appendix, Table II). This fact slightly altered the number of journals included in the list of 2013 (59 journals) (see Appendix, Table I).

Since only a minority of articles provided the day and month of print publication, the journal issue date was converted to days, based on average unit length. In this way, the date of publication was estimated: 1) as the 15th day of each month for monthly journals, and 2) the last day of the first month for bimonthly journals. In cases of irregular journals (i.e. with 8 or 10 issues per year) the date of publication was distributed among the yearly period (following Amat, 2008). When an article was published in print and online during the same period of time, the number of days was not considered (see online additional material 2013 and 2014).

4.4. STATISTICAL ANALYSES

Median and Interquartile Ranges (IQRs, 25% - 75%) were calculated with the variable *online-to-print* publication dates, obtained from the time distance between AOP and print publication for all articles and reviews of each journal. All data were tested for normality distribution and because the measures of the variable *online-to-print* publication dates were not normally distributed non-parametric tests were used.

The Spearman correlation test was used to examine the correlation between the median *online-to-print* publication delays and JIF, Immediacy Index and the size of journals, based on the number of articles published per year. The Mann-Whitney U test was used to compare the JIFs of journals with AOP or without AOP, and also to determine whether there were differences between journals published by commercial editors and scientific societies regarding the variable *online-to-print* publication delays. The Kruskal-Wallis test was used for intra-group comparison of distribution of journals in quartiles with the variable *online-to-print* publication.

In our tests the statistical significance was set at 5% and two-sided. The analyses were performed using the SPSS Statistics software (IBM), version 22.

5. RESULTS

The analysis revealed that a high proportion of *Urology and Nephrology* journals published their articles ahead of print 85.2% (n=52 out of 61),

with a great dispersion in the lapse of time *online-to-print* publication.

The median *online-to-print* publications was 86 days (IQR, 48-171) for journals published in 2013, and 95 days (IQR, 36-171) for journals published in 2014. The chronological distance between *online-to-print* publication dates showed a wide variability. In 2013, the range varied from journals with no-delays 16.9% (n=10) to journals whose median was more than 300 days 5.1% (n=3), with the most frequent interval being 43-84 days 23.7% (n=14) journals (see Fig. 1).

The chronological distance between *online-to-print* publications in 2014 varied from journals with no delays 14.8% (n=9) to 2 (3.2%) journals whose median was more than 300 days, with the most frequent interval being 169-210 days 18% (n=11) (see Fig. 2).

Regarding the study of the JIF (2015), the comparison between the JIF of journals with AOP and journals without AOP showed statistically significant differences (Mann-Whitney $z=-3.255$, $P=0.001$). The mean of the JIF for journals with AOP was $\bar{x}=2.953$, and for journals without AOP $\bar{x}=1.003$.

With respect to the correlation between JIF and the median days *online-to-print* publication delays, whose values for 2013 and 2014 were grouped, there was no statistically significant correlation ($r_s=0.158$, $P=0.233$).

In regard to Immediacy Index, the correlation between median days *online-to-print* delays and Immediacy Index was statistically significant ($r_s=0.280$, $P=0.029$) for journals published in 2014 (n=61) (see Fig. 3). However, for journals published in 2013 (n=56 with Immediacy Index), there was no significant Spearman correlation ($r_s=0.117$, $P=0.391$) (see Fig. 4).

In answer to the second research question, about whether there were differences of rank position of journals in quartiles (JCR) regarding *online-to-print* delays, the Kruskal-Wallis test did not identify statistically significant differences in 2013 ($X^2=1.814$, $DF=3$, $P=0.612$) and 2014 ($X^2=2.360$, $DF=3$, $P=0.501$). However, it is important to observe that journals without AOP were classified in the lower quartiles: 5 titles (Q4) and 4 titles (Q3), whereas 16 journals with 267 days of AOP, on average, were classified in the first quartile (Q1).

Regarding the relationship between *online-to-print* publication delays and the size of journals, in terms of the number of articles published per year by a journal, a significant Spearman correlation was observed, both for journals published in 2013 ($r_s=0.335$, $P=0.010$) and 2014 ($r_s=0.375$, $P=0.003$).

Figure 1. Frequency distribution of the median of days online-to-print publication delays in 2013

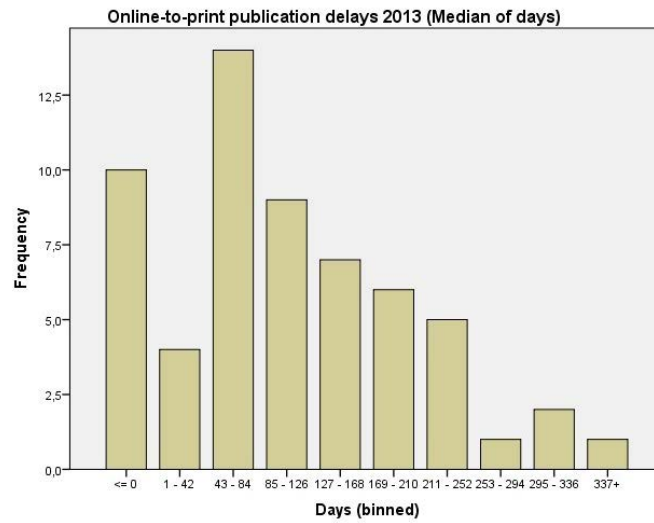


Figure 2. Frequency distribution of the median of days online-to-print publication delays in 2014

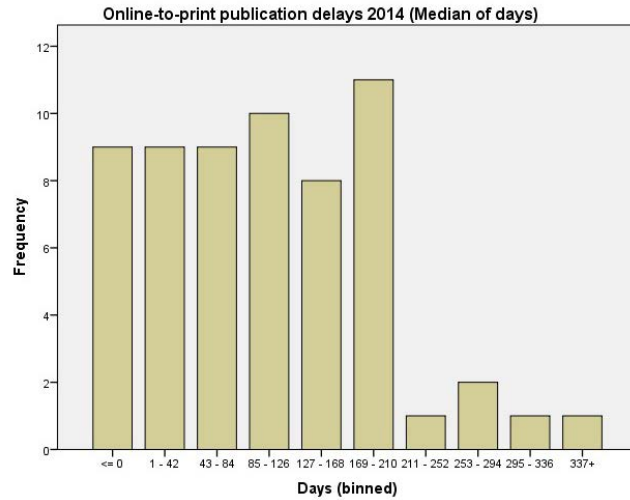


Figure 3. Online-to-print publication delays (median of days) and Immediacy Index (2013)

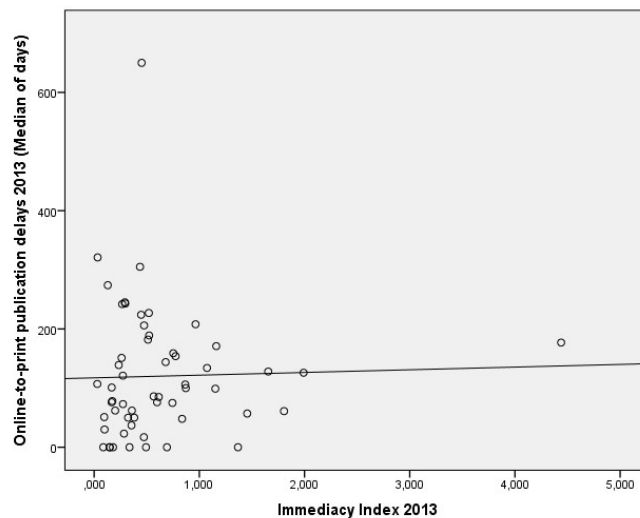
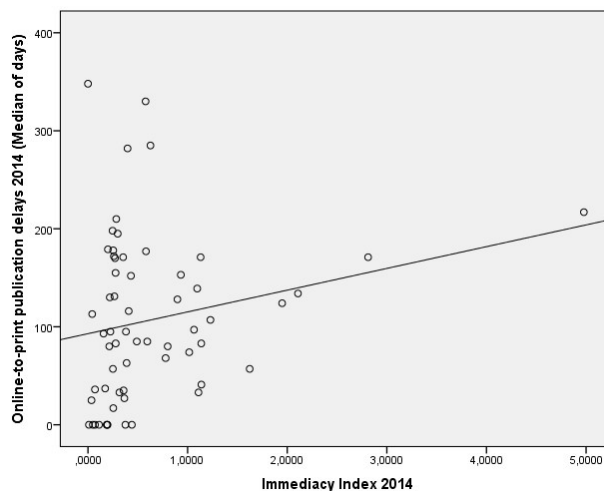


Figure 4. Online-to-print publication delays (median of days) and Immediacy Index (2014)

Another issue emerging from this research was whether there were differences regarding publication delays between journals published by commercial publishers (53/61) and by scientific societies (8/61) in relation to the variable of *online-to-print* publication delays. The Mann-Whitney U test identified statistically significant differences for journals published by commercial publishers and scientific societies in 2013 ($z=-2.109$, $P=0.035$, median *online-to-print* delays of publishers 101 days [IQR, 51-173] vs societies 31 days [IQR, 0-81]). For journals published in 2014 ($z= -2.119$, $P=0.034$, median *online-to-print* delays of publishers 97 days [IQR, 45-171] vs societies 34 days [IQR, 0-74]).

6. DISCUSSION

The analysis revealed that most publishers of *Urology* and *Nephrology* journals published articles ahead of print ($n= 52$ out of 61, 85.2%), with a great dispersion in the lapse time *online-to-print* publications (see Fig. 1 and Fig. 2). The period spent between online to print publication was almost 3 months (86 days), on average, for journals published in 2013 and more than 3 months (95 days), on average, for journals published in 2014.

Although most publishers included the online posting date in the full-text of articles, some exceptions were observed: 8 journals did not include the date of online posting, and 9 journals published the same articles with 2 different online posting dates, the date provided on the journal's website and the date provided by Elsevier' *Science Direct*. These practices reflected a lack of standardization in terms of how publishers report online dates.

Despite the role of time on the accumulation of citations, the analysis did not show significant correlation between JIF and *online-to-print* publication delays. This is probably due to the value of the JIF being determined by different parameters such as the journal size, the subject area, prestige, approach, availability or influential authors, and cannot objectively be reduced to a few variables (Alves-Silva et al., 2016). Although the extent to which online-to-print delays can affect a journal's JIF is difficult to determine, journals without AOP would have less opportunity to receive citations and consequently it may influence their JIF.

In order to contrast this result to other studies (Chen et al., 2013; Heneberg, 2013; Tort et al. 2012), these authors did not register statistically significant correlation between journals with AOP and JIF, although they mentioned the influence of AOP in terms of higher impact factors in comparison to journals without this feature, number of citations utilized for the calculation of JIF, and potential distortions concerning the evaluation of journals. Moed (2007) provided evidence that the inclusion of a paper in *arXiv* accelerated its citation probably due to this repository making papers available *earlier* accelerating communication, the process of reading and processing of information. This evidence was observed also by Craig et al. (2007) [mentioned by Heneberg (2013)] "preprint availability causes the citation counting process to start earlier although this earlier citation did not affect the final magnitude of citations accrued to a journal article". In this regard, Vanclay (2012) pointed out that "[AOP] offers little advantage for a journal that peaks early (e.g. *Nature*), but may substantially alter the JIF of slow-to-peak journals, such as *Ecology*".

Regarding the Immediacy Index, indicator that provides a useful perspective of the cutting-edge fields of research, the analysis showed a statistically significant correlation with *online-to-print* delays for the journals of 2014 (n=61), although without significant differences for the journals of 2013 (n=57 journals with Immediacy Index). Heneberg (2013), using the database Scopus, studied the number of citations received by articles with AOP ("in press" [Elsevier]) within the year when the article was assigned to a journal issue. According to Heneberg (2013), AOP ("in press" [Elsevier]) severely affects the calculation of Immediacy Index, although without significant differences.

As Immediacy Index is calculated on the number of citations that an article receives in the same year it is published, an article published early or with AOP has a better chance of being cited than one published later in the year. In fact, as Thomson Reuters (2012) states "Many publications that publish infrequently or late in the year have low Immediacy Indexes". On this point, the potential of technical conditions, such as frequency of publication and publication delays impact the Immediacy Index and have long been recognized and have limited the importance of this Index (Glänzel and Moed, 2002).

In answer to the second research question, about whether there were differences in the distribution of journals in quartiles (JCR) in relation to *online-to-print* delays, the Kruskal-Wallis test did not show any statistically significant differences, although it is observed that a higher median of AOP corresponds with the higher quartiles, and journals without AOP are located in the lower quartiles. However, this statement would need further analysis.

Regarding the correlation between *online-to-print* delays and the number of articles published per year by a journal, the results indicated a statistically significant association. The steady growth of the number of articles, the large backlog of articles awaiting publication in bound print volumes and the need to achieve a rapid and efficient scientific communication are in the origin of Advance Online Publication. In this context, the data reflected statistically significant differences between journals published by major commercial editors (Elsevier, John Wiley, Nature Publishing Group, Oxford, Springer, Taylor & Francis) with a large number of articles per year and long delays

versus journals published by scientific societies, with a smaller number of articles published and shorter delays.

It is important to note that commercial editors published 87% of *Urology and Nephrology* journals analysed. As Larivière et al. (2015) pointed out "The proportion of scientific output published in journals under their ownership [large commercial publishing houses] has risen steadily over the past 40 years, and even more so since the advent of the digital era".

7. CONCLUSION

The practice of making papers available online ahead of print publication has facilitated the rapid communication of scientific research, but at the same time it has resulted in bibliometric consequences, such as enlarging the citation time of JIF and Immediacy Index.

The research revealed that the longest periods of *online-to-print* publication are mainly produced by large commercial editors that publish journals with a large number of articles per year. This suggests that, to some extent, delays are justified or at least are convenient. However, it would be necessary to study the effects long terms that this gradual increase in the number of days of AOP is having on the number of citations and their contribution to the JIF values, on the differences created between journals with or without AOP, and possible consequences on other bibliometric indicators.

Another feature of some articles with AOP was the lack of standardization of online posting dates. The publishers reported journals with different posting dates for the same article and journals without any mention of the date of online posting. Thus, there is a need for transparency and standardization of posting dates in order to assure the calculation of bibliometric indicators for the evaluation and comparability of journals (Haustein et al., 2015).

Until such standards emerge it is important to have a better understanding of how AOP affects different fields and metrics, and whether there are areas where the *online-to-print* publication delay is sufficient to not only have an impact on lesser-used indicators such as the Immediacy Index but also the widely-adopted JIF.

8. REFERENCES

- Alves-Silva, E.; Porto, A.C.F.; Firmino, C.; Silva, H.V.; Becker, I.; Resende, L.; Borges, L.; Pfeffer, L.; Silvano, M.; Galdiano, M.S.; Silvestrini, R.; Moura, R. (2016). Are the impact factor and other variables related to publishing time in ecology journals? *Scientometrics*, vol. 108 (3), 1445-1453. <https://doi.org/10.1007/s11192-016-2040-0>
- Amat, C.B. (2008). Editorial and publication delay of papers submitted to 14 selected Food Research journals. Influence of online posting. *Scientometrics*, vol. 74 (3), p. 379-389. <https://doi.org/10.1007/s11192-007-1823-8>
- Björk, B.C.; Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. *Journal of Informetrics*, vol. 7 (4), 914-923. <https://doi.org/10.1016/j.joi.2013.09.001>
- Bordons, M.; Fernández M.T.; Gómez, I. (2002). Advantages and limitations in the use of impact factor measures for the assessment of research performance in a peripheral country. *Scientometrics*, vol. 53 (2), 195-206. <https://doi.org/10.1023/A:1014800407876>
- Bornmann, L.; Marx, W. (2016). The journal Impact Factor and alternative metrics. *EMBO reports*, vol. 17 (8), 1094-1097. <https://doi.org/10.15252/embr.201642823>
- Chen, H.; Chen, H.C.; Jhanji, V. (2013). Publication times, impact factors, and advance online publication in Ophthalmology journals. *Ophthalmology*, vol. 120 (8), 1697-1701. <https://doi.org/10.1016/j.ophtha.2013.01.044>
- Craig, I.D.; Plume, A.M.; McVeigh, M.E.; Pringle, J.; Amin, M. (2007). Do Open Access Articles Have Greater Citation Impact? A critical review of the literature. *Journal of Informetrics*, vol. 1 (3), 239-248. <https://doi.org/10.1016/j.joi.2007.04.001>
- Dong, P.; Loh, M.; Mondry, A. (2006). Publication lag in biomedical journals varies due to the periodical's publishing models. *Scientometrics*, vol. 69 (2), 271-286. <https://doi.org/10.1007/s11192-006-0148-3>
- Garfield, E. (2006). The history and meaning of the journal impact factor. *Jama*, vol. 295 (1), 90-93. <https://doi.org/10.1001/jama.295.1.90>
- Glänzel, W.; Moed, H.F. (2002). Journal impact measures in bibliometric research. *Scientometrics*, vol. 53 (2), 171-193. <https://doi.org/10.1023/A:1014848323806>
- Haustein, S.; Bowman, T.D.; Costas, R. (2015). When is an article actually published? An analysis of online availability, publication, and indexation dates. *15th International Conference on scientometrics and informetrics (ISSI), 29 June –3 September 2015*, pp. 1170-1179. Istanbul, Turkey. <https://arxiv.org/abs/1505.00796>
- Heneberg, P. (2013). Effects of print publication lag in dual format journals on scientometric indicators. *PLoS One*, vol. 8, (4), e59877. <https://doi.org/10.1371/journal.pone.0059877>
- Larivière, V.; Haustein, S.; Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLoS One*, vol. 10 (6), e0127502. <https://doi.org/10.1371/journal.pone.0127502>
- Lozano, G. A.; Larivière, V.; Gingras, Y. (2012). The weakening relationship between the impact factor and papers' citations in the digital age. *Journal of the American Society for Information Science and Technology*, vol. 63 (11), 2140-2145. <https://doi.org/10.1002/asi.22731>
- Moed, H.F. (2007). The effect of "open access" on citation impact: An analysis of ArXiv's condensed matter section. *Journal of the American Society for Information Science and Technology*, vol. 58 (13), 2047-2054. <https://doi.org/10.1002/asi.20663>
- Stuart, D. (2015). Metrics for an increasingly complicated information ecosystem. *Online Information Review*, vol. 39 (6), 848-854. <https://doi.org/10.1108/OIR-06-2015-0174>
- Thomson Reuters (2012). *Journal Citation Reports*; Available at: http://admin-apps.webofknowledge.com/JCR/help/h_jrnlnfo.htm. (Accessed 4 September 2016).
- Tort, A.B.L.; Targino, Z.H.; Amaral, O.B. (2012). Rising publication delays inflate journal impact factors. *PLoS One*, vol. 7 (12) e53374. <https://doi.org/10.1371/journal.pone.0053374>
- Vanclay, J. K. (2012). Impact factor: outdated artefact or stepping-stone to journal certification? *Scientometrics*, vol. 92 (2), 211-238. <https://doi.org/10.1007/s11192-011-0561-0>
- Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, vol. 94 (3), 851-872. <https://doi.org/10.1007/s11192-012-0775-9>

APPENDIX

Table I. Dataset of Urology and Nephrology Journals

Title of journals	2013				2014				JIF	Quartiles	Publishers commercial societies (1)	Scientific societies (2)
	Median days	Interquartile range	Immediacy Index	N articles	Median days	Interquartile range	Immediacy Index	N articles				
	Actas Urológicas Españolas	151	95-182	0,261	92	179	153-213	0,200				
Advance in Chronic Kidney Disease	151	95-182	0,261	92	179	153-213	0,200	99	0,964	Q4	1	
Aging Male	50	0-60	0,323	31	80	57-128	0,214	42	1,493	Q3	1	
American Journal of Kidney Diseases	128	100-157	1,655	219	124	100-152	1,949	201	6,369	Q1	1	
American Journal of Nephrology	17	0-24	0,473	129	33	24-46	0,315	128	2,605	Q2	1	
American Journal of Physiology-Renal Physiology	75	65-85	0,744	330	68	57-77	0,779	289	3,39	Q1	2	
Archivos Españoles de Urología	0	0	No Immediacy Index	118	0	0,009	0,009	113	0,307	Q4	1	
Asian Journal of Andrology	76	59-102	0,598	120	97	61-141	1,064	125	2,644	Q2	1	
BJU International	134	64-193	1,073	424	107	27-189	1,229	305	4,387	Q1	1	
Blood Purification	0	0	0,336	107	0	0	0,187	90	1,404	Q3	1	
Cardiorenal Medicine	0	0	0,179	28	0	0	0,194	31	1,698	Q3	1	
Clinical And Experimental Nephrology	121	183-258	0,273	213	282	181-301	0,397	114	1,945	Q2	1	
Clinical Genitourinary Cancer	154	118-167	0,774	91	172	135-199	0,26	96	2,599	Q4	1	
Clinical Journal of the American Society of Nephrology	99	77-115	1,152	252	74	47-85	1,016	242	4,657	Q1	2	
Clinical Nephrology	139	93-414	0,234	141	93	47-546	0,156	102	1,065	Q4	1	
CUAJ - Canadian Urological Association Journal	0	0	1,367	231	0	0	0,194	218	0,866	Q4	2	
Current Urology Reports	63	50-82	No Immediacy Index	92	57	34-62	0,25	84	1,597	Q3	1	
European Urology	177	143-223	4,439	189	217	148-327	4,978	228	14,976	Q1	1	
Hemodialysis International	245	204-281	0,292	87	170	125-199	0,271	107	1,495	Q3	1	
International Journal of Impotence Research	189	140-212	0,523	43	195	174-214	0,298	46	1,396	Q3	1	

International Journal of Urology	208	158-236	0,964	162	153	130-187	0,932	221	1,878	Q2	1
International Neurourology Journal	0	0	No Immediacy Index	32	0	0	0,375	32	1,344	Q3	2
International Urogynecology Journal	224	187-245	0,447	286	171	150-193	0,353	244	1,834	Q2	1
International Urology and Nephrology	242	81-328	0,267	236	155	108-184	0,277	336	1,292	Q3	1
Iranian Journal of Kidney Diseases	0	0	0,493	75	0	0	0,44	75	1,14	Q4	2
Journal of Endourology	73	41-103	0,274	237	95	78-106	0,381	246	2,107	Q2	1
Journal of Nephrology	Articles without online posting in 2013				116	53-194	0,409	88	1,352	Q3	1
Journal of Pediatric Urology	274	181-391	0,130	254	178	139-2015	0,253	217	1,17	Q4	1
Journal of Renal Nutrition	206	76-297	0,475	80	83	68-92	0,278	53	2,06	Q2	1
Journal of Sexual Medicine	85	73-125	0,614	324	85	75-103	0,595	315	2,844	Q2	1
Journal of the American Society of Nephrology	61	40-87	1,805	190	134	114-172	2,108	260	8,491	Q1	2
Journal of Urology	171	146-187	1,161	606	171	156-186	1,131	458	4,7	Q1	1
Kidney and Blood Pressure Research	0	0	0,087	92	17	0-32	0,254	71	2,908	Q2	1
Kidney International	126	96-153	1,991	232	171	143-194	2,812	242	7,683	Q1	1
LUTS - Lower Urinary Tract Symptoms	321	282-361	0,033	30	348	298-377	0	35	0,231	Q4	1
Minerva Urologica e Nefrologica	0	0	0,148	29	0	0	0,048	27	0,536	Q4	1
Nature Reviews Nephrology	57	47-62	1,455	55	57	47-62	1,623	53	9,463	Q1	1
Nature Reviews Urology	48	31-64	0,836	61	33	27-42	1,109	55	5,957	Q1	1
Nephrologie & Therapeutique	107	8-153	0,03	64	36	0-54	0,07	51	0,541	Q4	1
Nephrology	23	19-26	0,284	109	27	24-29	0,365	126	1,796	Q3	1
Nephrology Dialysis Transplantation	106	74-157	0,866	389	139	94-187	1,096	332	4,085	Q1	1
Nephrology Nursing Journal	0	0	0,148	34	0	0	0,111	45	0,734	Q4	2
Nephron Clinical Practice	51	0-72	0,096	94	37	0-53	0,172	85	1,471	Q3	1
Nephron Experimental Nephrology	Only 7 articles in 2013				0	0-39	0,071	42	1,531	Q3	1
Neurology Urodynamics	227	207-267	0,52	97	330	188-382	0,578	127	3,128	Q1	1
Pediatric Nephrology	144	122-214	0,68	225	177	148-270	0,582	250	2,338	Q2	1
Peritoneal Dialysis International	182	150-210	0,512	85	152	91-232	0,431	152	1,298	Q3	2
Progress en Urologie	101	47-137	0,167	228	113	71-172	0,042	212	0,562	Q4	1
Prostate	159	108-199	0,752	182	83	67-119	1,137	159	3,778	Q1	1

Prostate Cancer And Prostatic Diseases	100	73-131	0,871	62	80	53-102	0,8	55	3,803	Q1	1
Scandinavian Journal of Urology	243	189-274	0,296	81	198	172-222	0,247	81	1,346	Q3	1
Seminars in Dialysis	86	44-190	0,566	103	131	77-156	0,265	113	1,912	Q2	1
Seminars in Nephrology	0	-37 - 0	0,692	52	41	0-47	1,138	58	3,773	Q1	1
Therapeutic Apheresis and Dialysis	50	0-137	0,381	84	210	0-248	0,284	88	1,477	Q3	1
Urolithiasis	62	40-95	0,200	70	95	69-120	0,225	71	1,454	Q3	1
Urolog	30	0-63	0,099	163	25	0-36	0,035	172	0,331	Q4	1
Urologia Internationalis	76	41-115	0,166	169	130	81-178	0,219	160	1,313	Q3	1
Urologic Clinics of North America	78	64-95	0,173	52	85	66-123	0,49	51	1,835	Q2	1
Urologic Oncology- Seminars and Original Investigations	650	563-744	0,452	250	128	90-262	0,898	283	2,921	Q1	1
Urology	62	39-75	0,359	524	63	22-78	0,388	545	2,187	Q2	1
World Journal of Urology	305	197-435	0,436	236	285	226-312	0,627	212	2,397	Q2	1

Table II. Journals not included in the analysis

Journals not included	Reason for non-inclusion
Aktuelle Urology	No access available
BMC Nephrology	Online only
BMC Urology	Online only
Canadian Journal of Urology	Articles without date of online posting
Contributions to Nephrology	Monographic series (UlrichWeb)
Current Opinion in Nephrology and Hypertension.	Articles without date of online posting
Current Opinion in Urology.	Articles without date of online posting
European Urology Supplements	Supplements
International Brazilian Journal of Urology	Articles without date of online posting
Kidney International. Supplement.	Supplements
Nefrologia	Articles without date of online posting
Nephron	JIF not available
Nephron Physiology.	Only 11 articles published
Renal Failure	Only electronic since 2011
Revista de Nefrologia Dialisis y Trasplante	JIF and Immediacy Index not available
Urology Journal	Articles without date of online posting

ESTUDIOS / RESEARCH STUDIES

Modelos evaluativos de Metaliteracy y alfabetización en información como factores de excelencia académica

Miguel Ángel Marzal*, Jussara Borges**

*Universidad Carlos III de Madrid.

Correo-e: mmarzal@bib.uc3m.es | ORCID iD: <http://orcid.org/0000-0003-2039-234X>

**Universidade Federal da Bahia

Correo-e: jussarab@ufba.br | ORCID iD: <http://orcid.org/0000-0003-0157-8691>

Recibido: 01-06-2016; 2ª versión: 15-12-2016; Aceptado: 16-01-2017.

Cómo citar este artículo/Citation: Marzal, M. A., Borges, J. (2017). Modelos evaluativos de Metaliteracy y alfabetización en información como factores de excelencia académica. *Revista Española de Documentación Científica*, 40(3): e184. doi: <http://dx.doi.org/10.3989/redc.2017.3.1410>

Resumen: El trabajo analiza la transformación de la alfabetización en información a partir de la confluencia de las competencias en información con las competencias en comunicación. Esta confluencia se hace necesaria por el nuevo contexto que en Educación impone la eScience, la Web Social, el aprendizaje colaborativo y la imagen para la formación académica. La alfabetización en información debe coordinarse con la visual literacy, la dataliteracy, la transliteracy, la new media literacy, lo que hace posible el impacto de la metaliteracy en Educación. El trabajo estudia la definición de la metaliteracy en Educación Superior para demostrar la necesidad del diseño de un modelo de evaluación para unas competencias en metaliteracy, por lo que detalla el método de búsqueda de referentes y llegar a la presentación de una estructura de categorías que organicen la propuesta de indicadores apropiados, presentando la propuesta de un modelo de evaluación de competencias para metaliteracy.

Palabras clave: Metaliteracy; conocimiento de la información; alfabetización académica; modelo de indicadores; evaluación de competencias.

Evaluative models of Metaliteracy and Information Literacy as factors in academic excellence

Abstract: The paper analyzes the transformation of information literacy from the point of convergence of information competencies and communication competencies. This junction is needed due to the new context promoted by eScience, Social Web, collaborative learning, and the use of images in Education. Information literacy should be coordinated with visual literacy, data literacy, transliteracy and new media literacy, enabling the impact of metaliteracy in Education. The paper studies the definition of metaliteracy in Higher Education to demonstrate the need for designing an competence evaluation model in metaliteracy. The paper details the method of searching for references and presents a structure of categories for organizing a proposal for appropriate indicators, with a presentation of a proposal for a metaliteracy competency assessment model.

Keywords: Metaliteracy; information literacy; academic literacy; model of indicators; competencies evaluation.

Copyright: © 2017 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 evolución conceptual de la alfabetización en información ha transitado desde un conjunto de competencias individuales (buscar, evaluar, seleccionar, aplicar la información) hacia un sentido más social, que permita incluir competencias para la edición colaborativa y el aprendizaje en comunidad. La creciente disponibilidad de la información digital en red, unido al acceso a herramientas de creación y compartición de contenidos, hace posible transformar al usuario de la información en potencial creador de conocimientos (Tyner, 2008).

Esta potencialidad tecnológica se une a la mentalidad "postmoderna", en el marco de la Sociedad de la información y del conocimiento, que impulsa valores participativos. La "sociabilidad postmoderna" valora la compartición de contenidos, la conectividad constante entre las personas y la participación en procesos sociales de su interés. Sin duda cambia la noción "creación" en la edición, identificada como un trabajo en solitario mediante "técnicas narrativas lineales", ya que en el "postmodernismo" esta noción se entiende más como una multiplicidad de ideas, un *continuun narrativo* multiseccional, interactivo y dinámico, a través de estilos y formatos en ambientes distribuidos.

Para González Fernández-Villavicencio (2012) la sociedad actual posee un componente tecnológico fuerte, pero sobre todo exige formas diferentes de participar, comunicar y colaborar en red, características y propiedades que modifican sustancialmente las llamadas *competencias en información*. Parece necesario entonces incluir unas capacidades más relacionadas con aspectos comunicativos, sociales y participativos, destinadas a un aprendizaje social. Es aquí donde cobra todo su sentido la *competencia en comunicación*, en tanto que capacidad para generar y gestionar unas relaciones constructivas con otras personas en entornos digitales colaborativos.

Un aspecto íntimamente unido a la competencia en comunicación es la capacidad crítica de los individuos sobre su propio proceso de aprendizaje, una capacidad que debe prepararles para evaluar y reconocer las competencias que tienen, y las que necesitan desarrollar y "poner en conexión" para adquirir conocimiento. Emergen competencias que se orientan hacia la consecución de un pensamiento crítico sobre el propio progreso en el aprendizaje y hacia la comprobación de la eficacia en la aplicación del conocimiento obtenido.

Este contexto justifica el presente trabajo que pretende aproximarse a unas nuevas competencias infocomunicacionales, en el entorno de Educación Superior, convertidas en objetos de análisis

para una nueva versión de la alfabetización en información, la *metaliteracy*, que puede ofrecer unos beneficios si se dota de un marco evaluativo que mida el éxito de sus educandos, éxito que redunde en el de la propia institución educativa que los acoge, por lo que se hace necesario un modelo de indicadores propios.

1.1. La construcción de la Metaliteracy como nueva especialidad para nuevas competencias

La metaliteracy ha ido perfilándose hasta convertirse en una nueva especialidad competencial, con su propio objeto de estudio y metodología, pero siempre a partir de una derivación de anteriores *alfabetizaciones* (múltiples, por tanto), por lo que no parece adecuado presentar un estado de la cuestión, sino una mirada sobre los préstamos que ha ido adoptando de las anteriores alfabetizaciones y que estaban en consonancia con sus objetivos. Tratamos de presentar, pues, una visión evolutiva de cómo se ha ido fundamentando la metaliteracy y así abordar mejor su aplicación en evaluación competencial para Educación Superior.

El desarrollo de la web en general, pero específicamente de la Web 2.0 y su impacto, derivando en Web Social, hizo que desde mediados del primer decenio del siglo XXI las competencias en información y el propio cometido de la alfabetización en información necesitasen una redefinición. Tres fenómenos han colaborado con fuerza en esta evolución:

- a La evolución de las *competencias* sobre una información que varía con la propia evolución de la Web. Si en un momento se llegaba a plantear una dicotomía, dependiendo del ámbito de actuación, entre competencias informáticas y competencias informacionales, hasta el punto de que REBIUN definiría una convergencia entre ambas, como CI2, esta dualidad habría de evaporarse cuando se amplió la noción de competencias, dentro de las acciones de la Unión Europea, que reconocía en su Diario Oficial L394 de 2006 ocho competencias clave, una de ellas, la "competencia digital" para el aprendizaje permanente, que definía como "aquella que implica el uso crítico y seguro de las Tecnologías de la Sociedad de la Información para el trabajo, el tiempo libre y la comunicación. Apoyándose en habilidades TIC básicas: uso de ordenadores para recuperar, evaluar, almacenar, producir, presentar e intercambiar información, y para comunicar y participar en redes de colaboración a través de Internet". Este sesgo afectaba la consideración de las competencias desde la perspectiva educativa y de empleabilidad, por ello la Comisión

en su estrategia "Replantear la Educación" de 2012 reconoce a la competencia digital en un prerrequisito para que el educando optimice todo su proceso de aprendizaje y conocimiento. Se aprobaba el proyecto de DIGCOMP (desarrollado por JRC-IPTS IS), que se publicaría en el *Marco de desarrollo y comprensión de la competencia digital en Europa* (Ferrari, 2013), para establecer los componentes y marco conceptual normalizado, que derivase en acciones de aplicación normalizadas y conjuntas en la Unión. El documento tendría su dimensión educativa, dentro del Plan de Cultura Digital en la Escuela, en el *Marco Común de Competencia Digital Docente v. 2.0* (INTEF, 2013), reconociendo a la competencia digital cinco áreas competenciales: información, comunicación, creación de contenidos, seguridad, resolución de problemas. El Marco se ha actualizado y presenta sus resultados en 2017. Las competencias digitales, pues, desbordaban el objeto inicial de análisis de la alfabetización en información, que debía hallar una senda de especialización, como veremos.

- b La *e-Science*, asociada a la creciente importancia de los *big data* y que el británico National e-Science Centre definía como aquella que se apoya sobre una colaboración global distribuida, en un escenario web de cooperación, visualizable y con acceso a una colección enorme de datos (Szigeti y Wheeler, 2011). De hecho, en 2007, la ARL, en colaboración con la Joint Task Force on Library Support for e-Science, publicó la *Agenda for Developing e-Science in Research Libraries* (como se cita en Berman, 2013), que estimuló el desarrollo de especialidades dentro de la alfabetización en información, especialmente la *transliteracy*, *metaliteracy* y *data literacy*.
- c Proyección de la Web 2.0 en los *social media* y la imagen (3D, realidades virtual y aumentada), medios que reclaman adquirir competencias en comunicación en un entorno interactivo y colaborativo, fundamento de la *visual literacy* y la *new media literacy*. Este contexto se ha visto profundamente afectado por la conectividad mediante dispositivos móviles y por las posibilidades que ofrecen al facilitar la publicación, la compartición y la gestión colaborativa de contenidos, junto con la ampliación de espacios y formatos. Lin y otros (2013) anotan que la Web 2.0 "empodera" a los individuos para "remixar" el contenido e incluir sus propios valores e ideas en los productos existentes, más allá de participar en la co-construcción de ideas. El carácter descontextualizado,

"libre" (sin normas y reglas definidas en las propiedades de los mensajes), desestructurado, hacía aún más necesaria una alfabetización en información renovada (Godwin, 2009).

Este marco propiciatorio vino a sumarse al impacto de la *globalización* en Educación Superior, que se proyectaba en un mayor énfasis en el aprendizaje basado en problemas, el aprendizaje experimental y permanente, y una insistencia en una "educación interdisciplinar" (Kutner y Armstrong, 2012) que, en nuestra opinión, ya no hace referencia sólo a la convergencia y auxiliabilidad interdisciplinar entre distintas áreas de conocimiento, sino al hecho de que en la Web el aprendizaje para el conocimiento debe fundamentarse en la hipertextualidad la horizontalidad de sus contenidos, presentes en los repositorios de objetos digitales educativos. Parece necesaria una competencia para una "navegación" experta, capaz de dibujar un itinerario en la búsqueda y recuperación de contenidos digitales dentro de un mapa de conocimiento y que, como tal, permita planificar una argumentación y unos procesos inferenciales y deductivos que lleven al conocimiento (Brown, 2006).

1.2. Los componentes multialfabetizadores en los orígenes de la Metaliteracy

En la web el conocimiento parece asentarse, pues, en la construcción colectiva de "significados situados", una perspectiva que converge sin duda con el *Conectivismo*, que defiende la construcción del conocimiento y el significado mediante el diálogo y el debate (Siemens, 2010). El potencial que el Conectivismo podía tener en el desarrollo de programas de alfabetización en información ha sido analizado por Dunaway (2011), destacando su funcionalidad dentro de los nuevos entornos colaborativos. El impacto ha sido advertido e impulsado por los profesionales de la información, de modo que, cuando comenzó a plantearse la necesidad de renovar las Normas de alfabetización en información de ALA/ACRL (2016), el Conectivismo se convirtió en la base pedagógica más firme para su actualización, por su mayor solidez y eficacia para el aprendizaje y conocimiento en la Web 2.0 (Bobish, 2011). La expresión de los beneficios desde la perspectiva conectivista ha sido evidenciada por la práctica experimental en un curso formativo de alfabetización en información realizado en la University at Albany en 2008 (McBride, 2012).

Las formas tradicionales de aprendizaje están siendo progresivamente complementadas por formas más eficaces (una "ecología de conocimiento" conectivista) destinadas al aprendizaje social y colaborativo. El aprendizaje social se produce cuando

el individuo percibe la necesidad de un conocimiento a partir de motivaciones cotidianas, que el individuo trata de obtener en las redes sociales, que así permiten generar un ambiente propicio para la comprensión y el conocimiento. Se deduce, por tanto, que la capacidad de gestionar información y relacionarse con el otro a través de los recursos digitales, reclama la adquisición de competencias en información y en comunicación.

Desde esta perspectiva, en ambientes colaborativos también parece conveniente dar respuesta a la elaboración, presentación y distribución de contenidos, mediante una edición colaborativa (Moeller y otros, 2011). La competencia en comunicación, bien visible en entornos competenciales, se hizo sin embargo más compleja cuando las redes sociales se convierten en un instrumento poderoso de generación de opinión y gestor de tendencias, y cuando los dispositivos móviles se convierten en instrumentos de aprendizaje útiles (*m-learning*), una evolución que llevaría a primer término la necesidad de definir la funcionalidad de la identidad digital y la reputación en web en entornos digitales cooperativos y de comunicación (Giones-Valls y Serrat-Brustenga, 2010).

Este marco, que reclama unas competencias en información ampliadas por la comunicación, haría emerger la *new media literacy*, cuyo fundamento conceptual lo armaría H. Jenkins (2009). Él reconocía para esta nueva alfabetización un repertorio de destrezas colaborativas y conectivas, entre las que destacaba: la *navegación transmedia* (comprensión del relato surgido de distintos tipos de información, utilizando diferentes formatos y modalidades); *juicio* o *evaluación*; *apropiación* (interpretación de los constructos o sentido profundo de la información para su comprensión y reutilización); *juego* (resolución de problemas atendiendo a sus partes); *cognición distribuida* (interactuación con las herramientas de distinto tipo); *conducta* (adopción de diferentes identidades según el propósito); *inteligencia colectiva* (organización del propio conocimiento y su comparación con otros, para luego aplicar el conocimiento resultante en un fin común); *negociación* (destreza en debatir dentro de diferentes comunidades, reconociendo las diferentes perspectivas, proponiendo reglas alternativas).

Coherente con este sesgo de competencias enriquecidas y reinterpretadas, como acabamos de ver, habría de surgir la *transliteracy*, cuyo propósito se orienta a ejercitar una capacidad multimodal y multiseccional, pues faculta al individuo para ser competente en la lectura y escritura dentro de las redes sociales digitales, mediante una competencia transversal obtenida por el dominio experto de habilidades de comunicación en plataformas para

mensajes orales, escritos, impresos, audiovisuales (televisión, radio, cine).

1.3. La enunciación de la Metaliteracy y su convergencia con Educación

Este recorrido de "redefiniciones", ha desembocado en la *metaliteracy*. Realmente la *metaliteracy* no define una nueva competencia, sino que se refiere a una convergencia de alfabetizaciones, que habían ido emergiendo y que hemos ido presentando, pero con un aporte muy significativo: la capacidad evaluativa del individuo sobre sus propias competencias, especialmente sobre aquellas que facultan para una conciencia crítica sobre el propio comportamiento informacional en ambientes colaborativos y en medios sociales (Mackey y Jacobson, 2014).

Entre 2011 y 2014 Mackey y Jacobson, (2011) (2014), fueron puliendo el concepto de *metaliteracy* como "disciplina" hasta definir: un *campo de acción*, que tiene por objetivo impulsar la construcción coordinada de conocimientos a través de la adquisición, la producción y el intercambio de contenidos dentro de las comunidades que colaboran en línea; un *objeto de estudio* serán las competencias en información y comunicación como un todo; un *método* a partir de los enfoques tradicionales basados en habilidades para la alfabetización informacional y el uso experto de tecnologías emergentes; un *objetivo* como es la promoción de un marco integral para participar eficazmente en las redes sociales y comunidades en línea. El Conectivismo proporciona a la *metaliteracy* un saludable sustento pedagógico al incluir la "metacognición" entre las competencias, esto es, la capacidad del individuo de desarrollar competencias y adquirir conocimiento a partir de su propia reflexión sobre sus propias facultades alfabetizadoras: los individuos competentes participan activamente en construir y ajustar su propia red de aprendizaje, evaluando con eficacia qué elementos de la red son útiles y cómo mejor desarrollar esa red mediante las herramientas, procesos y elementos de acuerdo con sus necesidades.

La *metaliteracy* ha tenido una convergencia natural con la Educación Superior porque las competencias que desarrolla evolucionan hacia su conversión en elementos medibles para la excelencia académica por su impacto sobre la empleabilidad, innovación y actualización de conocimientos profesionales y técnicos, mediante el aprendizaje permanente. Este proceso ha tenido como instrumento idóneo, no una nueva alfabetización, sino una "etiqueta" para denominar las acciones y estrategias destinadas a fomentar el beneficio de las

alfabetizaciones en Educación Superior, la *academic literacy*, que pretende eliminar las deficiencias no sólo en la búsqueda y recuperación de información de los estudiantes universitarios, sino hacerlos competentes en el método para convertir la información en conocimiento, conforme al método científico, generar y publicar trabajos académicos y científicos (MacMillan y MacKenzie, 2012), para así obtener mejores rendimientos en el aprendizaje y una mayor cualificación profesional. Precisamente es la *academic literacy* la que ha proporcionado una metodología para incorporar la metaliteracy a los currículos universitarios, un aspecto que pretendía desarrollar, en forma de propuesta, el proyecto que presenta este trabajo.

La literatura científica y la experiencia de proyectos aplicados han demostrado cómo la *academic literacy* se hace exitosa por la cooperación educativa y formativa entre el cuerpo académico y el bibliotecario (Gunn y otros, 2011), que hace más fácil la integración en el currículo académico, sobre cuyo diseño, programación, beneficios e impacto ya existe una importante producción científica (Harris, 2013). Se ha publicado el diseño instruccional de un programa en la Universidad de Auckland (Mosen y Wang, 2014) para la cualificación educativa de los bibliotecarios, que luego ha dado como resultado un método para la integración de la *academic literacy* dentro del currículo en esta misma universidad, mediante cursos de método para la investigación (Adams y otros, 2016).

2. METODOLOGÍA DIDÁCTICA Y EVALUACIÓN DE COMPETENCIAS EN LA METALITERACY

Presentado el origen "compuesto y mixto" de la metaliteracy, su marco conceptual de definición y la *academic literacy* como su instrumento eficaz para una práctica docente en Educación Superior, el proyecto debía plantear la base metodológica didáctica para la integración en el currículo. Las investigaciones y prácticas existentes muestran que el mejor modo para esta integración debe partir del diseño metodológico de cursos formativos para aplicación de los "social media", capaces de generar comportamientos informacionales entre los estudiantes. Así se demuestra en el desarrollo de un curso en la primavera de 2011, *Rhetoric and Social Media*, en la University of Scranton, detallado por Witek y Grettano. En su artículo, las autoras presentan las "unidades didácticas", las tareas y los objetivos didácticos del curso hacia el uso educativo y competencial que los estudiantes podían hacer de las redes sociales, en correspondencia con las competencias que Mackey y Jacobson habían determinado para la metaliteracy (Witek y Grettano, 2014).

Los resultados obtenidos en este curso y el análisis de ellos, a la luz de la correspondencia de las competencias de la metaliteracy con las Normas de Alfabetización en información de ALA/ACRL (2016), que ofrece un marco de referencia amplio en la enunciación de sus cuatro normas, son los que ahora más nos interesan: 1) la personalización de archivos e información que permiten las redes sociales hace que la información fluya filtrada al estudiante, cuyo método de investigación cambia, porque más que una estrategia de búsqueda debe ser competente en estructurar lo recuperado, para lo que se aplican las competencias A (el investigador conoce el tipo de formato y medio de comunicación de las herramientas para hacer más accesible la información), B (pensamiento crítico para discriminar la utilidad de información para sus fines de conocimiento) y C (capacidad de comprender y crear un contexto de interpretación de una información que llega desestructurada) de la metaliteracy, unos principios que se acogen en la Norma 2 ALA/ACRL; 2) el comportamiento de los estudiantes respecto de la recuperación e interpretación de la información muestra que para establecer categorías semánticas fijas y precoordinadas no se basan en el contenido ni en la evaluación de fuentes, sino por la "autoridad" de quien ha compartido la información, lo que afecta a las competencias C y F (respeto a la privacidad, ética y propiedad intelectual al interpretar la información en unos entornos tan colaborativos) de la metaliteracy, competencias que se referencian con la Norma 3 ALA/ACRL; 3) la evaluación de la información es social en el entorno de las redes sociales, para lo que se aplican las competencias B, C, D (capacidad de evaluar críticamente contenidos dinámicos) y F de la metaliteracy, competencias relacionables con la Norma 4 ALA/ACRL; 4) la información no sólo es abierta sino que por su flujo diluye la función del productor y consumidor de la información, que se apoyan en las competencias E (capacidad de producir contenidos originales en distintos multimedia) y G (capacidad de compartir información) de la metaliteracy, competencias referidas en la Norma 1 ALA/ACRL.

La metodología didáctica ha encontrado como "espacio educativo" lo que Gee denominaba *affinity space*, un espacio con propiedades para la interacción de las "comunidades de práctica" (Gee, 2007).

En la metodología didáctica, además, ha sido muy relevante la conexión de las competencias derivadas de la metaliteracy con la *competencia de investigación*, aportada por Machin-Mastromatteo. El autor clasifica la competencia de investigación en dos categorías, "contenidos" e "interacción", ordenadas en cinco niveles competenciales: *introducción* donde se aprende los principios de la participación y

de la interferencia en la investigación; *acceso* donde se reflexiona sobre las propias prácticas en redes sociales y se ponen en relación con las de otros colegas; *uso* donde se reflexiona sobre cómo crear en redes sociales mediante el reconocimiento de las propias capacidades, para así determinar cómo generar y conservar contenidos e integrar distintos instrumentos para crear estos contenidos; *evaluación*, diseñando y aplicando unos criterios evaluativos fundamentados; *remate o perfil final* donde se reflexiona sobre la pertinencia de las hipótesis de investigación empleadas y las experiencias obtenidas en las cuatro fases anteriores.

Machin-Mastromatteo proyectó este diseño metodológico en una "arquitectura competencial", DO-RIS (Doing Online Relearning through Information Skills). Muy útil para nuestra investigación es su método en la recolección de datos mediante dos tipos de cuestionarios: un "cuestionario de diagnóstico" que permite analizar los resultados interpretados dentro de cuatro categorías (uso de las redes sociales; aprendizaje en estas redes; experiencias en el uso por cumplimiento de objetivos propuestos; percepción de existencia de distintas alfabetizaciones); un cuestionario "de contraste", cuyo objetivo es precisamente contrastar las respuestas, también dentro de cuatro categorías (publicación, desafíos y oportunidades de las redes sociales, aprendizaje; alfabetizaciones; la mutua modelación en el ambiente colaborativo), con las obtenidas en el primer cuestionario (Machin-Mastromatteo, 2012).

La investigación de Machin-Mastromatteo pone el acento, además, en una de las mayores funcionalidades de las competencias: la excelencia en la evaluación de contenidos, fuentes, web, etc., y en los medios sociales, respecto de las preferencias y los filtros que los individuos establecen. La capacidad de evaluación de contenidos, pues, se transforma en más exigente, en la medida en que se sitúa más próxima al proceso de acceso a la información, esto es, escoger fuentes, filtros, preferencias y canales de información, lo que ya está determinando parcialmente las informaciones accedidas hoy y en el futuro. La relevancia de la evaluación de contenidos es tanto mayor porque, a diferencia de la cultura impresa donde el trabajo de selección lo realiza el agente emisor (Lemos, 2009), en ambientes digitales pasa al receptor-consumidor la tarea de evaluar los contenidos recuperados y lo producido por los medios de comunicación de masas (como advierte la medialiteracy). Esto implica que la evaluación debe dirigirse también a la capacidad en comunicación.

Parece apropiado, en estas condiciones, aproximarse a los criterios, en principio útiles, para evaluar estas competencias informacionales y comunicacionales.

Buckingham (2008) apuntó al criterio sobre la autoridad de las fuentes en medios sociales, otorgando una mayor fiabilidad a las fuentes institucionales sobre las personales, aun cuando algunas empresas y gobiernos divulgan su información de forma incompleta, incorrecta y poco confiable, por lo que la autoridad de la fuente *per se* no es una garantía de calidad. Además, en redes sociales, con múltiples autorías, ¿cómo adoptar el criterio de la autoridad de la fuente?

De otro lado, los medios sociales, sin duda, están impactando en la construcción de los significados. Como observaron en su investigación Witek y Gretanno (2014), los estudiantes no categorizan la información basándose en criterios de evaluación de contenidos o de las fuentes, sino observando quién comparte la información con ellos o quién transmite la confiabilidad construida por las relaciones sociales.

La evaluación en redes sociales, por definición, es un factor más compartido que individual. Una vez que la información comunicada ya ha recibido comentarios o críticas de otros usuarios, esto impacta sin duda en su interpretación y evaluación, convirtiéndose en un modo de medición cualitativa de liderazgo, transparencia o influencia (González Fernández-Villavicencio, 2012). Otro factor y criterio evaluativo es el uso de los contenidos generados por interacción, en oposición a la información "acabada", generada unilateralmente. Bajo esta óptica, el comportamiento informacional es una variable evaluable muy significativa en redes sociales, por tanto en metaliteracy. Este criterio se fundamenta, además, en la noción de *cognición distribuida*, esto es, la capacidad de interactuar significativamente con las personas que apoyan nuestro aprendizaje mediante sus propios conocimientos o procedimientos y prácticas (Alper, 2013).

Los iniciales criterios evaluativos que apuntaban a la metaliteracy daban preferencia a que los individuos reconozcan sus debilidades y potencialidades, de modo que pudiesen mejor aprovechar los espacios de los medios sociales, sea para suplir esas debilidades, sea para buscar alianzas con las que puedan trabajar en colaboración. Es una constatación que se hacía explícita en las investigaciones de Brandão y Borges (2014), y de Thomas y Hodges (2015) respecto a la producción de contenidos. El primer trabajo mostraba las carencias de un grupo de estudiantes de nivel superior cuando necesitaban contextualizar y presentar la información producida por ellos. El segundo trabajo destacaba un aspecto importante en la producción de contenidos de trabajos científicos y textos académicos por los estudiantes, la competencia y una formación equivalente para exponer su propia opinión y una perspectiva original en el discurso académico.

3. UN METAMODELO DE EVALUACIÓN PARA LA METALITERACY: ELEMENTOS EVALUABLES

El proyecto, pues, reconocía como la aplicación metodológica didáctica en Educación Superior debía ir íntimamente relacionada con una evaluación consistente para la metaliteracy, tal como apuntaban algunos criterios y experiencias positivas en ambientes de comunicación. El carácter y naturaleza particulares de la metaliteracy aconsejaban buscar una evaluación propia, a partir de una perspectiva evaluativa diferente con una formulación apropiada de indicadores, asentada en un "modelo", que debe implicar la representación de la realidad conforme a un determinado propósito (Pidd, 2009).

Un primer hecho a considerar es, como hemos ido argumentando, que la metaliteracy armoniza con los valores de la cultura participativa, al mismo tiempo que destaca la consciencia que el individuo debe tener respecto del propio alcance, difusión e impacto de sus acciones sobre sí y sobre otras personas.

De otro lado, el carácter transversal y "compuesto" de la metaliteracy parece adecuarse mejor, es nuestro análisis en la investigación, a la formulación de un *metamodelo*, que en definición de Henderson-Sellers y González Pérez (2005) consiste en un conjunto de conceptos, normas y relaciones utilizados para definir una "familia" de metodologías relacionadas. En un esclarecedor artículo, Shenton y Hay-Gibson (2012) formulaban y exponían tres tipos de metamodelos, entre los que es el segundo el más apropiado para la metaliteracy, desde luego para nuestro propósito de investigación: este metamodelo, mediante la interoperabilidad de las metodologías que engloba, es capaz de representar al más alto nivel los conceptos semánticamente más relevantes, así como condensar-resumir los contenidos de los mensajes que se desee considerar. Precisamente Mackey y Jacobson vieron que las metodologías de este tipo de metamodelo parecían adecuarse muy bien a la metaliteracy, porque otorgan un marco para identificar similitudes en la semántica de conceptos a través de diferentes alfabetizaciones, permitiendo la unificación de los conceptos relacionados. No cabe duda que es un metamodelo que apoya con eficacia a los metabuscadores en Web.

La necesidad de proponer un modelo evaluativo en metaliteracy, a partir de un metamodelo útil, se hacía más interesante en Educación Superior cuando, emergiendo de la e-Science, algunos profesionales comenzaron a hablar, dentro de la *academic literacy*, de la *Science literacy*, entendida desde la óptica del comportamiento informacional: la competencia de una persona para la resolución de problemas y toma de decisiones por un sólido

conocimiento científico, por lo que incluye competencias en "cultura científica", "cultura cívica científica", "cultura práctica científica", "cultura estética científica" y "cultura de práctica científica".

En la formulación de un modelo evaluativo la metaliteracy, lógicamente, se inspiró en las alfabetizaciones con las que había convergido en su proceso de definición, por la propia naturaleza de la *Science literacy*, especialmente cuando se hizo necesario identificar los *objetos evaluables* para las nuevas competencias. Así, al tiempo que se producía la confluencia de la *Science literacy* con la "Ciencia Abierta" en los ámbitos académicos, provocando una complementariedad entre *transliteracy* y *metaliteracy* (Holden, 2012), los datos abiertos impactaban en la investigación científica, lo que permitía el desarrollo de la *dataliteracy* como el método idóneo para hacer competentes a los individuos en el acceso, interpretación, evaluación crítica, así como en la gestión experta y ética de los datos, acorde al desarrollo de los big data (Calzada y Marzal, 2013).

Otro objeto evaluable emergió por el desarrollo de las posibilidades técnicas para la visualización de la información en Web, así como un uso eficiente de la imagen para el conocimiento en Web a través de las Realidades Virtual y Aumentada. La *visual literacy* (perfectamente diferenciada de la *medialiteracy*, porque no cultiva la competencia en el uso experto de los medios, sino en comprender y analizar los componentes contextuales, culturales, éticos, estéticos, intelectuales y técnicos referidos al uso pero también, de forma importante, a la producción de materiales audiovisuales), debía ser promovida en los ámbitos académicos y científicos por convertirse los productos de las Realidades Virtual y Aumentada, entre otros, en objetos digitales educativos con una efectividad de primera magnitud en la generación de conocimiento y su transmisión. La *visual literacy*, por tanto, se dirigía, mediante estos objetos evaluables icónicos, tanto a las competencias en información, como también, evidentemente, a las competencias en comunicación, lo que la ha aproximado a la cooperación con la *new media literacy* y la *metaliteracy*. El *m-learning* en la Educación Superior y el uso de las imágenes por parte de los estudiantes, preferentemente en las redes sociales donde tienden además al remix, han impulsado estos objetos evaluables y sus competencias.

De nuevo, la relevancia de este objeto evaluable se ha evidenciado desde las buenas prácticas profesionales, como acreditan iniciativas en torno a las competencias de comunicación visual e icónica en bibliotecas universitarias (Denda, 2015). De hecho, asociaciones bibliotecarias de indudable im-

pacto, como las estadounidenses ALA/ACRL, optaron por editar las Normas de Visual Literacy para Educación Superior, en 2011. Son particularmente interesantes, para nuestro estudio, los indicadores que se proponen en las siete normas establecidas para medir el logro de las competencias en Visual Literacy (Hatwig y otros, 2013). Junto con estas iniciativas normativas, se han desarrollado acciones de buenas prácticas como el diseño programático e instruccional para un uso eficiente de las redes sociales por parte de los estudiantes, aplicado por L.M. Bridges en 2012, a partir de la propuesta de la alfabetización en medios sociales que publicaba Rheingold en 2010 para las universidades californianas de Berkeley y Stanford. El curso de Rheingold se organizaba en unidades didácticas entre las que se impartían competencias para el gaming, ética, comunidades virtuales, imágenes, privacidad, identidad, empleo o adicción a Internet.

La praxis didáctica de los cursos competenciales ha desembocado en el interés por otro objeto evaluable, una competencia básica, que es aquella que faculta a los estudiantes en el dominio experto de la *argumentación* para mejor obtener conocimiento de todo contenido digital, medible mediante el modelo Toulmin, según demostraron en la aplicación de su curso competencial sobre argumentación Radcliff y Wong (2015), en St. Mary's College y la California State University East Bay, ambas en California.

4. FUNDAMENTOS Y PRECEDENTES PARA UN MODELO DE EVALUACIÓN PARA METALITERACY

El análisis de objetos evaluables que podrían inspirar un modelo o metamodelo evaluativo para la metaliteracy, señalaba que la búsqueda podía inspirarse en los modelos de evaluación que, desde 2011, se estaban enunciando para la new media literacy. Debemos detenernos un momento sobre qué fundamentos conceptuales de la new media literacy pueden derivar en su conversión como referente para la evaluación en metaliteracy.

En un sólido trabajo, Chen y otros (2011) abordaban la fundamentación epistemológica de lo que debía ser y cómo debía entenderse la new media literacy, a partir de dos pilares de argumentación: a) la *caracterización* de la new media literacy atendiendo a criterios técnicos (modularidad, multimodalidad, carácter híbrido e interactividad en diferentes plataformas digitales), pero con especial énfasis en criterios socio-culturales (la "construcción" de los medios, ideología y valores integrados en los medios, finalidades establecidas para los medios); b) *un marco conceptual*, muy atractivo, que inte-

gra la concepción y comprensión de sus competencias en dos ejes, de abscisas y coordenadas: de un lado, en el eje de coordenadas, las competencias se inscriben en dos tipos o categorías de análisis, de forma yuxtapuesta, unas "funcionales" (dominio experto de los aspectos técnicos), otras "críticas" (interpretación a partir del contexto socio-cultural); de otro lado, en el eje de abscisas, estas mismas competencias se inscriben en otros dos tipos o categorías, también yuxtapuestas, unas capacitan para el "consuming" (eficacia en el consumo de la información, por una capacidad experta de selección de contenidos), otras capacitan para el "prosuming" (capacidad no sólo para asimilar la información, sino también de crear, a partir de lo asimilado), una categorización que en 1981 apuntara Toffler y que más tarde desarrollaría Buckingham (2003).

Esta interesante propuesta conceptual de modelo habría de ser desarrollada, como una propuesta metodológica y aplicativa, por Lin y otros (2013), orientada a los desafíos planteados por la Web 2.0. A las categorías enunciadas por Chen en sus dos ejes, arriba anotadas, sumaban en cada una de las categorías unos criterios, de manera que obtenían un marco conceptual más definido, más definitivamente evaluativo, cuando propusieron el "refinamiento" del marco conceptual mediante la asociación de diez "indicadores": cinco indicadores para el consuming, a saber, destrezas en el consuming, comprensión, análisis, síntesis, evaluación; cinco indicadores para el prosuming, esto es, destrezas en prosuming, distribución, producción, participación, creación.

Este trabajo fundamentaría un desarrollo definitivamente metodológico y evaluativo, a partir de la investigación de Lee y otros (2015), que recogía los indicadores de Lin y otros, arriba detallados, con el fin de proponer y aplicar un instrumento de medición para las competencias propias de la new media literacy. Es interesante cómo para crear el modelo de este instrumento de medida, los autores hacen un atractivo diseño previo que considera: a) unas *escalas de medición* para lo que parten de tres aspectos de la escala Likert, la "autonomía" (o "independencia") del usuario-educando dentro de la comunidad, el "acuerdo" con las normas y reglas de funcionamiento de la comunidad, y la "frecuencia" de intervención en las actividades propuestas, b) las *definiciones operacionales*, considerando las "destrezas" como punto de partida, luego la "productividad desde el consuming al prosuming, y la "críticidad" desde la competencia funcional a la crítica, la "sociabilidad" o la competencia desde el dominio experto de la Web 1.0 a la 2.0; c) *interpretación de datos* para la que los autores utilizan la taxonomía de Bloom

y los referentes de PISA 2009. A partir de este modelo generaron un cuestionario, cuya calidad de comprensión por los encuestados quedó acreditada mediante un pretestado, y luego se aplicó a un grupo-meta utilizando la plataforma Qualtrics.

El cuestionario tenía por objetivo comprobar la participación a través de los medios sociales en cuatro niveles: a) el *consumo funcional*, que se refiere a la habilidad de acceder y entender los contenidos mediáticos; b) el *consumo crítico*, referido a la habilidad de evaluar el contexto económico, político y cultural del contenido; c) el "prosumer" *funcional*, como habilidad para participar en la creación de contenidos mediáticos; d) "prosumer" *crítico*, cuando se tiene la capacidad de evaluar el contenido mediático durante su creación.

Método e instrumento de medición aparecían como muy útiles al proyecto, ya que al relacionar el "prosumer" crítico con un individuo que comprende su posición e identidad en la producción, distribución y participación a través de los medios, el concepto se aproxima al aspecto "metacognitivo" de la metaliteracy, esto es, para entrelazar sus valores sociales con una construcción y utilización crítica de contenidos digitales, el individuo necesita emplear una serie de competencias y, por ende, estar consciente de las que posee y de las que necesita buscar en red. La metaliteracy puede así dirigir el foco de su evaluación competencial también hacia la consideración de que el conocimiento puede lograrse por la producción, distribución y comunicación de la información en ambientes colaborativos (Jacobson y Mackey, 2013).

5. UNA PROPUESTA DE ESTRUCTURA DE UN MODELO PARA LA EVALUACIÓN DE COMPETENCIAS EN METALITERACY

Se llegaba así a la formulación de una propuesta de modelo evaluativo para metaliteracy, que presenta este trabajo, en el marco de un proyecto de investigación dentro de un programa de cooperación iberoamericano, cuyo objeto primordial de estudio eran las competencias infocomunicacionales consideradas desde la fundamentación conceptual de la metaliteracy.

El diseño del modelo se proyectó buscando establecer una relación directa entre la propuesta conceptual, cuyo proceso y referentes han sido ya expuestos, y su verificación práctica. Por esta razón se presenta el proceso de construcción de los indicadores del modelo, etiquetándolos y estructurándolos a partir de unas categorías y subcategorías diseñadas para el modelo, para luego definir los indicadores conforme a las competencias y objetivos competenciales que investigado-

res especialistas han ido reconociendo para la new media literacy y la metaliteracy.

La finalidad y funcionalidad de los indicadores partieron de la hipótesis de que deberían configurarse conforme a los siguientes determinantes: a) estuviese correctamente inscrito en su respectiva categoría y subcategoría y, por tanto, incluido dentro de una argumentación teórica (que sustenta la arquitectura de categorías del modelo) que lo justificase y lo relacionase sin ambigüedad con el concepto que se pretende medir; b) fuese efectivo y eficiente en medir el concepto (competencia) que pretende medir; c) asegurase su confiabilidad, en el sentido de que obtuviese iguales resultados con independencia del ámbito de aplicación, prestando, además, su interoperabilidad y comparación significativa con otros modelos evaluativos (Marzal, 2010). Los indicadores deberían medir e identificar las competencias infocomunicacionales necesarias, pero también el nivel que podrían tener los educandos en la consecución y dominio de estas competencias.

A partir de estos considerandos conceptuales, metodológicamente el diseño del desarrollo empírico del trabajo se organizó en torno a tres fases: a) diseño de un esquema para la categorización de las competencias infocomunicacionales, a partir de los fundamentos teóricos y proyectos antes expuestos; b) diseño y definición de los indicadores; c) aplicación empírica de los indicadores y su modelo, mediante la creación de un cuestionario ad hoc, para lo que se determinaría, como grupo-meta, un conjunto de estudiantes de postgrado en las áreas de Información y Documentación (másteres) de tres universidades. El presente trabajo aborda la fase a) y los fundamentos de la fase b).

La relación directa entre el modelo conceptual y el aplicativo estuvo inspirada en el diseño presentado por Lee y otros (2015), especialmente atractivo al señalar la "nueva brecha *digital*", cuyo motivo debía hallarse en la competencia (o ausencia de ella) de interacción con el "otro", por lo que sugiere indicadores desde el punto de vista de la competencia en comunicación, porque evaluarían precisamente la relación entre las personas respecto a producción y distribución, creación y participación.

El objetivo básico del proyecto, el diseño de un modelo estructurado de indicadores competenciales para metaliteracy, hizo que se necesitase implementar la propuesta modélica y metodológica de Lee y otros (2015), identificando y aplicando otras estrategias de diseño de modelo, para desarrollar adecuadamente las fases b) y c) del proyecto. Desde esta perspectiva del diseño modélico fueron particularmente útiles: la investigación

de Hargittai (2005) sobre el uso de los recursos de internet por parte de cien usuarios seleccionados aleatoriamente; la experiencia de Chang y otros (2011) con 1539 estudiantes que procedieron a autoevaluar sus competencias utilizando el Media Literacy Self-Assessment Scale (MLSS); y el modelo de la toma de decisiones para la evaluación de competencias en e-literacy o alfabetización tecnológica (Brandtweiner y otros, 2010).

Una vez que se articulara la estructura del modelo, ésta reclamaría una definición de indicadores propios, como venimos anunciando. Un primer referente natural fueron los indicadores contemplados dentro de la publicación de las *Normas*, entre aquellas más apropiadas para los objetivos del proyecto:

a) El diseño genérico de indicadores para la medición del desarrollo de la Sociedad de la información, por parte de la Unión Europea, en 2003, *SI-BIS (Statistical Indicators Benchmarking for Information Society)*.

b) Indicadores para la medición del desarrollo de las TIC en Educación por parte de UNESCO en 2009 (*Guide to measuring Information and Communication Technologies, ICT, in Education*).

c) Entre las Normas para Alfabetización en Información: 1) los indicadores de las Normas de Alfabetización en Información de ANZIIL, editados por A. Bundy (2004); 2) aquellos definidos para las *Directrices sobre el desarrollo de habilidades en información para el aprendizaje permanente* de IFLA, publicadas por J. Lau (2008); 3) los consignados en la publicación *Towards Information Literacy Indicators* de UNESCO, por R. Catts y J. Lau (2008), así como los que reconocía la Comisión Europea en *Measuring Digital Skills across the EU: EU wide Indicators of Digital Competence* (2014); 4) los indicadores que se definían para la mediailiteracy por parte de UNESCO, *Towards Media and Information Literacy Indicators* (2011); 5) los que proponían Ferrés e Piscitelli (2012), o los propuestos en el programa competencial de UNESCO y Athabasca University, *e-Media and Information Literacy Course* (2015), disponible en línea a partir de la plataforma moodle; 6) la enunciación de los "objetivos competenciales, que Mackey y Jacobson establecían en su monografía sobre Metaliteracy (2014); 7) la propuesta que Borges (2014) presenta como una formulación conceptual para las competencias infocomunicacionales; 8) los indicadores que la ALA/ACRL, reconociendo los decisivos cambios en la Web desde su primera edición de Normas de Alfabetización en Información en 2000, se fundamentaban en el *Framework for Information Literacy in Higher Education* (2015) y publicaban como nuevo marco normativo (que derogaba el de

2000), en 2016, *Information Literacy Competency Standards for Higher Education* (<http://www.ala.org/acrl/standards/informationliteracycompetency>), sumamente interesante porque reconoce la alfabetización y competencias tecnológicas como el fundamento para las competencias informacionales y las digitales, a partir del concepto de *fluency*, otorga un notorio protagonismo al apoyo pedagógico, pero además hace un análisis relevante sobre la evaluación de estas competencias, desde una perspectiva útil en nuestro trabajo. Parece una vía interesante donde analizar las competencias digitales, planteadas en la Unión Europea, con una dimensión próxima a la metaliteracy.

Todos los anteriores considerandos confirmaban la necesidad de que los indicadores respondiesen a un modelo de definición bien estructurado, a través de la división en tres clases (indicadores de Competencias en Información, Competencias en Comunicación, Competencias en Metaliteracy), así como por la identificación clara por su etiqueta ("título" del indicador) y por sus objetivos de medición competencias y de actitudes para el comportamiento informacional, de modo que asegurase un análisis e interpretación "estandarizados" y, además, se proyectasen con posterioridad en unas preguntas sólidas para los cuestionarios.

6. MODELO DE INDICADORES

El proyecto, por tanto, acometió la definición del modelo de indicadores en dos fases: en una primera fase se diseñó una estructura de los indicadores más útiles al proyecto; en una segunda fase, a partir de la estructura antes diseñada, se integraba en un único modelo los indicadores de medición de las competencias anotadas, dentro de la estructura de consuming y prosuming, alfabetización funcional y crítica para el consumo de contenidos.

6.1. Estructura del modelo de indicadores por competencias

Se recogen en tablas, según hemos anotado, las tres áreas competenciales, información, comunicación y metaliteracy, dentro de cada cual se procedía a señalar para cada indicador la competencia a medir, su definición (el objeto o campo de la competencia) y los objetivos competenciales (expresión en % del nivel de consecución de la competencia y su finalidad) a los que se refería. El nivel de consecución de competencia debe plantearse en una fase posterior de aplicación metódica de los indicadores, por cuanto es un nivel y no una calificación o nota: la competencia se logra o no, y el nivel lo que debe indicar es el percentil de excelencia en la consecución de la competencia.

Presentamos, entonces, las tres áreas competenciales en sendas tablas, que para mayor coherencia del discurso del trabajo, aquellas referidas a los Indicadores de Competencias en Información y de Competencias en Comunicación se publican en Anexos, reservando la Tabla I de Indicadores de Metaliteracy al cuerpo de texto.

6.2. Estructura integrada de modelo de indicadores para Metaliteracy

A partir de la estructura propuesta, según se ha argumentado en el trabajo, se procedió a implementar este modelo conforme al diseño que se contemplaba para indicadores de new media literacy y los "goals" competenciales que se reconocían para la metaliteracy, resultando una "estructura integrada" de un modelo evaluativo, del que derivarán las baterías de preguntas integrantes de cuestionarios que servirán como instrumento de medición de las competencias en metaliteracy.

Se estructuraron cuatro clases de competencias, articuladas en categorías de competencias que, bajo ellas, registran los indicadores propuestos:

1. Primera Clase: Consumo eficaz de Contenidos

Engloba dos categorías competenciales:

I. Alfabetización funcional para el consumo de contenidos, que registra como sus indicadores:

- **Habilidades para el consumo informativo:** Destrezas técnicas para el consumo de contenidos

- **Comprensión:** Captación del significado de los contenidos de los media

II. Alfabetización crítica para el consumo de contenidos, que consta de los siguientes indicadores:

- **Análisis:** Habilidades para procesar analíticamente los contenidos mediante unos criterios objetivos y regulados
- **Síntesis:** Habilidades para crear un mapa semántico de los contenidos recuperados, sintetizar sus clústeres semánticos y etiquetarlos conforme a los intereses de conocimiento
- **Evaluación:** Competencia para el cuestionamiento y evaluación crítica de los mensajes

2. Segunda Clase: Producción eficaz de contenidos

También incluye dos categorías competenciales:

I. Alfabetización funcional para la producción de contenidos, cuyos indicadores son tres:

- **Prosuming:** Destrezas técnicas para la producción de contenidos
- **Distribución:** Habilidades para la distribución y diseminación eficaz de contenidos
- **Producción:** Competencias para duplicar, reutilizar o remezclar contenidos

II. Alfabetización crítica para la producción de contenidos, con dos indicadores:

- **Participación:** Capacidad para participar interactiva y críticamente en entornos de medios participativos

Tabla I. Indicadores de Metaliteracy

Competencia	Definición	Objetivos
Transversalidad y autoevaluación	Competencia en autoevaluar las propias competencias en información y comunicación para conectarlas con las de los otros	<ul style="list-style-type: none"> ✓ Autoevaluar las propias competencias para acceder, evaluar y gestionar la información ✓ Reconocer la propia responsabilidad en la producción y generación de conocimiento y percibir el impacto de este conocimiento ✓ Autoevaluar los productos de información generados ✓ Identificar las personas o grupos que puedan sumar sus competencias a las propias ✓ Realizar una autoevaluación para determinar las propias motivaciones en aprender e investigar y evaluar si las propias acciones son adecuadas para este fin ✓ Definir expectativas a corto, medio y largo plazo ✓ Evaluar los resultados del aprendizaje y conocimiento obtenidos ✓ Considerar los factores y circunstancias que influyen en el alcance de los objetivos y sus resultados ✓ Analizar las potencialidades y los desafíos de los recursos electrónicos para el conocimiento ✓ Comprender la naturaleza, características y propiedades de la comunicación e información en medios digitales para fortalecer sus competencias

- **Creación:** Competencias para tener la iniciativa en la creación de contenidos y/o acrecentar la comprensión crítica de los valores socioculturales integrados, así como cuestiones ideológicas

3. Tercera Clase: Interacción

Una clase competencial, cuyas categorías derivan directamente en indicadores:

I. Establecer y mantener la comunicación:

Capacidad para establecer y mantener la comunicación con otras personas o entidades, adecuando el canal de comunicación y el lenguaje y registro idiomático al interlocutor

II. Crear lazos sociales:

Competencia para desarrollar relaciones sociales saludables en ambientes digitales, basadas en el respeto a la diversidad y en la afirmación de la identidad

III. Construir conocimiento en colaboración:

Competencia para generar conocimiento y trabajar en colaboración

4. Cuarta Clase: Metaliteracy

Una clase cuyas categorías (también derivan directamente en indicadores) y su definición se han hecho conforme a los "Goals", que Mackey y Jacobson (2014) han definido para la Metaliteracy.

I. Evaluación crítica:

Competencia para evaluar críticamente el contenido, incluyendo el contenido dinámico en línea, que cambia y se transforma

II. Privacidad, ética y propiedad intelectual:

Competencia para distinguir los contenidos que pueden ser diseminados y reutilizados, reconocer los derechos intelectuales con seguimiento de normas, reconocer la propia responsabilidad ética y legal en la documentación y contenidos compartidos, como la repercusión de la propia difusión de información

III. Colaboración:

Competencia para compartir informaciones y colaborar en una variedad de ambientes participativos

IV. Aprendizaje a lo largo de la vida:

Competencia para conectar estrategias de enseñanza e investigación con procesos de aprendizaje a lo largo de la vida y con objetivos personales, académicos y profesionales

7. CONCLUSIONES Y DESARROLLOS FUTUROS

Las iniciativas dirigidas hacia la alfabetización en información deben evolucionar en un futuro inmediato para incluir competencias que contemplan valores como la participación, compromiso en

prácticas comunicativas constructivas y la constante autoevaluación, como también competencias específicas para la lectura eficaz de la imagen en entornos crecientemente icónicos. Estos valores, junto con el iconismo, son también nodales para la noción de metaliteracy, bajo la cual los educandos son impelidos a colaborar, crear y compartir información usando y expandiendo los recursos existentes. La metaliteracy es reconocida como una alfabetización especializada dentro de la alfabetización en información, que conserva su función y papel nuclear, en la percepción de los especialistas.

Desde esta óptica son varias las conclusiones a las que ha querido llegar este trabajo. En primer lugar, la alfabetización en información con el fin de conservar su carácter de pivote primordial en eliminar las brechas digitales para la inclusión social, así como pilar de desarrollo para la sociedad del conocimiento, ha comenzado a especializarse en distintos ámbitos donde las competencias en información son fundamentales, uno de ellos, sin duda, es convertirse en factor inexcusable para la excelencia académica de la Educación Superior y, por ende, de las universidades. Este impulso de especialización ha hecho emerger la alfabetización académica, que ha hallado el diseño de sus programas competenciales y su espacio en los Academic Skills Centres. Sin embargo, la propia evolución y naturaleza del modelo educativo competencial en Educación Superior ha hecho cambiar las propiedades de las competencias en información.

Así llegamos a la segunda conclusión del trabajo. En el contexto de una alfabetización académica debían redefinirse las competencias en información, un camino que ha sido posible por la confluencia con dos alfabetizaciones realmente relevantes: la visual literacy para imágenes, y la new media literacy para las competencias en comunicación. Ha sido esta senda de convergencia la que ha ido diseñando el proyecto y, como resultado este trabajo, resultando una percepción interesante de la metaliteracy.

Esta perspectiva conceptual de la metaliteracy, nos lleva necesariamente a la tercera conclusión de este trabajo. Todo programa competencial de alfabetización debe ir inevitablemente unido a una evaluación, por indicadores competenciales, que garantice la consecución de las competencias requeridas. El proyecto, y por tanto este trabajo, procedía a proponer un modelo de indicadores competenciales para la metaliteracy, pero, coherente con su argumentación, derivado de los modelos de indicadores de la alfabetización en información, complementada por la new media literacy. A partir de investigaciones y experiencias aplicativas anteriores (especialmente desde la new media

literacy) se presenta la propuesta de un modelo para el debate académico.

El modelo, por lo demás, nace con una evidente vocación de investigación aplicada: derivar en instrumentos apropiados de medición, que puedan ser aplicados a colectivos concretos de estudiantes de Educación Superior, para analizar y debatir sobre sus resultados. Es un futuro desarrollo en que el proyecto está inmerso.

9. REFERENCIAS

- Adams, C.; Buetow, S.; Edlin, R.; Zdravkovic, N.; Heyligers, J. (2016). A collaborative approach to integrating Information and Academic Literacy into the Curricula of Research Methods Courses. *The Journal of Academic Librarianship*, vol. 42 (3), 222-231 <https://doi.org/10.1016/j.acalib.2016.02.010>
- ALA/ACRL (2011). *Visual literacy competency standards for higher education*. <http://www.ala.org/acrl/standards/visualliteracy> [23/01/2017]
- ALA/ACRL (2016). *Information Literacy Competency Standards for Higher Education*. <http://www.ala.org/acrl/standards/informationliteracycompetency> [24/01/2017]
- Alper, M. (2013). Developmentally appropriate new media literacies: Supporting cultural competencies and social skills in early childhood education. *Journal of Early Childhood Literacy*, vol. 13 (2), 175-196. <https://doi.org/10.1177/1468798411430101>
- Berman, E. (2013). Transforming Information Literacy in the Sciences through the lens of e-Science. *Communications in Information Literacy*, vol. 7 (2), 161-170.
- Bobish, G. (2011). Participation and pedagogy: Connecting the social web to ACRL learning outcomes. *The Journal of Academic Librarianship*, vol. 37 (1), 54-63 <https://doi.org/10.1016/j.acalib.2010.10.007>
- Borges, J. (2014). Competências infocomunicacionais na atuação política de organizações da sociedade civil. *Tendências da Pesquisa em Ciência da Informação*, vol. 7 (2), 81-102
- Brandão, G.; Borges, J. (2014). Emprego das competências em informação pelos estudantes de arquivologia da universidade federal da bahia. *Ágora*, vol. 24 (49), 277-310.
- Brandtweiner, R.; Donat, E.; Kerschbaum, J. (2010). How to become a sophisticated user: a two-dimensional approach to e-literacy. *New Media & Society*, vol. 12 (5), 813-833 <https://doi.org/10.1177/1461444809349577>
- Bridges, L.M. (2012). Librarian as professor of Social media Literacy. *Journal of Library Innovation*, vol. 3 (1), 48-65
- Brown, T.H. (2006). Beyond constructivism: Navigationism in the knowledge era. *On the Horizon*, vol. 14 (3), 108-120 <https://doi.org/10.1108/10748120610690681>

8. AGRADECIMIENTOS

Este trabajo ha sido financiado por el Consejo Nacional de Desarrollo Científico y Tecnológico – Brasil.

ACKNOWLEDGEMENTS

This work has been funded by the National Council for Scientific and Technological Development - Brazil

- Buckingham, D. (2003) *Media Education: Literacy, Learning and Contemporary Culture*. Polity Press: Cambridge
- Buckingham, D. (2008). Defining digital literacy-what do young people need to know about digital media? En C. Lankshear; M. Knobel (editores), *Digital literacies* (pp. 73-90). New York: Peter Lang.
- Bundy, A. (ed.) (2004). *Australian and New Zealand Information Literacy Framework principles, standards and practice*. ANZIIL: Adelaide
- Calzada, F.J.; Marzal, M.A. (2013). Incorporating Data Literacy into Information Literacy Programas: Core Competencies and Contents. *Libri*, vol. 63 (2), 123-134 <https://doi.org/10.1515/libri-2013-0010>
- Catts, Y; Lau, J. (2008). *Towards Information Literacy Indicators*. UNESCO: París
- Chang, C.S.; Liu, E.Z.F.; Lee, C.Y.; Chen, N.S.; Hu, D.C.; Lin, C.H. (2011). Developing and validating a media literacy self-evaluation scale (MLSS) for elementary school students. *The Turkish Online Journal of Educational Technology*, vol. 10 (2), 63-71
- Chen, D-T. ; Wu, J. ; Wang, Y-M. (2011). Unpacking media literacy. *Journal on Systemics, Cybernetics and Informatics*, vol. 9 (2), 84-88
- Comisión Europea (2003). *SIBIS: Statistical Indicators Benchmarking Information Society. Indicator Handbook*. http://www.sibis-eu.org/files/Sibis_Indicator_Handbook.pdf [24/01/2017]
- Comisión Europea. *Measuring Digital Skills across the EU: EU wide Indicators of Digital Competence*. <https://ec.europa.eu/digital-single-market/en/news/measuring-digital-skills-across-eu-eu-wide-indicators-digital-competence> [24/01/2017]
- Denda, K. (2015). Developing Interview Skills and Visual Literacy: A New Model of Engagement for Academic Libraries. *Portal: Libraries and the Academy*, 15(2), 299-314 <https://doi.org/10.1353/pla.2015.0024>
- Dunaway, M.K. (2011). Connectivism. Learning theory and pedagogical practice for networked information landscapes. *Reference Services Review*, vol. 39 (4), 675-685 <https://doi.org/10.1108/00907321111186686>
- Ferrari, A. (2013). *DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe*.

- European Commission Joint Research Centre. Institute for Prospective Technological Studies [eds. Y. Punie; B. Brecko]: Luxemburgo
- Ferrés, J.; Piscitelli, A. (2012). La competencia mediática : propuesta articulada de dimensiones e indicadores. *Comunicar*, v. 19 (38), 75-82 <https://doi.org/10.3916/C38-2012-02-08>
- Gee, J.P. (2007). *Good Video Games and Good Learning: Collected Essays on Video games Learning and Literacy*. Peter Lang; New York <https://doi.org/10.3726/978-1-4539-1162-4>
- Giones-Valls, A.; Serrat-Brustenga, M. (2010). La gestión de la identidad digital: una nueva habilidad informacional y digital. *BiD: textos universitaris de biblioteconomia i documentació*, 24. <http://bid.ub.edu/24/giones2.htm> [24/01/2017].
- Godwin, P. (2009). Information Literacy and Web 2.0: is it just hype? *Program: electronic library and information systems*, vol. 43 (3), 264-274 <https://doi.org/10.1108/00330330910978563>
- González Fernández-Villavicencio, N. (2012). Alfabetización para una cultura social, digital, mediática y en red. *Revista Española de Documentación Científica*, 35 (Monográfico), 17-45. <https://doi.org/10.3989/redc.2012.mono.976>
- Gunn, C.; Hearne, S.; Sibthorpe, J. (2011). Right from the start: A rationale for embedding academic literacy skills in university courses. *Journal of University Teaching & Learning Practice*, vol. 8 (1), 1-6
- Hargittai, E. (2005). Survey measures of web-oriented digital literacy. *Social Science Computer Review*, vol. 23 (3), 371-379 <https://doi.org/10.1177/0894439305275911>
- Harris, B. R. (2013). The New ACRL Information Literacy Competency Standards. Revising reception. *Communications in Information Literacy*, vol.7 (2), 139-145.
- Hattwig, D.; Bussert, K.; Medaille, A.; Burgess, J. (2013). Visual literacy standards in higher education: new opportunities for libraries and student learning. *Portal: Libraries and the Academy*, vol. 13 (1), 61-89 <https://doi.org/10.1353/pla.2013.0008>
- Henderson-Sellers, B.; González-Pérez, C. (2005). A comparison of four process metamodels and the creation of a new generic standard. *Information and Software Technology*, vol. 47 (1), 49-65 <https://doi.org/10.1016/j.infsof.2004.06.001>
- Holden, I. (2012). Predictors of Students' Attitudes toward Science Literacy. *Communications in Information Literacy*, vol. 6 (1), 107-123
- Instituto Nacional de Tecnologías Educativas y Formación del Profesorado (INTEF) (2013). *Marco Común de la Competencia Digital Docente v. 2.0*. Plan de Cultura Digital en la escuela: Ministerio de Educación, Cultura y Deporte: Sevilla.
- Jacobson, T. E.; Mackey, T. P. (2013). Proposing a metaliteracy model to redefine information literacy. *Communications in Information Literacy*, vol.7 (2), 84-91.
- Jenkins, H. et al. (2009). *Confronting the Challenges of Participatory Culture. Media Education for the 21st Century*. MIT Press; Cambridge (Mas.)-London
- Kutner, L.; Armstrong, A. (2012). Rethinking Information Literacy in Globalized World. *Communications in Information Literacy*, vol. 6 (1), 24-33.
- Lau, J. (2008). *Directrices sobre desarrollo de habilidades informativas para el aprendizaje permanente*. IFLA. <https://www.ifla.org/files/assets/information-literacy/publications/ifla-guidelines-es.pdf>
- Lee, L.; Chen, D.; Li, J.; Lin, T. (2015). Understanding new media literacy: The development of a measuring instrument. *Computers & Education*, vol. 85, 84-93. <https://doi.org/10.1016/j.compedu.2015.02.006>
- Lemos, A. (2009). Cibercultura como território recombinate. *A Cibercultura e Seu Espelho: Campo De Conhecimento Emergente e Nova Vivência Humana Na Era Da Imersão Interativa*. São Paulo: ABCiber, 38-46.
- Lin, T.; Li, J.; Deng, F.; Lee, L. (2013). Understanding new media literacy: An explorative theoretical framework. *Journal of Educational Technology & Society*, vol. 16 (4), 160-170.
- Machin-Mastromatteo, J.D. (2012). Participatory action research in the age of social media: literacies, affinity spaces and learning. *New Library World*, vol. 113 (11-12), 571-585. <https://doi.org/10.1108/030748012111282939>
- Mackey, T. P.; Jacobson, T. E. (2011). Reframing information literacy as a metaliteracy. *College & Research Libraries*, vol. 72 (1), 62-78. <http://crl.acrl.org/index.php/crl/article/view/16132/17578>
- Mackey, T. P.; Jacobson, T. E. (2014). *Metaliteracy: Reinventing information literacy to empower learners*. Facet; London
- MacMillan, M.; MacKenzie, A. (2012). Strategies for integrating Information Literacy and Academic Literacy. Helping undergraduate students make the most of scholarly articles. *Library Management*, vol. 33 (8-9), 525-535. <https://doi.org/10.1108/01435121211279885>
- Marzal, M.A.(2010). La evaluación de los programas de alfabetización en información en educación superior: estrategias e instrumentos. *Revista de Universidad y Sociedad del Conocimiento (RUSC)*, vol. 7 (2), 28-38
- McBride, M.F. (2012). Reconsidering Information Literacy in the 21st Century: the redesign of an Information Literacy Class. *Journal of Educational Technology Systems*, vol.40 (3), 287-300. <https://doi.org/10.2190/ET.40.3.e>
- Moeller, S.; Joseph, A.; Lau, J.; Carbo, T. (2011). *Towards Media and Information Literacy Indicators*. Background Document of the Expert Meeting, 4-6 November 2010, Bangkok (Thailand). UNESCO; Paris

- Moselen, C.; Wang, L. (2014). Integrating Information Literacy into Academic Curricula: A professional development programme for librarians at the University of Auckland. *The Journal of Academic Librarianship*, vol. 40 (3), 116-123. <https://doi.org/10.1016/j.acalib.2014.02.002>
- Pidd, M. (2009). *Tools for Thinking: Modelling in Management Science*. Wiley; Chichester
- Ponjuán, G.; Pinto, M.; Uribe Tirado, A. (2015). Conceptualización y perspectivas de la alfabetización informacional en iberoamérica: Un estudio delphi. *Information Research*, vol. 20 (3), 1-26.
- Radcliff, S; Wong, E. (2015). Evaluation of sources: a new sustainable approach. *Reference Services Review*, vol. 43 (2), 231 – 250. <https://doi.org/10.1108/RSR-09-2014-0041>
- Rheingold, H. (2010). *Tools for Thought: The History and Future of Mind-Expanding Technology, The Virtual Community: Homesteading on the Electronic Frontier*. MIT Press
- Shenton, A.K.; Hay-Gibson, N.V. (2012). Information behaviour meta-models. *Library Review*, vol. 61 (2), 92-109. <https://doi.org/10.1108/00242531211220735>
- Siemens, G. (2010). *Conociendo el conocimiento*. Traducido por Nodos Ele: Quintana, E.; Vidal, D.; Torres, L.; Castrillejo, V.; Santamaria, F.; Alonso, N. <http://yoprofesor.org/2015/12/06/conociendo-el-conocimiento-george-siemens-descarga-gratuita/> [24/01/2017]
- Szigeti, K.; Wheeler, K. (2011). Essential Readings in E-Science. *Issues in Science and Technology Librarianship*, (64). <https://doi.org/10.5062/F400001J>
- Thomas, A. B.; Hodges, A. R. (2015). Build sustainable collaboration: Developing and assessing metaliteracy across information ecosystems. Paper presented at the *ACRL 2015*, Portland, Oregon. 78-94.
- Tyner, K. (2008). Audiencias, intertextualidad y nueva alfabetización en medios. *Comunicar: Revista Científica de Comunicación y Educación*, vol. 15 (30), 79-85. <https://doi.org/10.3916/c30-2008-01-012>
- UNESCO (2009). *Guide to measuring Information and Communication Technologies, ICT, in Education*. Institute for Statistics: Quebec
- UNESCO (2011). *Towards Media and Information Literacy Indicators*. UNESCO: París
- UNESCO; Athabasca University (2015). *e-Media and Information Literacy Course*. [http://www.unesco.org/new/en/communication-and-information/crosscutting-priorities/gender-and-media/women-make-the-news-2016/register-for-online-mil-course/\[24/01/2017](http://www.unesco.org/new/en/communication-and-information/crosscutting-priorities/gender-and-media/women-make-the-news-2016/register-for-online-mil-course/[24/01/2017)
- Witek, D.; Grettano, T. (2014). Teaching metaliteracy: A new paradigm in action. *Reference Services Review*, vol. 42 (2), 188-208. <https://doi.org/10.1108/RSR-07-2013-0035>

ANEXOS

Anexo I. Indicadores de Competencia en Información

Competencia	Definición	Objetivos
Percibir una necesidad de información	Competencia para determinar cuándo necesita información y cuánta le es necesaria para satisfacer sus necesidades en diferentes contextos: académico, profesional, social, cotidiano, etc.	<ul style="list-style-type: none"> ✓ Percibir que hay cuestiones cotidianas que pueden ser resueltas por el acceso a la información ✓ Reconocer cuándo necesita información para aprender ✓ Tomar conciencia de las necesidades de información para la optimización de su actividad académica ✓ Percibir cuándo necesita información complementaria para entender mejor una cuestión, hipótesis, etc. ✓ Conseguir delimitar la extensión y el formato de la información necesaria en diferentes contextos ✓ Determinar cuál es el eje conductor (información esencial) y cuáles son los contenidos complementarios
Acceder a la información	Competencia para establecer estrategias adecuadas con las que localizar la información necesaria, considerando las diversas fuentes y formatos de información	<ul style="list-style-type: none"> ✓ Reconocer las principales y más idóneas fuentes de información para sus necesidades informativas ✓ Acceder a bancos y bases de datos especializadas para buscar información ✓ Planear el proceso de búsqueda de información, determinando sistemas, estrategias y términos de búsqueda, adecuados para una recuperación de información eficaz ✓ Establecer preferencias y filtros para la recuperación de información ✓ Reconocer las propiedades y caracteres distintivos de los diferentes formatos de información (texto, imagen, sonido, vídeo) ✓ Gestionar adecuadamente y mantener el plan de navegación dentro de un recurso, entre recursos, en la Web ✓ Comprender el funcionamiento específico de cada sistema de recuperación de información ✓ Determinar si la información recuperada es pertinente, adecuada y suficiente, revisando la estrategia de búsqueda si fuera necesario

Evaluar información	Competencia para comprender, analizar e interpretar críticamente la información, según unos criterios claros y adecuados	<ul style="list-style-type: none"> ✓ Comprender la información recuperada ✓ Discernir opiniones y hechos contrastados ✓ Interpretar con eficacia contenidos en distintos contextos ✓ Identificar el propósito de la información ✓ Adoptar criterios adecuados para seleccionar el contenido pertinente ✓ Identificar la autoridad de la fuente ✓ Comparar contenidos de distintas fuentes para evaluar la certidumbre en los hechos ✓ Analizar críticamente las informaciones obtenidas ✓ Evaluar la confiabilidad, corrección y veracidad de la información ✓ Examinar las relaciones entre informaciones
Organización de la información	Competencia para gestionar la información y contenidos recuperados, organizándolos y asignándolos los metadatos suficientes para su representación	<ul style="list-style-type: none"> ✓ Registrar las fuentes o referencias de la información-contenido seleccionados ✓ Establecer un mapa semántico y un mapa conceptual entre los contenidos, para fijar un mapa semántico de relaciones y clústeres ✓ Integrar los contenidos en esquemas mentales para su comprensión y apropiación ✓ Sintetizar la información y el contenido recuperados ✓ Describir y etiquetar con eficacia la semántica de los contenidos para una posterior recuperación o reutilización ✓ Tomar conciencia de la inestabilidad de la información en Web y, por tanto, ser capaces de establecer un sistema para su preservación ✓ Apostillar y ampliar los contenidos editados por otros
Producción de Contenidos	Competencia para editar contenidos en distintos formatos, aprovechando, reutilizando y referenciando el conocimiento existente	<ul style="list-style-type: none"> ✓ Crear contenido original ✓ Reaprovechar y reutilizar contenidos, generado por otros para nuevas producciones ✓ Respetar los principios de propiedad intelectual, con referencias y uso ético de información ✓ Establecer enlaces entre informaciones dispersas (creación de hipertextos e hipermedia) ✓ Desarrollar contenidos en colaboración ✓ Desarrollar una perspectiva de análisis original y propia ✓ Utilizar con competencia diferentes formatos de edición para mejor representar conceptos y expresar ideas ✓ Disponer y editar los contenidos en espacios colaborativos

Anexo II. Indicadores de Competencia en Comunicación

Competencia	Definición	Objetivos
Establecer y mantener la comunicación	Competencia en establecer y mantener una comunicación eficaz y provechosa con personas y entidades	<ul style="list-style-type: none"> ✓ Identificar cuál es el medio, canal o instrumento más adecuado para transmitir un mensaje dependiendo del contenido y contexto ✓ Disfrutar de aplicaciones y programas de comunicación para establecer interacción con personas ya conocidas o con nuevos contactos, identificando la más adecuada para la comunicación ✓ Comprender y hacerse comprender en los mensajes de comunicación ✓ Conseguir la atención ajena ✓ Expresarse con claridad y objetividad, considerando los hábitos de recepción de la audiencia ✓ Adecuar el lenguaje/código de comunicación/formato al receptor ✓ Integrar más formatos de representación que el textual, para comunicar conceptos e ideas ✓ Considerar el contexto (social, religioso, cultural, económico) del otro ✓ Contextualizar las informaciones y contenidos recibidos, evaluando la perspectiva del emisor ✓ Adaptar el mejor modo de representación de una argumentación en la comunicación (prezzi, ppt, etc. entre otras muchas formas) ✓ Crear un espacio propio para la comunicación digital con otros (blog, redes, etc.) ✓ Entender cómo se organiza la información y contenidos dentro del contexto del tópico de una comunidad virtual o una disciplina ✓ Identificar la potencial audiencia de la propia comunicación y comunicarla de forma específica ✓ Propiciar medios para que los otros puedan responder a la propia comunicación

Crear lazos sociales	Competencia en el desarrollo de redes sociales	<ul style="list-style-type: none"> ✓ Construir una identidad digital para participar en grupos ✓ Reconocer cómo la visibilidad, reputación y privacidad en internet se convierten en factores claves de la identidad digital ✓ Participar en comunicaciones y redes sociales para interactuar con los otros y establecer relaciones sociales en ambiente digital ✓ Compartir informaciones útiles según los intereses y necesidades del interlocutor ✓ Reconocer y respetar las normas de comportamiento para la interacción en ambiente virtual ✓ Argumentar y defender las propias ideas y convicciones de manera clara y respetuosa en ambiente virtual, respetando la opinión ajena para llegar a la resolución de problemas ✓ Reconocer la diversidad cultural (multi e interculturalidad), respetando el uso y el valor de la comunicación para los distintos grupos humanos ✓ Racionalizar y fundamentar los puntos de vista de una argumentación en un debate ✓ Practicar la lectura social y colaborativa ✓ Comunicarse de forma escrita según los parámetros establecidos en los medios formales de comunicación científica ✓ Cumplir con los aspectos formales de un proyecto de investigación en ambientes y espacios digitales de especialistas
Construir conocimiento en colaboración	Competencia para generar conocimiento colaborativo	<ul style="list-style-type: none"> ✓ Reconocer la importancia de cooperar en proyectos de aprendizaje colaborativo ✓ Demostrar la disposición para participar con otros en la experiencia del aprendizaje ✓ Identificar las dinámicas y también espacios y herramientas para favorecer la colaboración ✓ Comprender la dinámica discursiva y argumentativa de una comunidad ✓ Establecer alianzas de trabajo y movilizar redes sociales para lograr ayuda ✓ Participar interactiva y críticamente en foros, comenzando un debate relevante, promoviendo la resolución de un problema o proponiendo un modelo o marco conceptual de interpretación, colaborado en la finalización de tareas en plazo ✓ Participar en la edición de documentos colaborativos ✓ Colaborar, cooperar y trabajar en la formulación de hipótesis y en desarrollo de proyectos que requieran la participación y asesoría de otros especialistas ✓ Cumplir las tareas individuales, encomendadas en el equipo ✓ Reconocer y utilizar las potencialidades de las TIC en el proceso colaborativo (Drive, Dropbox, etc.) ✓ Desarrollar una estrategia de uso eficaz de Web 2.0 y redes sociales en ambientes colaborativos de comunicación ✓ Actuar como eficaz intermediario entre una información pertinente y sus potenciales usuarios interesados
Evaluar la Comunicación	Competencia para autoevaluar las propias competencias en información y conectar con las competencias de los otros	<ul style="list-style-type: none"> ✓ Evaluar los aspectos de privacidad y seguridad antes de publicar información, más si es acerca de terceros ✓ Comprender la propia responsabilidad legal en las publicaciones ✓ Evaluar el interés del receptor antes de reenviar comunicaciones indiscriminadamente ✓ Prestar atención a aspectos ortográficos, gramaticales y de edición en los textos, atender a variedades lingüísticas y registros idiomáticos distintos, como también discernir los mensajes indeseables ✓ Autoevaluar la propia capacidad de comunicación ✓ Prestar atención a los riesgos de la comunicación en ambiente digital ✓ Considerar el impacto de la divulgación de una comunicación