



## Observatorio

En la era de la web de los datos: primero datos abiertos, después datos masivos

Tony Hernández-Pérez

Gestión del conocimiento: orígenes y evolución

Mario Pérez-Montoro

## Artículos

Consumption of public institutions' open data by Spanish citizens  
M. Gértrudix, M. C. Gertrudis-Casado and S. Álvarez-García

Social sciences and humanities research data and metadata: A perspective from thematic data repositories

Nancy D. Gómez, Eva Méndez and Tony Hernández-Pérez

Grado de cumplimiento de las *Leyes de transparencia, acceso y buen gobierno* y de reutilización de los datos de contratación de la Administración central española

Pilar Beltrán-Orenes y Esther Martínez-Pastor

Data management in audiovisual business: *Netflix* as a case study  
E. P. Fernández-Manzano, E. Neira and J. Clares-Gavilán

Civic participation and interactive documentaries: A contribution to the open government model

Laura Cortés-Selva and Marta Pérez-Escolar

Análisis computacional de las características ortotipográficas y paralingüísticas de los tweets periodísticos

Esteban Vázquez-Cano, Javier Fombona y Carlos Bernal

*Joupreneur*: An original methodology for raising entrepreneurial intentions among journalism students

Manuel Goyanes y Marcello Serra

Situación de los centros de documentación en los medios escritos andaluces

Natalia Meléndez-Malavé y Rocío Hirschfeld-Suárez

Prevalencia y capacidad de reconocimiento del plagio académico entre el alumnado del área de economía

Rubén Comas-Forgas y Jaume Sureda-Negre

## Análisis

Técnicas *big data*: análisis de textos a gran escala para la investigación científica y periodística

C. Arcila-Calderón, E. Barbosa-Caro y F. Cabezuelo-Lorenzo

Cómo analizar el impacto de los datos de investigación con métricas: modelos y servicios

A. Ferrer, E. A. Sánchez, R. Alexandre y F. Peset

Uso de datos en la creatividad publicitaria: el caso de *Art, Copy & Code* de *Google*

David Selva-Ruiz y Lucía Caro-Castaño

Análisis de la recomendación entre iguales en la reputación online de las organizaciones

Lucía Benítez-Eyzaquirre

Tecnologías para la incorporación de objetos 3D en libros de papel y libros digitales

C. Carbonell, J. L. Saorín, C. Meier, D. Melián y J. De-la-Torre

## Software

Diseño de un agregador para la gestión de los *big data* informativos

Manuel Blázquez-Ochando

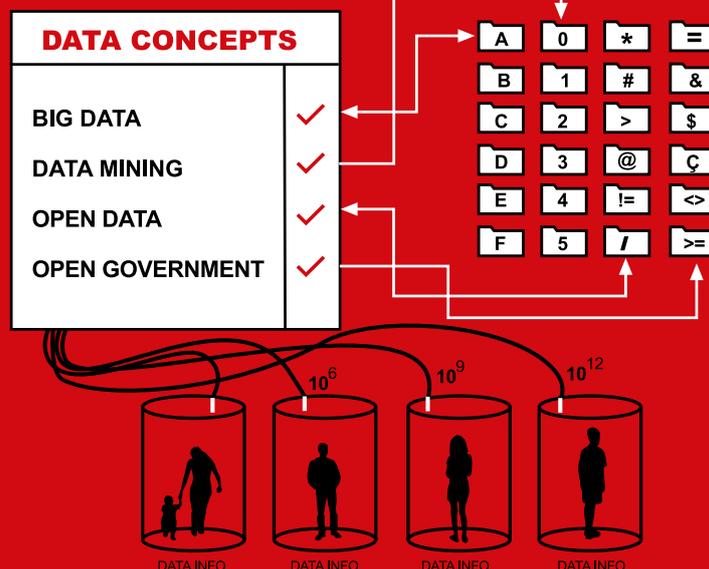
## Indicadores

Propuesta de rankings de universidades españolas en redes sociales

Carmen Zarco, Salvador Del-Barrio y Óscar Cordón



2016, julio-agosto, v. 25, n. 4, pp. 513-710  
eISSN 1699-2407  
CODEN: PINFF2



# DATOS

03:18 AM - READING DATA



# El profesional de la información

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

El profesional de la información es una revista de  
la editorial EPI SCP

Apartado 32.280 - 08080 Barcelona, España

Tel.: +34 - 609 352 954

<http://www.elprofesionaldelainformacion.com>

## Redacción

El profesional de la información

Apartado 32.280

08080 Barcelona, España

Tel.: +34 - 934 250 029

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

## Publicidad

Tel.: +34 - 608 491 521

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

## Suscripciones

El profesional de la información

Apartado 32.280

08080 Barcelona, España

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

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

## Servicios online

María T. Moreno: [mt.moreno@ono.com](mailto:mt.moreno@ono.com)

Roxana Dinu: [roxanadinu@gmail.com](mailto:roxanadinu@gmail.com)

## Diseño

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

Director artístico: Moisés Mañas

## Maquetación

Isabel Olea

[isabel.iolea@gmail.com](mailto:isabel.iolea@gmail.com)

## Distribución online

Recyt. Fundación Española para la Ciencia y la  
Tecnología

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

Depósito legal: B. 12.303-1997

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

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

## DIRECCIÓN EDITORIAL

Tomàs Baiget

EPI SCP

<http://www.baiget.com>

## SUBDIRECTOR

Javier Guallar

Univ. de Barcelona / Univ. Ramon Llull

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

## COORDINACIÓN EDITORIAL

Carlos Tejada-Artigas

Universidad Complutense de Madrid

[tejada@ccdoc.ucm.es](mailto:tejada@ccdoc.ucm.es)

Luis Rodríguez-Yunta

CCHS, CSIC, Madrid

[luis.ryunta@cchs.csic.es](mailto:luis.ryunta@cchs.csic.es)

## REDACTORA JEFE

Isabel Olea

[isabel.iolea@gmail.com](mailto:isabel.iolea@gmail.com)

## REDACCIÓN

Natalia Arroyo-Vázquez

Universidad Isabel I

[natalia.arroyo@gmail.com](mailto:natalia.arroyo@gmail.com)

Ricardo Eíto-Brun

Grupo GMV

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

Javier Leiva-Aguilera

Catorze.com

<http://www.javierleiva.info>

Toon Lowette

Grid Electronic Publishing

[toon@grid.be](mailto:toon@grid.be)

Roser Lozano

CRAI Universitat Rovira i Virgili

[roser.lozano@urv.cat](mailto:roser.lozano@urv.cat)

Jorge Serrano-Cobos

MASmedios

[jorgeserrano@gmail.com](mailto:jorgeserrano@gmail.com)

## REVISIÓN DE LENGUA INGLESA

Laurie Bridges

Oregon State University Libraries and Press

[laurie.bridges@oregonstate.edu](mailto:laurie.bridges@oregonstate.edu)

## COMITÉ CIENTÍFICO

Ernest Abadal

Universitat de Barcelona, Barcelona.

Isidro F. Aguillo

Centro de CC Humanas y Sociales, CSIC, Madrid.

Adela d'Alòs-Moner

Consultora, Barcelona.

Carlos B. Amat

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

Jesús Bustamante

Biblioteca, Cedefop, Salónica, Grecia.

Carlota Bustelo-Ruesta

Consultora, Madrid.

Atilio Bustos

Pontificia Univ. Católica de Valparaíso, Chile.

Lluís Codina

Universitat Pompeu Fabra, Barcelona.

Emilio Delgado-López-Cózar

Universidad de Granada, Granada.

Javier Díaz-Noci

Universitat Pompeu Fabra, Barcelona.

Assumpció Estivill

Universitat de Barcelona, Barcelona.

Antonia Ferrer-Sapena

Universidad Politécnica de Valencia, Valencia.

António Fidalgo

Universidade da Beira Interior, Portugal.

Francisco-Javier García-Marco

Universidad de Zaragoza, Zaragoza.

Johannes Keizer

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

Jesús Lau

Universidad Veracruzana, Veracruz, México.

Mari-Carmen Marcos

Universitat Pompeu Fabra, Barcelona.

Pere Masip

Blanquerna, Univ. Ramon Llull, Barcelona.

Charles McCathieNevile

Yandex, Moscow, Rusia.

José-Antonio Moreiro-González

Universidad Carlos III de Madrid.

Enrique Orduña-Malea

Universidad Politécnica de Valencia, Valencia.

Juan-Antonio Pastor

Universidad de Murcia, Murcia.

Fernanda Peset

Universidad Politécnica de Valencia, Valencia.

Joan Roca

Minnesota State University, Mankato, USA.

Ramón Salaverria

Universidad de Navarra, Pamplona.

Tomás Saorin

Universidad de Murcia, Murcia.

Robert Seal

Loyola Univ. Chicago, Evanston, Illinois, USA.

Ernesto Spinak

Consultor, Montevideo, Uruguay.

Emir-José Suaidén

IBICT, Brasília, Brasil.

Daniel Torres-Salinas

Universidad de Navarra, Pamplona.

Jesús Tramullas

Universidad de Zaragoza, Zaragoza.

## BASES DE DATOS

### Academic search premier (Ebsco)

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

### Catálogo de Latindex

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

### Communication source (Ebsco)

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

### Dialnet (Universidad de La Rioja)

[http://dialnet.unirioja.es/servlet/revista?&clave\\_revista=469](http://dialnet.unirioja.es/servlet/revista?&clave_revista=469)

### ERIH Plus

<https://dbh.nsd.uib.no/publiseringsskanaler/erihplus/periodical/info.action?id=449601>

### ISI Social sciences citation index, Social SCI, WoS (Thomson Reuters)

[http://wokinfo.com/products\\_tools/multidisciplinary/webofscience/ssci](http://wokinfo.com/products_tools/multidisciplinary/webofscience/ssci)  
Impact Factor 2015 = 0,710

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

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

### ISOC, Índice español de ciencias sociales y humanidades

<http://bddoc.csic.es:8080/ver/isoc.do>

### ISTA, Information science and technology abstracts (Ebsco)

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

### Lisa, Library and information science abstracts (CSA)

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

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

<http://www.libraryresearch.com>

### Scopus (Elsevier) Scimago Journal Rank 2015 = 0,422

<http://www.scopus.com>

## ACCESO A LOS TEXTOS COMPLETOS

### Recyt (2000-)

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

### Digitalia Publishing (2011-)

<http://www.digitaliapublishing.com>

### Ebscohost Electronic Journals Service (2000- )

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

### e-Libro (2011-)

<http://www.e-libro.net>

### Library, information science & technology abstracts with full text (2000- )

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

### Library, information science & technology abstracts with full text, LISTA (2000- )

<http://www.libraryresearch.com>

### Academic search complete (2000- )

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

### Business source complete (2000- )

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

### El profesional de la información (1992-embargo 3 años)

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

## EPI EN REDES SOCIALES

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

[http://twitter.com/revista\\_EPI](http://twitter.com/revista_EPI)

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

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

## PLATAFORMA DE PRODUCCIÓN OJS

### Recyt, Repositorio español de ciencia y tecnología (Fecyt)

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

## Tema central: Datos

### OBSERVATORIO

- 517 **En la era de la web de los datos: primero datos abiertos, después datos masivos**  
Tony Hernández-Pérez
- 526 **Gestión del conocimiento: orígenes y evolución**  
Mario Pérez-Montoro

### ARTÍCULOS

- 535 **Consumption of public institutions' opendata by Spanish citizens**  
Manuel Gértrudix, María-Carmen Gertrudis-Casado and Sergio Álvarez-García
- 545 **Social sciences and humanities research data and metadata: A perspective from thematic data repositories**  
Nancy-Diana Gómez, Eva Méndez and Tony Hernández-Pérez
- 557 **Grado de cumplimiento de las Leyes de transparencia, acceso y buen gobierno y de reutilización de los datos de contratación de la Administración central española**  
Pilar Beltrán-Orenes y Esther Martínez-Pastor
- 568 **Data management in audiovisual business: Netflix as a case study**  
Eva-Patricia Fernández-Manzano, Elena Neira y Judith Clares-Gavilán
- 578 **Civic participation and interactive documentaries: A contribution to the open government model**  
Laura Cortés-Selva and Marta Pérez-Escolar
- 588 **Análisis computacional de las características ortotipográficas y paralingüísticas de los tweets periodísticos**  
Esteban Vázquez-Cano, Javier Fombona y César Bernal
- 599 **Joupreneur: An original methodology for raising entrepreneurial intentions among journalism students**  
Manuel Goyanes y Marcello Serra
- 606 **Situación de los centros de documentación en los medios escritos andaluces**  
Natalia Meléndez-Malavé y Rocío Hirschfeld-Suárez
- 616 **Prevalencia y capacidad de reconocimiento del plagio académico entre el alumnado del área de economía**  
Rubén Comas-Forgas y Jaume Sureda-Negre

### ANÁLISIS

- 623 **Técnicas big data: análisis de textos a gran escala para la investigación científica y periodística**  
Carlos Arcila-Calderón, Eduar Barbosa-Caro y Francisco Cabezuelo-Lorenzo
- 632 **Cómo analizar el impacto de los datos de investigación con métricas: modelos y servicios**  
Antonia Ferrer-Sapena, Enrique-Alfonso Sánchez-Pérez, Rafael Aleixandre-Benavent y Fernanda Peset
- 642 **Uso de datos en la creatividad publicitaria: el caso de Art, Copy & Code de Google**  
David Selva-Ruiz y Lucía Caro-Castaño
- 652 **Análisis de la recomendación entre iguales en la reputación online de las organizaciones**  
Lucía Benítez-Eyzaguirre
- 661 **Tecnologías para la incorporación de objetos 3D en libros de papel y libros digitales**  
Carlos Carbonell-Carrera, José-Luis Saorín, Cecile Meier, Dámari Melián-Díaz y Jorge De-la-Torre-Cantero

### SOFTWARE

- 671 **Diseño de un agregador para la gestión de los big data informativos**  
Manuel Blázquez-Ochando

### INDICADORES

- 684 **Propuesta de rankings de universidades españolas en redes sociales**  
Carmen Zarco, Salvador Del-Barrio-García y Óscar Cordón

### EVALUADORES (2014-2015)

### 703 AGENDA

### 710 INFORMACIÓN PARA LOS AUTORES

### **Asóciate y disfruta de estas ventajas:**

- Regalo de dos libros cada año.
- Matrícula gratuita en cursos y otras actividades formativas.
- Descuento en suscripciones a revistas del sector.
- Participación en sorteos y promociones.
- Información sobre las actividades de la Fundación.
- Desgravación fiscal.

... Y, SOBRE TODO, LA SATISFACCIÓN DE MEJORAR NUESTRA SOCIEDAD A TRAVÉS DEL FOMENTO DE LA LECTURA.

### **¿Quieres colaborar?**

Hazte socio/a y recibirás dos libros de regalo cada año

(cuota mínima: 25 euros al año)

### **¿Quiénes somos?**

Una entidad sin ánimo de lucro creada en el año 2000 con sede en Málaga, España (C/Donoso Cortés, 6), con CIF G92215540, e inscrita en el Registro del Ministerio de Cultura con el n. 440.

### **¿Qué hacemos?**

- Fomento de la lectura con menores hospitalizados.
- Cooperación internacional.
- Español para inmigrantes.
- Actividades de formación e investigación.
- Revista "Mi Biblioteca".
- MiniBibliotecas Alonso Quijano.
- Recursos sobre lectura en la web.

**Me gusta leer**

Puedes inscribirte por teléfono 952 23 54 05 o a través de nuestra web: [www.alonsoquijano.org](http://www.alonsoquijano.org)

## EN LA ERA DE LA WEB DE LOS DATOS: PRIMERO DATOS ABIERTOS, DESPUÉS DATOS MASIVOS

In the age of the web of data: first open data,  
then big data

**Tony Hernández-Pérez**



**Tony Hernández-Pérez** es doctor en ciencias de la información y profesor del *Departamento de Biblioteconomía y Documentación* de la *Universidad Carlos III de Madrid* en donde dirige el programa de doctorado en documentación. Su labor docente e investigadora está ligada al grupo *TecnoDoc* incluyendo asignaturas de web social, gestión de contenidos web, metadatos, búsqueda y recuperación de información, e-learning y documentación periodística y audiovisual.  
<http://orcid.org/0000-0001-8404-9247>

*Universidad Carlos III de Madrid*  
*Facultad de Humanidades, Comunicación y Documentación*  
C/ Madrid, 128. 28903 Getafe (Madrid), España  
[tony@bib.uc3m.es](mailto:tony@bib.uc3m.es)

### Resumen

Repaso de los conceptos y las tecnologías asociadas a la evolución de una web de documentos a una web de datos. Papel que las bibliotecas públicas y académicas están jugando o pueden jugar respecto a los datos masivos, los datos abiertos y los datos abiertos vinculados. Se subraya la importancia estratégica que para el futuro de las bibliotecas pueden tener los datos abiertos y especialmente los datos abiertos vinculados.

### Palabras clave

Datos masivos; Datos abiertos; Datos abiertos vinculados; Bibliotecas públicas; Bibliotecas académicas.

### Abstract

Review of the concepts and technologies associated with the transition from a web of documents to a web of data. The role that public and academic libraries are playing or may play about big data, open data, and linked open data is described. The strategic importance of open data and linked open data (LOD) for the future of libraries is emphasized.

### Keywords

Big data; Open data; Linked open data; LOD; Public libraries; Academic libraries.

**Hernández-Pérez, Tony** (2016). "En la era de la web de los datos: primero datos abiertos, después datos masivos". *El profesional de la información*, v. 25, n. 4, pp. 517-525.

<http://dx.doi.org/10.3145/epi.2016.jul.01>

## 1. Introducción

Las bibliotecas y los medios de comunicación viven tiempos de crisis. Se cierran medios de comunicación, se reducen o congelan presupuestos en las bibliotecas y se despide o no se renueva a muchos de sus profesionales. Ocurre lo obvio, como consecuencia de que las bibliotecas han perdido el monopolio de los libros como medio de transmisión de conocimiento y cultura. Y los medios de comunicación tradicionales el de la producción y distribución de noticias. En

el mundo digital en que nos movemos ahora la información se consume principalmente a través de dispositivos móviles e internet (**Negredo; Vara-Miguel; Amoedo**, 2016) y ni la biblioteca ni los medios poseen ya en exclusiva el valor de la información como agente transformador de conocimiento y cultura.

*Google* y las redes sociales se han convertido en las principales puertas de acceso a la información. En este contexto de crisis y de sensación de pérdida de funciones, en el que

un *youtuber* puede tener más seguidores que un medio de comunicación, los periodistas y los bibliotecarios, con cada vez menos recursos, intentan “reinventarse” ampliando el espectro y el formato de noticias a publicar. En el caso de las bibliotecas, ampliando o creando nuevos servicios: apoyo a la publicación en revistas y a la petición de sexenios para investigadores, servicios de repositorios de investigación, apoyo a la alfabetización digital, al e-learning, etc. Son de alabar los esfuerzos por evolucionar, por innovar y casi por sobrevivir.

La Web, además, está pasando de ser un medio en el que publicar y compartir documentos —en el que hasta ahora nos sentíamos más o menos cómodos— a un medio en el que lo que se publica y comparte cada vez más son datos ya que incluso los documentos vienen marcados para poder tratarlos como datos. ¿No es paradójico? Después de pensar que pasábamos de la sociedad de la información a la sociedad del conocimiento regresamos a los orígenes de los procesos de automatización y parece que los datos vuelven a ser esenciales: *big data*, *data mining*, *open data*, *research data*, *linked data*, *open linked data*, *data science*, *data literacy*, *social data*, *data infrastructure*, *data privacy*..., son algunos de los términos frecuentes en la bibliografía científica de casi cualquier campo científico e incluso en los medios no especializados.

No, no se trata de ninguna paradoja sino de la evolución hacia la madurez de las tecnologías de la información: aumento y abaratamiento de la producción en formato digital, aumento de las capacidades personales y tecnológicas, de almacenamiento, de procesamiento, de acceso, de análisis, de enlace y de distribución..., que hace que ahora la pirámide del conocimiento de la que hablaba Russell Ackoff (Rowley, 2007) dato - información - conocimiento se cierre y sea realmente un ciclo. Y de la evolución de la Web como un medio para publicar documentos a un medio para compartir datos estructurados.

Naciones Unidas (2014) habla de una revolución de datos que tiene que ver con la gran cantidad de ellos que se generan y se almacenan cada día, y cuyo procesamiento puede servir como evidencia para la toma de decisiones bien fundamentadas. La OCDE (2015), el gobierno de EUA (Obama, 2009) y la Comisión Europea (2014) conceden la máxima importancia a la gestión de datos de todo tipo y la consideran



Figura 1. Nube de etiquetas relacionadas con los datos

una parte fundamental de la economía digital que regirá el mundo en los próximos años.

La nueva fiebre del oro son los datos. De acuerdo con IBM:

“cada día se crean 2,5 trillones de bytes de datos —supone que el 90% de los datos en el mundo de hoy se ha creado en los últimos dos años. Proceden de todas partes: sensores utilizados para recopilar información sobre el clima, mensajes a sitios de medios sociales, fotos y vídeos digitales que compartimos, registros de transacciones de compra o señales de GPS del teléfono móvil, por citar unos pocos” (IBM, 2016).

Todos estos datos conforman los “datos masivos” o *big data*.

Y en esa carrera por no perder el pie muchos se abrazan a los nuevos términos o conceptos que aparecen, como los citados antes. Uno de esos términos, de moda desde hace unos años, es *big data* (BD). Conste que no hay escepticismo respecto al mismo. No es una moda y está aquí para quedarse, sin duda.

“ No está tan claro que los profesionales de la información, bibliotecarios o comunicadores, tengan que tratar habitualmente con *big data*, más bien no ”

Lo que ya no está tan claro es que los profesionales de la información, bibliotecarios o comunicadores, puedan considerar que los *big data* serán algo con lo que tratarán habitualmente, más bien no. Muy pocos podrán integrarse, ni siquiera a medio plazo, en equipos multidisciplinares que trabajen con datos masivos y habrá que distinguir entre *data scientist*, *data analyst* y otras nuevas denominaciones que seguro irán apareciendo.

## 2. Datos masivos o *big data* ¿Concierne hoy a las grandes bibliotecas?

Los datos masivos son fundamentalmente datos transaccionales, millones de búsquedas en tiempo real, millones de transacciones de compras o millones de posibles combinaciones químicas. Aunque no existe un estándar respecto a cuán grande debe ser un conjunto de datos, para considerarlo datos masivos y sólo a efectos didácticos se puede convenir que:

- Ficheros con datos de hasta 10 GB (decenas de miles de filas y columnas), procesables en *Excel* o *R* y manejables en la memoria de un solo ordenador pueden considerarse “datos pequeños”.
- Ficheros de entre 10 GB y 1 TB (millones de páginas web) ya requieren para su manejo bases de datos algo especializadas y no funcionan bien en *Excel*, pero se pueden almacenar y trabajar desde un disco duro externo del ordenador.
- Ficheros mayores de 1 TB ya se consideran datos masivos, aunque cada vez se tiende a hablar más de petabytes o de exabytes. Para manejarlos normalmente se requieren bases de datos distribuidas y es necesario almacenarlos

en múltiples ordenadores que deben ser, por ejemplo, capaces de gestionar billones de clicks, como los que hacen decenas de miles de personas cuando juegan a *Candy Crush*.

A principios de los años 2000 los datos masivos se relacionaban con un modelo basado en las 3 V's:

- Volumen (en el sentido de que la cantidad de datos de los que se dispone exceden la capacidad de procesamiento de las bases de datos convencionales);
- Velocidad (en el sentido en que el tiempo en que se recopila y se procesa tal volumen de información es muy rápido, a veces incluso instantáneo); y
- Variedad (en el sentido de que las fuentes de las que proceden y los tipos de datos que se procesan pueden ser muy diferentes).

El *National Institute of Standards and Technology (NIST, 2015)* amplía su definición con una cuarta V, Variabilidad (los cambios que se producen en las características de los datos).

Al modelo se le han añadido algunas otras V's:

- Veracidad (en el sentido de que los datos deben ser fiables) o
- Valor (en el sentido de que la transformación de los datos en información y conocimiento es lo que ofrece una ventaja competitiva a las organizaciones que trabajan con estas masas ingentes de datos) pero sólo las primeras cuatro son reconocidas por *NIST*.

Trabajar con este tipo de datos requiere una arquitectura tecnológica especial para un eficiente almacenamiento, manipulación y análisis de los datos, una arquitectura muy vinculada a *Hadoop*<sup>1</sup>, al aprendizaje automático y a una infraestructura con un sistema de ficheros diferente a las que conocemos, con algoritmos que funcionan de forma paralela y distribuida y con bases de datos *NoSQL*<sup>2</sup>.

« Sólo grandes organizaciones en sectores como farmacia, banca, energía o tecnología, entre otras, son capaces de tratar y aprovechar los *big data* »

Más allá de los tecnicismos, en la actualidad sólo las grandes organizaciones en sectores como las farmacéuticas, la banca, grandes empresas de energía o tecnológicas, entre otras, son capaces de permitirse una tecnología tan compleja. Se requiere personal informático muy especializado y con capacidades no sólo informáticas, sino con una alta especialización matemática o estadística, y con competencias analíticas, económicas y de comunicación para no sólo interpretar y dar valor a los datos, sino ser capaz de sintetizarlos y comunicarlos de forma clara, fiable y eficiente.

¿Será necesario para bibliotecarios y profesionales de la información o la comunicación conocer las técnicas para trabajar con datos masivos? No parece que tal circunstancia pueda darse en un horizonte a medio plazo. Los grandes centros de datos en donde se puede trabajar con tecnologías de datos masivos requieren personal con alta cualifica-

ción informática y con un perfil muy multidisciplinar –como hemos dicho- y, sobre todo, una infraestructura tecnológica alejada de lo que conocemos. Ni los profesionales de la información o comunicación han sabido prever la demanda de formación en campos tan especializados y, a la vez, tan multidisciplinarios.

Pero el mundo no son sólo datos masivos o *big data*. Existen muchos conjuntos de datos a los que se puede sacar valor y en donde los profesionales de la información y de la comunicación quizá tengan un gran papel que jugar: los datos abiertos.

### 3. Datos abiertos

Por datos abiertos se entiende

“datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen” (**Dietrich; Gray; McNamara; Poikola; Pollock; Tait; Zijlstra**, n.d.).

Abierto no significa gratis, sino a un coste razonable o proporcional a su valor. Reutilizables significa que deben estar disponibles en una forma conveniente para poder agregarlos a otros conjuntos de datos. Y redistribuibles significa que dichos datos deben ser provistos de licencias o términos de acuerdo que permitan usarlos, sin otras restricciones comerciales o de ningún otro tipo.

Desde el punto de vista técnico, Tim Berners-Lee añadió a sus notas técnicas sobre *linked data* un sistema de marcas basado en estrellas (**Berners-Lee, 2009**) para concienciar sobre la importancia de que los datos sean realmente abiertos y, a ser posible, vinculados:

- 1 estrella: si los datos están disponibles en la Web, en cualquier formato, y con una licencia abierta;
- 2 estrellas: cuando además los datos se ofrecen en un formato estructurado, por ejemplo, en *Excel* en vez de imágenes o pdfs;
- 3 estrellas: como dos estrellas, pero además los datos se ofrecen en un formato no propietario, por ejemplo, ficheros de datos separados por comas (CSV) en vez de *Excel*;
- 4 estrellas: cuando además de todo lo anterior se usan estándares (RDF, y Sparql) para poder identificar cosas y para que otros puedan apuntar hacia ellas; y
- 5 estrellas: los datos vinculados, todo lo anterior y además cuando los datos se enlazan con otros para proporcionar contexto.

Los datos abiertos actuales generalmente son públicos, generados por las administraciones públicas. Por eso muchas veces se les confunde con gubernamentales o como parte del llamado e-gobierno o gobierno electrónico. Que sean públicos no significa que sean muy abiertos, de hecho, muchos documentos y datos gubernamentales en nuestro país son abiertos, pero sólo tienen 1 estrella (mucho pdf, poco reutilizables). Pueden existir también datos abiertos generados por empresas, pero las entidades privadas temen perder cierto control sobre su negocio liberando datos que les afectan, por lo que el grueso

de los datos abiertos se puede decir que son datos generados por organismos públicos.

*Open Knowledge International*, antes *Open Knowledge Foundation (OKFN, n.d)* distingue 7 tipos de datos abiertos:

- 1) Culturales: datos sobre trabajos culturales y objetos, por ejemplo, títulos y autores- generalmente recopilados por GLAM (galerías, bibliotecas, archivos y museos);
- 2) Científicos: se producen como parte de la investigación científica, desde la astronomía a la zoología.
- 3) Finanzas: sobre cuentas públicas (gastos e ingresos) e información sobre mercados financieros.
- 4) Estadísticos: producidos por las oficinas estadísticas, como datos del censo o indicadores socioeconómicos clave.
- 5) Meteorológicos: los muchos tipos de información utilizada para comprender y predecir el tiempo y el clima.
- 6) Medioambientales: sobre el medio ambiente natural, como presencia y niveles de polución, calidad de ríos y mares...
- 7) Transporte: horarios, rutas, estadísticas de puntualidad, etc.

Cuando se habla de la ventaja de los datos abiertos a menudo se mencionan tres tipos de beneficios (**Hernández-Pérez; García-Moreno, 2013**):

- Transparencia para el buen funcionamiento de las sociedades democráticas, puesto que permite conocer qué hacen los gobiernos. Como ejemplo de transparencia se puede mencionar el proyecto de la *Fundación Civio* "¿Dónde van mis impuestos?" (*Civio, n.d.*) en donde, a partir de los datos de los *Presupuestos generales del Estado* (por cierto, extraídos a partir de ficheros de datos abiertos no muy reutilizables) el ciudadano puede visualizar información sobre gastos e ingresos del Estado.
- Aporte de valor comercial y social, puesto que permite la creación de negocios y servicios innovadores basados en esos datos, como las aplicaciones de muchas ciudades españolas para ofrecernos información sobre si se debe esperar un autobús o la demora hace que convenga más tomar otro medio de transporte. O las aplicaciones con información sobre el tiempo, la calidad del aire, o de las aguas en las playas.
- Participación y compromiso de los ciudadanos puesto que al estar más informados pueden implicarse más en los procesos de toma de decisiones, hacer sugerencias sobre presupuestos participativos o dar ideas sobre el uso que se puede asignar a terrenos sin usar de una comunidad autónoma.

No se trata, por tanto, de trabajar con datos masivos sino de aprender a trabajar con datos abiertos y, preferiblemente, vinculados.

#### 4. Las bibliotecas universitarias y la gestión de datos científicos

El mundo de los datos ha irrumpido con fuerza en las bibliotecas académicas desde que la *OCDE* (2007) publicara un documento en el que instaba a los líderes políticos y científi-

cos a impulsar políticas para facilitar el acceso a los datos de investigaciones financiadas con fondos públicos. Las razones esgrimidas en el informe aludían a que los retos de la humanidad son cada vez más globales (salud, cambio climático, etc.) y que había que aprovechar las nuevas tecnologías para lograr mejorar el retorno de la inversión que hacen las instituciones públicas cuando financian la investigación.

Desde entonces las agencias de financiación de la investigación más importantes del mundo han ido tomando iniciativas para incrementar la puesta a disposición del público de los datos de investigación. Existe un amplio consenso (**Hossain; Dwivedi; Rana, 2016**) en que la transparencia y la reproducibilidad de las investigaciones son clave para una ciencia abierta y de excelencia. Transparencia en este contexto significa que la metodología publicada en los resultados de investigación ha sido la más adecuada y se aplicó de forma correcta. La reproducibilidad se refiere a la capacidad de los investigadores de aprovechar los datos para confirmar o refutar los resultados, con el mismo conjunto de datos u otros diferentes, agregados o no, que permitan generar nuevos resultados.

La liberación de estos datos abiertos de investigación beneficia a la institución y a la sociedad en general porque:

- acelera la investigación, puesto que se puede ahorrar tiempo en comparar y contrastar datos y resultados, y ayuda a reducir duplicidades o investigaciones redundantes;
- ahorra costes, puesto que grupos pequeños pueden aprovechar los datos de grupos más grandes o viceversa, se pueden agregar datos pequeños; y
- porque comparar y contrastar los datos también supone un ahorro económico y de tiempo. Permite detectar fraudes y que se inicien o se reconsideren líneas de investigación basadas en datos erróneos.

“ La liberación de los datos abiertos de investigación beneficia a la institución y a la sociedad porque acelera la investigación, ahorra costes y permite detectar fraudes y reconsiderar líneas de investigación ”

Cualquiera de los informes de prospectiva que tomemos señala la gestión de los datos científicos como uno de los temas que más interés suscita en el mundo de las bibliotecas académicas: la *Association of College & Research Libraries (ACRL, 2014)* o el informe *Horizon (Johnson; Adams-Becker; Estrada; Freeman, 2015)* señalan la gestión de datos científicos como una de las tendencias que más van a crecer en los próximos tres años. Probablemente recaiga sobre las bibliotecas responder a las presiones de las agencias de financiación sobre los investigadores para que publiquen sus datos de investigación. La UE, por ejemplo, acaba de aprobar una comunicación (*Comisión Europea, 2016*) en la que declara que, a partir de 2017, todos los datos científicos producidos en el marco del *Programa Horizon 2020* deberán estar abiertos por defecto.

Tabla 1. Cuestiones pendientes para las bibliotecas universitarias

<ul style="list-style-type: none"> <li>- Realizar guías y protocolos para implementar planes de gestión de datos, como los ya publicados por el <i>Consortio Madroño</i> (Madroño, n.d.)</li> <li>- Orientar y formar a los investigadores sobre la conveniencia de formatos y estándares a utilizar para poder preservar sus datos a largo plazo.</li> <li>- Revisar sus infraestructuras y sus procesos y decidir si conviene almacenar en repositorios institucionales o en repositorios de cada disciplina, los datos generados en el seno de su institución.</li> <li>- Ayudar a los investigadores a que sus datos sean visibles y puedan ser citados.</li> </ul>	<ul style="list-style-type: none"> <li>- Orientar y formar a los investigadores sobre dónde y cómo poder encontrar planes de gestión o datos de investigación aplicables a sus investigaciones.</li> <li>- Orientar y formar a los investigadores sobre dónde depositar (qué repositorios) y cómo (esquemas de metadatos a utilizar en función del área de conocimiento) para que sus datos sean descubribles.</li> <li>- Establecer políticas que aseguren que los investigadores cumplen con las obligaciones y requisitos (anonimización de los datos, fechas de liberación de datos, etc.) que establecen las agencias de financiación y las autoridades académicas.</li> <li>- Enlazar los datos de investigación con otros registros académicos (tesis, artículos u otros objetos digitales):</li> </ul>
---	--

A las bibliotecas universitarias les tocará, sin duda, resolver cuestiones como las indicadas en la tabla 1.

**Pasquetto et al.** (2015) resumen muy bien los ocho ejes principales alrededor de los cuales gira el interés respecto a los datos de investigación (figura 2), que pasamos a comentar:

Las definiciones sobre *open data* (OD) se refieren al concepto de “abierto” en el contexto del mundo de la investigación y de las políticas que deben implementarse para considerar que los datos sean técnica y legalmente abiertos.

Las fuentes de los datos abiertos de investigación (*open research data*) son los estudios que distinguen este tipo de datos de otros como los gubernamentales, y describen las iniciativas (repositorios, revistas, etc.) que tienen lugar en los distintos campos.

Al hacer accesibles los datos de investigación se obtienen beneficios económicos, educativos (para formar a nuevos científicos), sociales –como el apoyo que puede suponer para países menos desarrollados-, y se impulsa la ciencia hecha por los ciudadanos, además de la mencionada, transparencia, reproducibilidad y otros principios relacionados con la ciencia abierta.

Escala de compartición de datos se refiere a: quiénes, cuánto y cómo se comparten, puesto que desde el punto de vista internacional se trata de una práctica que aún está en su infancia. No hay incentivos para publicar datos, aunque ahora se trabaja en el reconocimiento de sus citas (data citation).

Los aspectos sobre la propiedad y las licencias se refieren a la ética y a la privacidad. Los datos y los metadatos pueden revelar información personal muy sensible. Y la propiedad de los datos puede ser un foco de problemas cuando existen conflictos entre la industria privada y la administración pública por la propiedad de los mismos.

Ventajas e inconvenientes de los distintos medios que se utilizan para publicar los datos. Algunos repositorios comerciales garantizan el acceso, pero no la preservación a largo plazo. Los repositorios institucionales son más difíciles de gestionar debido a la diversidad de las áreas de conocimiento, a la falta de cultura entre investigadores. Algunas revistas obligan a depositar en repositorios concretos; otras permiten almacenar los datos en los propios servidores de los grupos de investigación.

El acceso técnico a los datos aborda el tema de las relaciones entre los conjuntos de datos (*datasets*) y otros productos de investigación, desde los resultados a los programas que se han utilizado; desde las bases de datos generadas a partir de la recolección de datos a la metodología que se ha empleado para el análisis o la visualización de dichos datos.

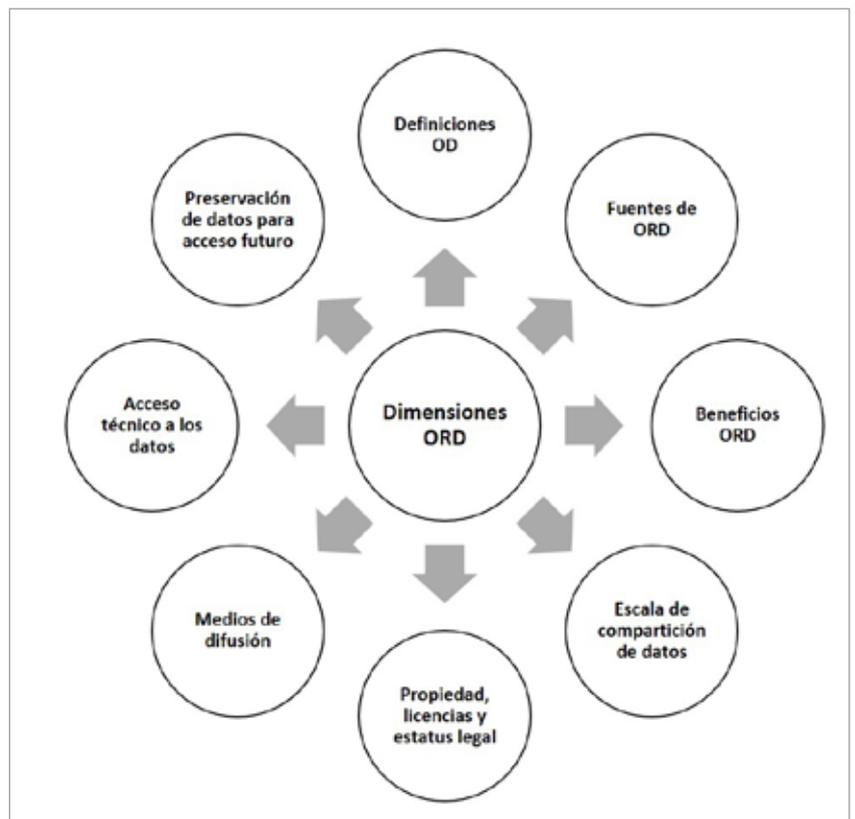


Figura 2. Ejes principales de investigación sobre datos científicos. Basado en **Pasquetto et al.** (2015). OD = *Open data*, ORD = *Open research data*

Y sobre la reutilización de los mismos y su capacidad de estar vinculados.

Por último, la preservación de los datos es algo que preocupa especialmente a archiveros y bibliotecarios, quién tendrá la responsabilidad de la gestión de los datos, de su almacenamiento -si se almacenan en repositorios externos-, de su curación, acceso o mantenimiento sostenible. Además del problema de la propia tarea de describir con metadatos estos *datasets*.

### 5. Bibliotecas públicas y datos abiertos

También las bibliotecas públicas tienen un papel que jugar con los datos abiertos, aunque no sea con el tipo específico de datos de investigación. Las bibliotecas no dejan de innovar para intentar seguir atrayendo a su comunidad de usuarios a su espacio. Un reto difícil considerando que la mayor parte de la población satisface muchas de las necesidades que antes suplía la biblioteca desde prácticamente cualquier sitio con acceso a internet.

Estamos viviendo cómo las bibliotecas se transforman arquitectónicamente, quitando espacio a las estanterías de libros y liberándolos para ordenadores, impresoras 3D o cómodos espacios en donde realizar otro tipo de actividades: desde la alfabetización digital a los cuentacuentos o a talleres de escritura. Se trata, como decía Barbara Striping, presidenta de la *American Library Association* entre 2013 y 2014, de lograr que las bibliotecas

“transformen su relación con las comunidades y darles el poder para que éstas sean creadoras de información y no solo consumidoras” (De-Boer, 2015).

¿En tiempos de sobreabundancia de información, cómo puede la biblioteca pública contribuir a aportar valor social e implicar a los ciudadanos en los procesos de toma de decisiones de la comunidad de usuarios a los que atiende mediante datos abiertos? Un ejemplo, el *Ayuntamiento de Madrid* publica datos diarios de contaminación acústica desde su web de medio ambiente:

<http://goo.gl/ieOX2N>

Los datos se publican por barrios y por tramo horario (mañana, tarde y noche). En zonas con problemas de contaminación acústica, ¿resulta imaginable que la biblioteca proporcione formación, ayude a descargar los datos y permita hacer estudios más específicos sobre los focos de contami-

nación para demostrar la elevada contaminación acústica en su área de trabajo y proponer al ayuntamiento algunas iniciativas para reducirla?

Quizá hoy no resulte muy imaginable. No se trata de un reto fácil, las bibliotecas han dispuesto durante mucho tiempo de millones de datos que no se han sabido explotar. Con honrosas excepciones, como el proyecto *Library analytics toolkit* (Dulin; Spina, 2013), la mayor parte de las bibliotecas no han sabido explotar el sistema de recomendación que tan de moda ha puesto *Amazon* (“si usted leyó esto quizá le interese esto otro”) u ofrecer a un estudiante de primero de derecho un listado con los títulos de los libros de primer curso de carrera más prestados el año anterior. O mostrar si en los últimos cinco años se compran muchos libros de matemáticas, pero no se presta ninguno. Los datos no han sido, al menos hasta ahora, una gran prioridad.

Existen muchos más ejemplos, datos sobre accidentes con lesiones, con información sobre el lugar, tipo de accidente, horas, días de la semana..., o sugerencias y reclamaciones presentadas por los ciudadanos para proponer mejoras en los servicios o para reclamar o protestar sobre el normal o anormal funcionamiento de los servicios municipales que permitirían demostrar con datos fehacientes que no se hace caso de una sugerencia o reclamación recurrente de los ciudadanos, etc.

El rol de las bibliotecas públicas respecto a los datos abiertos se podría resumir en los aspectos que se indican en la tabla 2.

Ahora sí que resulta inimaginable, existen muchas barreras: falta formación en tratamiento de datos, por eso muchas instituciones académicas se lanzan a reformar o crear nuevos planes de estudio, introduciendo cursos de “ciencia de datos”, visualización, estadística o programación básica. Pero si no somos conscientes de que todos, teóricos y profesionales, tenemos que seguir reciclándonos cada día, la ola que trae *Google* y los nuevos hábitos de búsqueda de información acabará con muchas bibliotecas y centros de documentación.

Si en esencia, el papel de las bibliotecas públicas consiste en proporcionar acceso a información y conocimiento a la comunidad a la que sirve. Si su misión es lograr que los ciudadanos a los que sirven tengan un espacio que contribuya a lograr una sociedad más democrática, con menos desigual-

Tabla 2. Rol de las bibliotecas públicas respecto a los datos abiertos

- Identificar y recopilar los conjuntos de datos ( <i>datasets</i> ) abiertos publicados y que afecten a su comunidad.	- Analizar, visualizar y crear contenidos que pongan en valor los <i>datasets</i> y promuevan la concienciación y la participación ciudadana.
- Planificar e identificar los conjuntos de datos necesarios para comprender y reflejar la realidad y los problemas que afecten a sus comunidades.	- Abogar ante las administraciones pertinentes la recolección y publicación de datos que puedan afectar y ser de interés para su comunidad de usuarios.
- Identificar iniciativas de otras bibliotecas y comunidades sobre uso de datos abiertos que podrían ser de aplicación en la comunidad de usuarios a los que atiende.	- Formar a usuarios en la identificación y tratamiento de datos cuantitativos y en la visualización o narrativa de dichos datos.
- Organizar y estimular eventos ( <i>hackathon</i> , etc.) que ayuden a crear aplicaciones que permitan utilizar estos conjuntos de datos.	- Crear conciencia sobre el valor de los datos y metadatos en el mundo digital y su incidencia en la privacidad y los derechos del individuo.

dades económicas (las de quienes no pueden adquirir libros y otros materiales) y menos desigualdades socio-culturales (las de quienes tienen las habilidades y el conocimiento para tomar decisiones bien fundamentadas), entonces las bibliotecas, además de ofrecer infraestructura y recursos físicos y tecnológicos, deberían intensificar su rol como formadoras en programas de alfabetización: básica, digital, informacional y de datos.

Habrá que esperar aún algunos años para ver cómo van evolucionando las bibliotecas, como espacios físicos y como espacios culturales. De lo que no cabe duda es que más que estanterías con los libros, la biblioteca si quiere sobrevivir tendrá que convertirse en una plataforma (*library as platform*) (Weinberger, 2012) en la que los bibliotecarios no sólo deberán salir más allá de las estanterías. Quizá haya llegado el momento de, además de ir a buscar a los usuarios en tantas redes sociales, salir más allá de los muros de la biblioteca para encontrar a los usuarios en sus comunidades.

Si reconocemos que entre las misiones principales de las bibliotecas están: promover la igualdad, el acceso y la participación abierta entre la ciudadanía, entonces tendremos que abordar las estrategias que sean necesarias para que las bibliotecas puedan ser centros de participación ciudadana. Centros en los que las bibliotecas y los bibliotecarios juegan el papel de conectores, de activistas sociales fortaleciendo las relaciones sociales de su comunidad.

Tendremos que abordar las estrategias que sean necesarias para que las bibliotecas puedan ser centros de participación ciudadana en los que los bibliotecarios juegan el papel de conectores, de activistas sociales

La necesidad de redefinir los espacios en las bibliotecas debe servir como estrategia para además de seguir siendo un lugar perfecto para la lectura y el estudio, un lugar de juegos, un lugar en el que construir cosas (*makerspaces*) y un lugar en el que se generan contenidos de interés para la comunidad, desde apps a visualizaciones o narrativas sobre su entorno. Y los profesionales de la información deben convencerse de que los usuarios no son sólo consumidores, que además se han ido a otras tiendas, sino creadores y ciudadanos que pueden ver en las bibliotecas un espacio comunitario de innovación y desarrollo social.

*The Aspen Institute* (2014), en un interesante informe sobre una nueva visión de las bibliotecas públicas lo resume muy bien: es necesario redefinir el papel de las bibliotecas como instituciones que inspiren el aprendizaje, hagan crecer el capital social de sus poblaciones y sean capaces de crear oportunidades entre la comunidad de usuarios a las que atiende. Entre la lista de estrategias que cita para lograr estos objetivos menciona:

- conectar los recursos de otras agencias o bibliotecas a la biblioteca como plataforma, más que reinventar la rueda o ir siempre sola;

- apoyar fuertemente el rol del aprendizaje para que tenga lugar en los espacios físicos o a través de la biblioteca como plataforma; o
- hacer del acceso a la información gubernamental un modelo para curar datos abiertos.

Y es que: las bibliotecas públicas o cambian o sucumben a que sus espacios sean ocupados.

## 6. *Linked open data* (LOD): datos abiertos vinculados

Si decíamos al principio que cada día la web evoluciona desde una web de documentos hacia una web de datos no podemos dejar de resaltar la importancia que tiene en el mundo de la web semántica, en el que las bibliotecas han jugado un papel crucial, el concepto de datos abiertos vinculados. Como decían Zeinstra y Keller (2011), “los datos pueden estar abiertos, pero no vinculados. Los datos pueden estar vinculados, pero no abiertos”.

Aclaremos antes que a veces se utilizan datos enlazados y datos vinculados como sinónimos. De hecho, cuando se habla de uno de los proyectos más importantes de datos vinculados en España, el de la *Biblioteca Nacional* (*datos.bne.es*) se habla de “Datos enlazados en la BNE”. Para Xavier Ajenjo, la diferencia entre el término enlazado y vinculado está en relación con el tipo de enlace entre dos objetos digitales. En su opinión, que compartimos, si se conectan mediante un URL (*uniform resource locator*, web sintáctica), debería hablarse de datos enlazados; y debería utilizarse datos vinculados

“para aquellos recursos que tengan una relación mediante una URI (*uniform resource identifier*), que es lo que ocurre en la web semántica” (Ajenjo, 2012).

Un URL es un caso particular de URI.

El concepto de datos vinculados hace referencia a que los datos, por ejemplo, las diferentes palabras de un documento web que nombra a personas, están vinculados a través de técnicas de web semántica, especialmente RDF (*resource description framework*). RDF es un conjunto de especificaciones del *World Wide Web Consortium* (W3C) cuyo fin es describir recursos mediante triples o tripletas en forma de expresiones sujeto-predicado-objeto. Ejemplo: La afirmación ‘El cielo tiene color azul’ se puede representar en RDF con una tripleta compuesta por un sujeto (cielo), un predicado (un atributo, color) y un valor (azul).

Así es como se está pasando de la web de documentos, con simples textos y palabras, a la web de datos en la que se describe la función de cada palabra. La tecnología LOD permite identificar dentro de un documento el dato que se refiere a una persona, a un lugar, a un acontecimiento, etc., y además describirlo, representarlo y vincularlo con la información que se tiene de esa persona en otros sitios web, sin importar el idioma o el sitio web o URL físico en el que se encuentre.

Existe mucho publicado sobre datos vinculados a partir de las ideas iniciales de Berners-Lee (2009). Las reglas para trabajar con datos vinculados son básicamente cuatro:

- “1) Utilizar URIs como nombres para las cosas (las URIs buenas para la web semántica son las que no cambian);
- 2) Utilizar URIs http (URIs desreferenciables) para que la gente pueda ver esos nombres;
- 3) Cuando alguien busca una URI, proporcionar información útil usando los estándares apropiados (RDF, Sparql); y
- 4) Incluir enlaces a otras URIs, de tal forma que se pueda recuperar más información” (Méndez; Greenberg, 2012).

Las aplicaciones en el mundo de las bibliotecas son múltiples. **Agenjo y Hernández** (2016) acaban de publicar un resumen de un informe del año 2016 de *Library technology reports* sobre el estado actual de los datos vinculados en bibliotecas, archivos y museos en el mundo. Estos autores, muy activos en *Europeana* y otros proyectos españoles (**Hernández**, 2015) aprovechan la ocasión para, con buen criterio, mencionar algunos de los proyectos españoles realmente *linked open data*, como el ya mencionado de la *BNE* y algunos otros.

Antes ya mencionamos el informe *NMC Horizon, Library edition 2015* (**Johnson; Adams-Becker; Estrada; Freeman**, 2015) para señalar que una de las tendencias que destacaba en el mundo de las bibliotecas académicas era la gestión de datos científicos. Pues bien, otra de estas tendencias es “la web semántica y los datos vinculados” y augura que *Bibframe*, un modelo de datos basado en los principios de datos vinculados, será el próximo estándar para asignar y gestionar metadatos bibliográficos sustituyendo a MARC. Permitirá visibilizar mucho más la información contenida en los catálogos de las bibliotecas. En el mismo informe se afirma que:

“los motores de búsqueda más populares apenas tocan el 10% de internet puesto que el 90% restante son sitios web que no se indexan porque la mayor parte de estos datos están en catálogos de bibliotecas en formatos que no pueden ser buscados o detrás de áreas seguras que no pueden ser accedidas por los robots” (p. 42).

Los datos demuestran que las bibliotecas, nacionales, académicas y públicas, están aumentando la exposición de datos vinculados (**Smith-Yoshimura**, 2016). Se publican fundamentalmente datos bibliográficos, pero también ficheros de autoridades, colecciones digitales, datos geográficos, datos sobre objetos de museos y algunos más. Y aunque existen muchas barreras: alta curva de aprendizaje que supone trabajar con datos vinculados; inconsistencia de los datos con los que a veces se trabaja, o, simplemente, falta de recursos, la tendencia a trabajar con datos vinculados no es una moda pasajera y será del todo necesaria si queremos “competir” para que nuestros recursos sean visibles en la Web.

## 6. Conclusiones

Los ciudadanos han encontrado en *Google* y las redes sociales una alternativa válida para satisfacer sus necesidades de información. Ante esta situación las bibliotecas se están viendo forzadas a redefinir y reorientar sus espacios, sus procesos y sus servicios para adaptarse y poder ofrecer sus recursos de forma virtual, además de seguir prestando sus infraestructuras físicas y tecnológicas para no perder la esencia de servicio cultural que siempre han tenido.

Las bibliotecas necesitan salir de las estanterías y lograr un mayor compromiso social con los ciudadanos a los que atienden. Con cada vez menos recursos, tienen que demostrar el valor que aportan a la sociedad. En la carrera por aumentar y ofrecer nuevos servicios es necesario una mayor implicación con los investigadores y con los ciudadanos. Los datos abiertos pueden ser una oportunidad para ello y eso requiere seleccionar sobre qué servicios se debe poner el acento. Los *big data* están fuera del alcance de la mayor parte de las bibliotecas, públicas y académicas, pero los datos abiertos pueden ser una gran oportunidad para impulsar el papel de formadoras que tradicionalmente han jugado las bibliotecas.

En las bibliotecas académicas, concienciando, formando y gestionando, desde planes de gestión de datos de investigación hasta el descubrimiento y preservación de los mismos, para ayudar a los investigadores a satisfacer la presión de las agencias de financiación para liberar los datos de investigación y contribuir así a mejorar y acelerar una verdadera ciencia abierta. En las bibliotecas públicas, aprovechando las posibilidades que ofrecen los datos abiertos para generar contenidos que atañen muy directamente a sus comunidades y propiciando espacios de innovación y de participación ciudadana. Para todas ellas, exponer sus datos en forma de datos abiertos vinculados será una de las mejores formas de poner en valor sus fondos y hacerse más visible ante la sociedad. Los datos abiertos y especialmente los datos abiertos vinculados entran dentro de la categoría de estos “nuevos” servicios, sin excluir a los demás. Renovarse o morir. No hacer nada no puede ser una opción.

## Nota

1. *Apache Hadoop* es una plataforma de software de código abierto para el almacenamiento y procesamiento distribuidos de grandes conjuntos de datos en clusters de ordenadores. Todos los módulos de *Hadoop* están diseñados en el supuesto de que los fallos de hardware son comunes y deben gestionarse automáticamente dentro de *Hadoop*.

2. *NoSQL* (no SQL o no relacional) es un programa para elaborar bases de datos con un sistema de almacenamiento y recuperación que no se basa en relaciones tabulares como las utilizadas en las bases de datos relacionales. Han existido desde la década de 1960, pero no se las llamó *NoSQL* hasta que a principios del siglo XXI aumentó su popularidad por necesidades de las empresas web 2.0 como *Facebook*, *Google* y *Amazon*. Se utilizan cada vez más en *big data* y aplicaciones web en tiempo real.

## 7. Bibliografía

**Agenjo, Xavier** (2012). “Tómalo con filología: no es lo mismo enlazado que vinculado”. *Archivos de la lista de distribución de IweTel*.

<https://listserv.rediris.es/cgi-bin/wa?A2=IWETEL;a9862cc8.1209D>

**Agenjo, Xavier; Hernández, Francisca** (2016). “El estado de los datos vinculados en bibliotecas en 2015”. *Blok de bid*. <http://www.ub.edu/blokdebid/es/content/el-estado-de-los-datos-vinculados-en-bibliotecas-en-2015>

*The Aspen Institute* (2014). “Rising to the challenge: Re-envisioning public libraries”, *Dialogue on public libraries*, 80 pp.

- <http://csreports.aspeninstitute.org/documents/AspenLibrariesReport.pdf>
- Berners-Lee, Tim** (2009). "Linked data - Design issues". <https://www.w3.org/DesignIssues/LinkedData.html>
- Civio, *Fundación Ciudadana* (n.d). "¿Dónde van mis impuestos?". <http://www.dondevanmisimpuestos.es>
- Comisión Europea* (2014). "Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las regiones. *Hacia una economía de los datos próspera*. COM (2014). 442 final". <http://goo.gl/OVHTCO>
- Comisión Europea* (2016). "Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. European cloud initiative - *Building a competitive data and knowledge economy in Europe*. COM (2016). 178 final". <http://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=COM:2016:178:FIN&rid=2>
- ACRL (2014). "Top trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education". *College & research libraries news*, v. 75, n. 6, pp. 294-302. <http://crln.acrl.org/content/75/6/294>
- De-Boer, Jeroen** (2015). "The business case of FryskLab, Europe's first mobile library FabLab". *Library Hi Tech*, v. 33, n. 4, pp. 505-518. <http://dx.doi.org/10.1108/LHT-06-2015-0059>
- Dietrich, Daniel; Gray, Jonathan; McNamara, Tim; Poikola, Antti; Pollock, Rufus; Tait, Julian; Zijlstra, Ton** (n.d). "¿Qué son los datos abiertos?". *Open data handbook*. <http://opendatahandbook.org/guide/es/what-is-open-data>
- Dulin, Kim; Spina, Carli** (2013). "Library analytics toolkit | Harvard Library Lab". <https://osc.hul.harvard.edu/liblab/projects/library-analytics-toolkit>
- Hernández, Francisca** (2015). "Los cambios en *Europeana data model* 5.2.5 y 5.2.6". *Blok de bid*. <http://www.ub.edu/blokdebid/es/content/los-cambios-en-europeana-data-model-525-y-526>
- Hernández-Pérez, Tony; García-Moreno, María-Antonia** (2013). "Datos abiertos y repositorios de datos: nuevo reto para los bibliotecarios". *El profesional de la información*, v. 22, n. 3, pp. 259-263. <http://dx.doi.org/10.3145/epi.2013.may.10>
- Hossain, Mohammad A.; Dwivedi, Yogesh K.; Rana, Nripendra P.** (2016). "State-of-the-art in open data research: Insights from existing literature and a research agenda". *Journal of organizational computing and electronic commerce*, v. 26, n. 1-2, pp. 14-40. <http://dx.doi.org/10.1080/10919392.2015.1124007>
- IBM (2016). "What is big data?". <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>
- Johnson, Larry; Adams-Becker, Samantha; Estrada, Victoria; Freeman, Alex** (2015). *NMC Horizon report: 2015 Library edition*, 60 pp. <http://cdn.nmc.org/media/2015-nmc-horizon-report-library-EN.pdf>
- Madroño, Consorcio* (n.d). *Pagoda: PIA de Gestión de Datos del Consorcio Madroño*. <http://www.consorcioadrono.es/pagoda>
- Méndez, Eva; Greenberg, Jane** (2012). "Datos enlazados para vocabularios abiertos: marco global de HIVE". *El profesional de la información*, v. 21, n. 3, pp. 236-244. <http://dx.doi.org/10.3145/epi.2012.may.03>
- Naciones Unidas* (2014). *A world that counts: Mobilising the data revolution for sustainable development*, 32 pp.. <http://goo.gl/94ehcc>
- Negredo, Samuel; Vara-Miguel, Alfonso; Amoedo, Avelino** (2016). *Digital news report .es 2016: Cambios decisivos en el consumo de noticias digitales*, 84 pp. <http://www.digitalnewsreport.es>
- NIST (2015). *NIST big data interoperability framework: Vol. 1, Definitions*. Big Data Public Working Group, Definitions and Taxonomies Subgroup. <http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf>
- Obama, Barack** (2009). "Transparency and open government". *The White House*. <https://www.whitehouse.gov/node/698>
- OECD (2007). *OECD Principles and guidelines for access to research data from public funding*. OECD. <http://goo.gl/Rh2Bhb>
- OECD (2015). *Data-driven innovation*, OECD Publishing, ISBN: 978 92 64 22934 1 <http://goo.gl/pt5cWe>
- OKFN (n.d). *Open knowledge: What is open?* <https://okfn.org/opendata>
- Pasquetto, Irene V.; Sands, Ashley E.; Borgman, Christine L.** (2015). "Exploring openness in data and science: What is "open," to whom, when, and why?". *Procs of the Association for Information Science and Technology*, v. 52, n. 1, pp. 1-2. <http://dx.doi.org/10.1002/ptra.2015.1450520100141>
- Rowley, Jennifer** (2007). "The wisdom hierarchy: representations of the DIKW hierarchy". *Journal of information science*, v. 33, n. 2, pp. 163-180. <http://dx.doi.org/10.1177/0165551506070706>
- Smith-Yoshimura, Karen** (2016). "Analysis of international linked data survey for implementers". *D-Lib magazine*, v. 22, n. 7/8. <http://dx.doi.org/10.1045/july2016-smith-yoshimura>
- Weinberger, David** (2012). "Library as platform". *Library journal*. Sept. 4<sup>th</sup> <http://lj.libraryjournal.com/2012/09/future-of-libraries/by-david-weinberger>
- Zeinstra, Maarten; Keller, Paul** (2011). *Open linked data and Europeana*. <http://goo.gl/aKYRCc>

# GESTIÓN DEL CONOCIMIENTO: ORÍGENES Y EVOLUCIÓN

## Knowledge management: origins and evolution

Mario Pérez-Montoro



**Mario Pérez-Montoro** es doctor en filosofía y ciencias de la educación por la *Universitat de Barcelona* y posgraduado en organización de sistemas de documentación por la *Universitat Politècnica de Catalunya*. Ha realizado estudios de postgrado en el *Istituto di Discipline della Comunicazione* de la *Università di Bologna* (Italia), y ha sido profesor visitante del *CSLI (Center for the Study of Language and Information)* de la *Stanford University* (California, EUA) y de la *School of Information* de la *UC Berkeley* (California, EUA). Investiga sobre arquitectura y visualización de la información. Ha publicado los libros: *The phenomenon of information* (Scarecrow Press, 2007), *Arquitectura de la información en entornos web* (Trea, 2010), y junto a Lluís Codina, *Navigation design and SEO for content-intensive websites: A guide for an efficient digital communication* (Elsevier Chandos, 2016), entre otros. Es profesor del *Departamento de Biblioteconomía, Documentación y Comunicación Audiovisual* de la *Universitat de Barcelona*.

<http://orcid.org/0000-0003-2426-8119>

*Universitat de Barcelona, Departament de Biblioteconomía, Documentación y Comunicación Audiovisual*  
Melcior de Palau, 140. 08014 Barcelona, España  
[perez-montoro@ub.edu](mailto:perez-montoro@ub.edu)

### Resumen

Desde sus orígenes, la disciplina de la gestión del conocimiento ha experimentado una evolución particular, arrancando con fuerza en un primer momento, presentando un sólido desarrollo posterior, pero viviendo actualmente un lánguido estancamiento. En este trabajo se analiza ese itinerario. Para cubrir este objetivo presentamos el contexto fundacional de la disciplina, para, seguidamente, revisar la evolución de la misma a través de varios indicadores como el interés que despierta en las últimas décadas y la producción científica anual (monografías y artículos de revistas indizadas). También se describe el ecosistema de la formación universitaria centrada en la gestión del conocimiento.

### Palabras clave

Gestión del conocimiento; Orígenes; Evolución; Interés general; Producción científica; Formación universitaria.

### Abstract

Since its origins, the discipline of knowledge management has had a unique evolution: going strong at first, followed by solid development, but currently falling into languid stagnation. In this paper we analyze this trajectory by describing the founding context of the discipline and reviewing its evolution using two indicators, the heightened interest in recent decades and the annual scientific production (monographs and articles in indexed journals). Finally, the ecosystem of university training that is focused on knowledge management is described.

### Keywords

Knowledge management; Origins; Evolution; General interest; Scientific production; University education.

**Pérez-Montoro, Mario** (2016). "Gestión del conocimiento: orígenes y evolución". *El profesional de la información*, v. 25, n. 4, pp. 526-534.

<http://dx.doi.org/10.3145/epi.2016.jul.02>

## 1. Introducción

Corrían los años finales de la década de los 90 del siglo pasado cuando unos académicos universitarios publicaron en forma de coautoría dos obras (**Nonaka; Takeuchi**, 1995; **Davenport; Prusak**, 1998) que cimentarían los conceptos principales de una nueva disciplina emergente en ese momento: la gestión del conocimiento. Ya desde sus inicios se

entendió como una disciplina encargada de diseñar e implementar sistemas cuyo objetivo es identificar, capturar y compartir sistemáticamente el conocimiento involucrado dentro de una organización de forma que pueda ser convertido en valor para la misma.

Han pasado ya casi 20 años desde ese momento bibliográfico fundacional y la disciplina ha podido experimentar una

evolución particular, arrancando con fuerza en un primer momento, presentando un sólido desarrollo posterior, pero viviendo actualmente un lánguido estancamiento.

En este trabajo analizamos ese itinerario. Primeramente presentamos el contexto fundacional de la disciplina, para, a continuación, revisar su evolución a través de dos indicadores:

- interés que despierta en las últimas décadas;
- producción científica anual (monografías y artículos de revistas indizadas).

Finalmente se describe el ecosistema de la formación universitaria centrada en gestión del conocimiento.

La gestión del conocimiento es la disciplina encargada de diseñar e implementar sistemas con el objetivo de identificar, capturar y compartir el conocimiento de una organización de forma que pueda ser convertido en valor para la misma

## 2. Orígenes

A partir de la Segunda Guerra Mundial algunas empresas comienzan a desarrollar una serie de estrategias para mejorar su producción basadas en el aprendizaje a partir de la experiencia. Tienen su máxima expresión teórica en la obra del economista **Kenneth Arrow** (1962). Algunos autores (**Prusak**, 2001; **Senge**, 1990) han visto en estas estrategias un claro antecedente intelectual del movimiento de la gestión del conocimiento.

Sin embargo, este movimiento se extiende y se consolida de forma definitiva en la segunda mitad de los 90. No existe un solo factor que pueda justificar por sí mismo el escenario que ha generado el nacimiento de esta disciplina. Pero existen dos variables, una tecnológica y otra con un carácter más directamente económico, que pueden justificar ese escenario:

- han ido apareciendo una serie de tecnologías de la información y de la comunicación que han permitido el acceso, la gestión y el uso intensivo de información y conocimiento hasta niveles nunca antes conocidos;
- se ha consolidado un nuevo entorno económico de globalización de los mercados y una nueva cultura de la competitividad (**Gorey**; **Dobat**, 1996). En ese escenario, el conocimiento y los activos intangibles —no sólo los materiales— de las empresas comienzan a entenderse como ese valor añadido que puede garantizar el correcto funcionamiento y la supervivencia de las mismas en los mercados globalizados (**Stewart**, 1997; **Leonard-Barton**, 1995; **Sveiby**, 1997).

En ese contexto fundacional, cristalizaron dos escuelas que darían lugar a las dos maneras de entender la disciplina de la gestión del conocimiento:

1) Se entiende el conocimiento como un proceso, como lo denomina **Karl Sveiby** (2001). Es la escuela oriental (o ja-

ponesa) —derivada de las ideas defendidas, entre otros, por **Nonaka** y **Takeuchi** (1995), por citar algunas cabezas visibles— y centra sus investigaciones en el conocimiento tácito y su creación. Se nutre de disciplinas teóricas como la psicología, sociología y pedagogía. Dentro de esta escuela o *knowledge-process group*, se concibe la empresa no de una manera mecanicista sino como un organismo vivo que interacciona con el entorno. En este contexto, el conocimiento se entiende más como un proceso psicológico, relacionándolo con la experiencia. Se defiende la posibilidad del conocimiento y su socialización como operaciones centrales para la gestión del mismo (**Nonaka**, 1991; **Nonaka**; **Konno**, 1998). Dentro de este enfoque, las acciones encaminadas a la gestión del conocimiento suelen realizarse desde el departamento de recursos humanos.

2) La otra escuela entiende el conocimiento como un objeto, como lo denomina también **Karl Sveiby** (2001). Se trata de la escuela más occidental —derivada de las ideas defendidas por **Davenport** y **Prusak** (1998), por citar sólo un par de autores— y centra sus investigaciones en el conocimiento explícito y la gestión del mismo; nutriéndose de disciplinas teóricas como la teoría y la gestión de la información (**Davenport**, 1997). Dentro de esta segunda escuela o *knowledge-object group* se identifica la empresa como un mecanismo de procesar información. En este contexto, el conocimiento se entiende como una producción obtenida a partir de la gestión adecuada de la información y el valor del mismo se consigue cuando se logra transformar en conocimiento explícito representado en documentos que pueden ser tratados y gestionados. Dentro de este enfoque alternativo, las acciones encaminadas a la gestión del conocimiento suelen realizarse desde el departamento de sistemas y gestión de información.

Al comienzo de la gestión del conocimiento se distinguieron dos escuelas: la oriental, que trataba el conocimiento como proceso, y la occidental, que lo consideraba un objeto obtenido a partir de la información

Como cabía esperar, en la evolución de la disciplina, la tensión entre las dos escuelas se ha ido disipando y en el campo de la consultoría se combinan simultáneamente, y en una misma estrategia, las ventajas que ofrecen los dos enfoques implementando soluciones que van desde los recursos humanos a la gestión de la información, pasando por los desarrollos tecnológicos.

## 3. Evolución

Tras su consolidación en la década de los noventa, la disciplina ha presentado una evolución muy similar a otras propuestas científicas y de gestión introducidas en el pasado y que estuvieron de moda durante un tiempo. Ejemplos fueron la auditoría de la información, la vigilancia tecnológica, la inteligencia competitiva o la ecología de la información. De esta manera, tras un primer momento de explosión, la

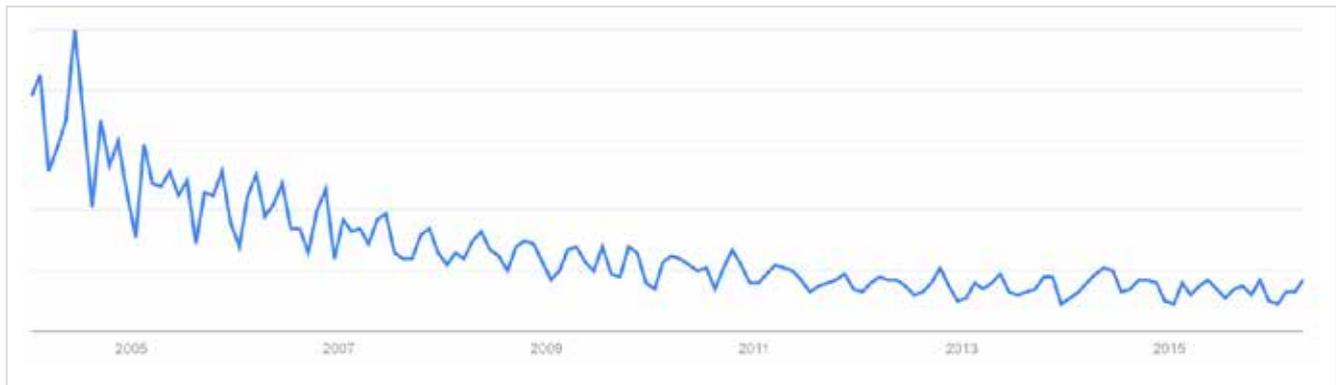


Figura 1. Gráfica asociada con la expresión “knowledge management” obtenida en *Google Trends* (junio 2016)

disciplina ha disfrutado de un período de expansión, para entrar más tarde en un estadio de paulatino descenso del interés generado.

Podemos encontrar varios indicadores que permiten dibujar ese escenario histórico expresado. Principalmente, indicadores obtenidos de la utilización de una serie de herramientas y conjuntos de datos que nos permiten visualizar los patrones evolutivos de la atracción despertada por un tema o disciplina concreta presentada a lo largo de un período de tiempo concreto.

Para recoger la evolución de esa atracción vamos a analizar cómo ha ido variando el interés general despertado por la disciplina, la producción científica anual (monografías y artículos de revistas indizadas) generada sobre la misma y la propuesta universitaria dirigida a formar a los profesionales en la gestión del conocimiento.

### 3.1. Interés general

Estudiar el interés por un tema determinado no es una tarea sencilla. Apostar por métodos de investigación participativos como la entrevista o el cuestionario puede ofrecer interesantes resultados, pero los costes y el tiempo requerido pueden hacer inviable el estudio.

Sin embargo, podemos apostar por métodos no participativos, más sencillos e indirectos. Así, por ejemplo, cuando alguien está interesado en un tema concreto acude a un buscador para informarse. Por tanto, podemos estudiar la evolución cuantitativa de las consultas realizadas en un buscador como un indicador de cómo ha evolucionado el interés general de la población por ese tema.

*Google* es, sin duda, el buscador más utilizado en el mundo para localizar información en internet. Dentro de las herramientas asociadas a ese buscador, la empresa ofrece *Google Trends*.

<https://www.google.com/trends>

*Google Trends* es un servicio gratuito online de *Google* que permite visualizar la popularidad o interés de búsqueda en internet de varias palabras y frases (hasta 5 términos de forma simultánea) a lo largo del tiempo (desde 2004 hasta la actualidad) respecto a una escala de 0 a 100. El principio que subyace es bien simple: cuantas más veces se escriban las palabras relacionadas con un tema en la caja de búsqueda

de *Google*, mayor interés suscita el tema expresado por esa ecuación de búsqueda. Esta visualización nos ofrece la posibilidad de identificar los patrones y cambios de interés por el tema producidos en ese período o apreciar cómo varía según la zona geográfica. También suministra un mapa de calor geográfico que muestra gráficamente el índice del volumen de búsquedas en las distintas regiones de cada país y una lista de las palabras relacionadas con cada término buscado que presentan un mayor crecimiento.

Realizando la búsqueda con el término compuesto “knowledge management” en *Google Trends* se obtiene la gráfica recogida en la figura 1. Según esta gráfica, el interés sobre esta disciplina presenta una línea dentada (o en forma de sierra) que muestra momentos de tiempo en los que éste aumenta alternados de manera cíclica con otros en los que disminuye. Sin embargo, en términos absolutos, la tendencia es que, tras un período de aumento de interés que culmina en 2004, el tema va perdiendo interés lineal y progresivamente a lo largo de los años hasta alcanzar las cuotas más bajas en la actualidad. De 2004, con un valor de interés por encima de 80 (sobre 100 posibles), ha pasado a presentar un valor de interés en 2016 por debajo de 15 (sobre 100 posibles).

“ Tras un primer momento de explosión, la disciplina ha disfrutado de un período de expansión, para entrar más tarde en un estadio de paulatino descenso del interés generado ”

Podemos visualizar también la evolución del interés por el tema de la gestión del conocimiento comparándolo con el interés que suscita otra disciplina emergente, la de los *big data*, que actualmente se presenta como uno de los temas más candentes tanto en contextos académicos como profesionales. Para realizar esa comparación, en *Google Trends* hemos añadido el segundo término compuesto “big data” (línea roja) a la primera búsqueda realizada con el término compuesto “knowledge management” (línea azul), obteniendo la gráfica recogida en la figura 2. Como se desprende de la gráfica, el interés por los *big data*, de la misma forma que por la gestión del conocimiento, presenta una línea dentada

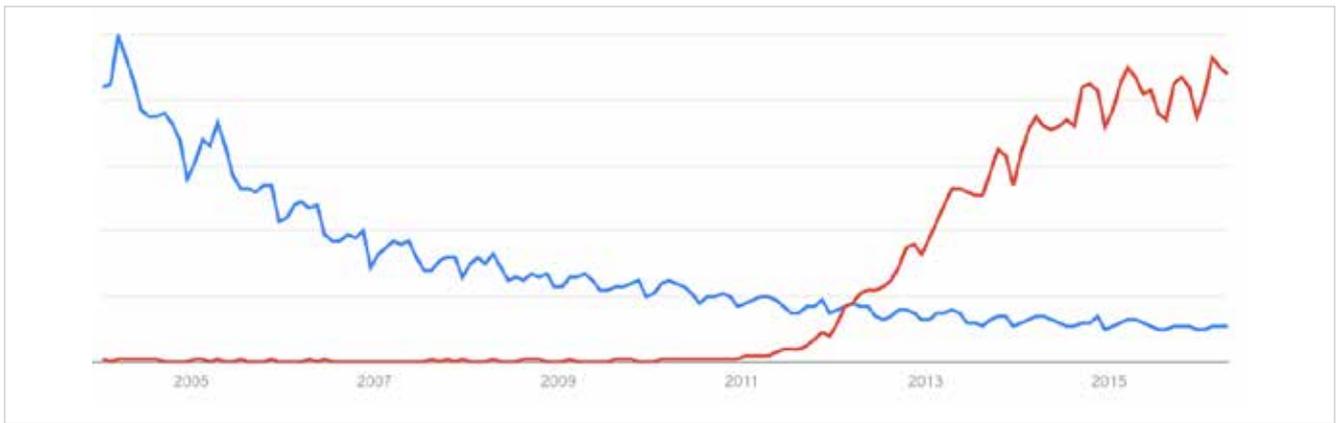


Figura 2. Gráfica asociada con las expresiones “knowledge management” (línea azul) y “big data” (línea roja) obtenida en Google Trends (junio 2016)

(aunque mucho menos escarpada) que muestra momentos de tiempo alternos en los que éste aumenta y otros en los que disminuye. Sin embargo, en términos absolutos, la tendencia es que, a partir de 2011, el interés por este tema se dispara de forma acusada hasta el día de hoy. En este sentido, en sólo 5 años, desde 2011, con un valor de interés cercano a 0 (sobre 100 posibles), ha pasado a presentar un valor de interés en 2016 por encima de 75 (sobre 100 posibles). Destaca también que a partir de 2012 el interés por los *big data* supera el mostrado por la gestión del conocimiento.

Este análisis se puede completar comparando el interés geográfico mostrado por ambas disciplinas. Así, si observamos la figura 3 podemos comprobar que entre los siete primeros países del ranking de interés sobre el tema de la gestión del conocimiento no se encuentra ninguno de los que lideren los avances científicos ni las propuestas innovadoras en el ámbito de la gestión. El ranking de siete países está liderado por Kenia (con un valor de interés de 100 (sobre 100 posibles) y lo cierra la India (con un valor de interés de 34 (sobre 100 posibles)), incluyendo en ese intervalo otros países como Sudáfrica, Sri Lanka, Malasia, Hong Kong y Singapur.

Sin embargo, si generamos la gráfica de los siete primeros países del ranking de interés sobre el tema de los *big data* (figura 4) el resultado es distinto. En este caso entre estos

siete primeros países se encuentran Estados Unidos y Corea del Sur, dos de los países que lideran los avances científico-tecnológicos. En esta gráfica también comprobamos que el ranking está liderado por la India (con un valor de interés de 100 (sobre 100 posibles)) y lo cierra Estados Unidos (con un valor de interés de 28 (sobre 100 posibles), incluyendo en ese intervalo otros países como Singapur, Hong Kong, Corea del Sur, Taiwán y Sri Lanka.

### 3.2. Producción científica

La producción científica sobre un tema puede ser identificada como un fiel indicador del interés que suscita una materia. Sin embargo, no siempre es así. La producción científica puede verse alterada por factores exógenos (líneas de investigación impuestas en convocatorias de financiación o intereses institucionales, por ejemplo, entre otras) que hagan variar el tema y el volumen producido en un período de tiempo concreto. Por estos factores, en momentos determinados se puede hablar de alguna manera de un cierto grado de “burbuja” académica alrededor de un tema. Pensemos, por ejemplo, cómo ha aumentado de forma casi lineal la producción científica sobre el tema de la web semántica en los últimos años y cómo paradójicamente ha disminuido, en cambio, el número de proyectos reales en los que se ha implementado esa tecnología (W3C, 2012).



Figura 3. Gráfica con las regiones que muestran mayor interés frente a la ecuación “knowledge management” obtenida en Google Trends (junio 2016)



Figura 4. Gráfica con las regiones que muestran mayor interés frente a la ecuación “big data” obtenida en Google Trends (junio 2016)

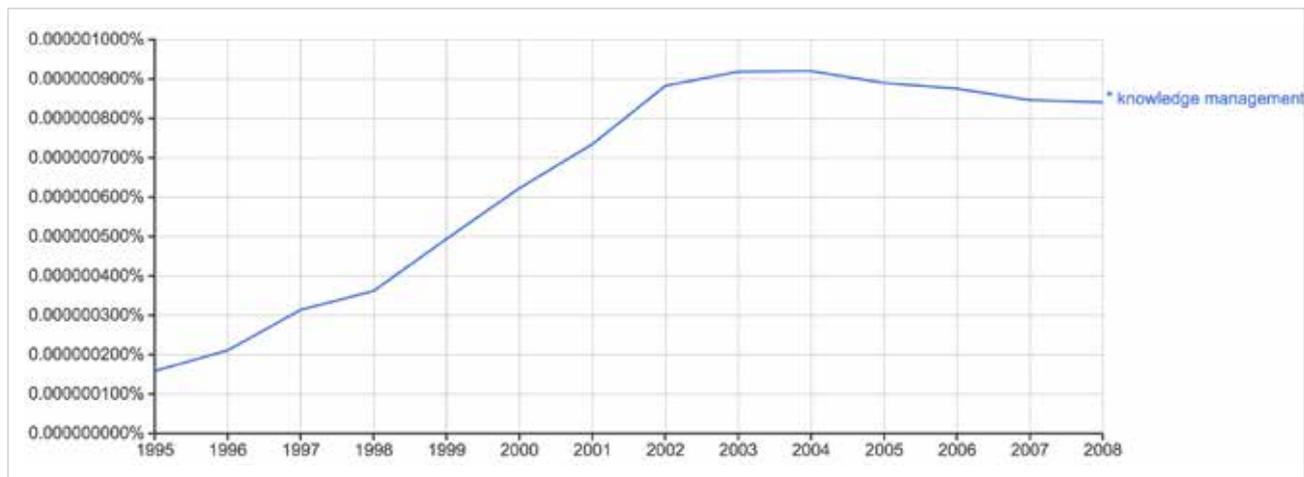


Figura 5. Gráfica asociada con la expresión “knowledge management” dentro del corpus lingüístico anglosajón, entre los años 1995 y 2008, obtenida en Google Ngram Viewer (junio 2016)

Pero dejando al margen esos factores de distorsión, podemos utilizar la variación anual en la producción científica (monografías y artículos) sobre el tema de la gestión del conocimiento como un indicador de la evolución de esta disciplina.

**Producción de monografías**

Para poder analizar de forma cómoda la producción de monografías sobre el tema de la gestión del conocimiento podemos utilizar, de nuevo, otra propuesta de la familia de Google: Google Books Ngram Viewer. <https://books.google.com/ngrams>

Google Books Ngram Viewer es un servicio gratuito online de Google que permite visualizar la evolución de la producción de monografías (a las que Google tiene acceso a partir de su proyecto Google Books) sobre un tema concreto. Para entregarnos esa visualización representa en una gráfica de series temporales la evolución de la frecuencia de los términos de búsqueda (términos separados por comas o n-gramas, de ahí el nombre) en los libros indizados dentro del corpus lingüístico y a lo largo del rango de años elegidos.

Si realizamos una búsqueda con el término compuesto “knowledge management” dentro del corpus lingüístico anglosajón y entre los años 1995 y 2008 (límite superior que puede alcanzar el rango) en Google Ngram Viewer obtenemos la gráfica recogida en la figura 5. En ella podemos observar cómo, a partir de 1995, aumenta de forma casi lineal la producción de monografías que incluyen el término compuesto hasta 2002. De 2002 a 2004, aumenta un poco esa producción alcanzando su máximo valor. Pero es a partir de ese año cuando el número anual de libros sobre el tema comienza a decaer paulatinamente alcanzando su nivel más bajo en 2008 (el sistema no permite obtener esa evolución en los años posteriores), justo por debajo de los datos obtenidos en 2002.

Ese patrón de la evolución del tratamiento del tema en las monografías en lengua inglesa se reproduce también, aunque con un desfase de 2 años, en las monografías que incluyen el término “gestión del conocimiento” dentro del corpus de libros en español (figura 6). En este segundo caso, la evolución casi lineal arranca en 1997 y se extiende hasta 2004. Un año más tarde la producción se estanca para subir levemente hasta 2006. A partir de ese año y hasta 2008

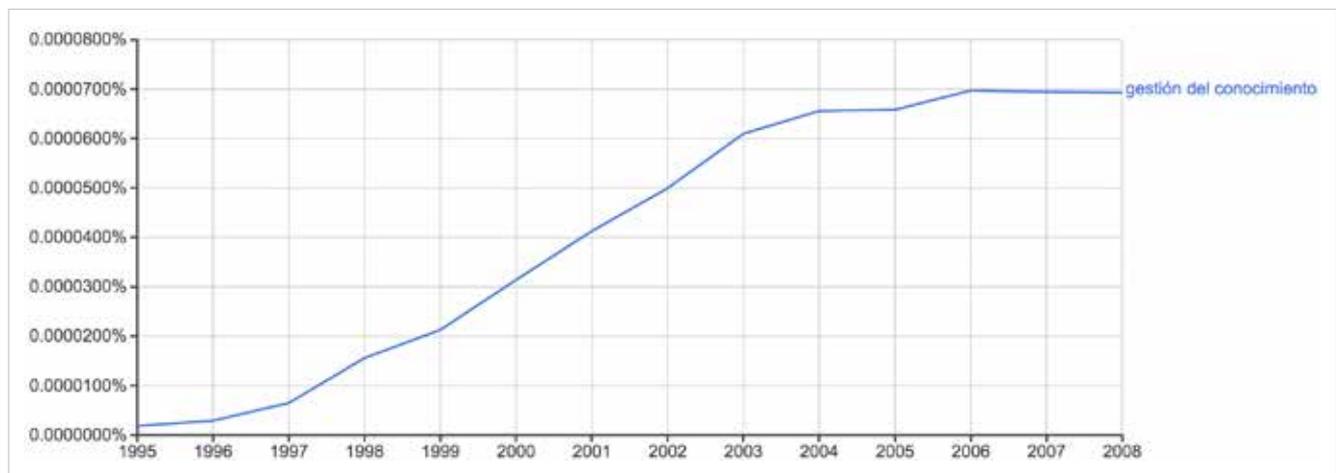


Figura 6. Gráfica asociada con la expresión “gestión del conocimiento” dentro del corpus lingüístico español, entre los años 1995 y 2008, obtenida en Google Ngram Viewer (junio 2016)

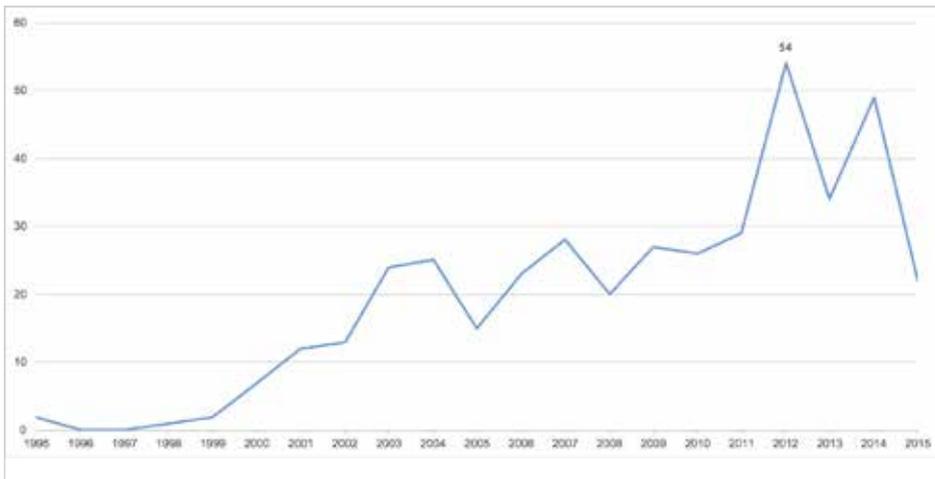


Figura 7. Gráfica de la evolución de monografías sobre gestión del conocimiento publicadas en España durante el período 1995-2015

(el sistema no permite obtener esa evolución en los años posteriores), el patrón apunta a un descenso paulatino de la producción.

La evolución global e internacional de la publicación de monografías dedicada al tema de la gestión del conocimiento (figura 6) contrasta, al menos en parte, con la evolución de esa misma producción centrada exclusivamente en el territorio español. Para realizar esa comparación podemos recurrir a la base de datos del ISBN de libros editados en España, publicada por el *Ministerio de Educación, Cultura y Deporte*.

Si realizamos una búsqueda con el término compuesto “gestión del conocimiento” en esa base de datos y los ordenamos por fecha de publicación descartando el año 2016, obtenemos que desde 1995 hasta finales de 2015 se han publicado en España 429 monografías con un número ISBN asociado. Con ese resultado, podemos construir una gráfica donde se muestra la evolución temporal de la producción anual de monografías dotadas de ese número estándar en España y en ese período de tiempo (figura 7). Según esa gráfica, tras un comienzo titubeante, la publicación de monografías sobre la disciplina en España se dispara en un primer período de 1999 (con la publicación de 2 monografías) a 2004 (con la publicación de 24); un patrón en cierto grado similar al mostrado por la publicación internacional en los corpus en lengua inglesa (figura 5) y en lengua española (figura 6). Sin embargo, y a diferencia de la evolución internacional en las dos lenguas, a continuación se produce un período de aumento (con forma dentada) en la producción anual de volúmenes que culmina en 2012 con la publicación de 54 monografías. Este aumento en el período de 2004 a 2012 puede justificarse por la proliferación de conferencias y congresos relacionados con la disciplina que terminaron generando con sus actas un número importante de monografías con ISBN. Después de 2012, de la misma manera que ocurría en la producción internacional en inglés y español, la publicación en España entra en declive (aunque también con forma dentada) alcanzándose en 2015 una producción de sólo 22 monografías, similar a la que se produjo en 2008.

## Producción de artículos científicos

El análisis de la evolución anual de la producción científica sobre la disciplina no estaría completa si no tratáramos también esa misma evolución en la publicación de artículos por parte de investigadores y académicos.

Una buena estrategia para analizar la evolución anual de la producción de artículos científicos es examinar la producción registrada por las principales bases de datos que indizan las revistas donde esos investigadores publican sus aportaciones científicas.

Entre esas bases de datos destacan, por encima del resto, *Web of Science* y *Scopus*. Estas bases de datos no recogen toda la producción en forma de artículos realizada por los científicos, pero dan cuenta de una parte importante que, según declaran, se ajusta a unos estándares de calidad establecidos.

*Web of Science* es una web de las compañías *Onex* y *Baring Private Equity Asia* que da acceso a varias bases de datos bibliográficas y recursos de análisis de la información que permiten evaluar y analizar el rendimiento de la investigación. No proporciona el texto completo de los documentos indizados, pero sí herramientas de análisis que permitan valorar su calidad científica.

Si realizamos una doble búsqueda con el término compuesto “*knowledge management*”, primero en el campo Tema y después en el campo Título entre los años 1995 y 2015, obtenemos una serie de datos con los que construir la gráfica recogida en la figura 8. En ella podemos observar como a partir de 1995 aumenta suavemente, de forma casi lineal, la producción anual de artículos que versan sobre el tema de la gestión del conocimiento o que incluyen en su título el nombre en inglés de la disciplina hasta el año 2006. En los tres años posteriores, aumenta de forma ostensible el ritmo de esa producción anual alcanzando su máximo valor en 2009: con 731 artículos indizados que incluyen la expresión “*knowledge management*” en su título y con 1.810 artículos indizados que tratan sobre el tema de la gestión del conocimiento. Pero es a partir de ese año cuando el número anual de esos dos tipos de artículos comienza a decaer paulatinamente para, tras un pequeño repunte 2013, desplomarse hasta alcanzar en 2015 valores cercanos a los de 2006.

*Scopus*, por su parte, es la otra gran plataforma de resúmenes y referencias bibliográficas de literatura científica que pertenece a la multinacional *Elsevier*. Incluye más de 55 millones de registros indizados y el área de humanidades y ciencias sociales se encuentra mejor representada que en la base de datos *Web of Science*.

Si realizamos una búsqueda con el término compuesto “knowledge management”, en el campo Título (Title) entre los años 1995 y 2015 en *Scopus*, obtenemos la gráfica de la figura 9, que ofrece un patrón evolutivo muy similar al de la base de datos anterior. En esta nueva figura podemos observar como a partir de 1995 aumenta de forma casi lineal la producción anual de artículos que incluyen en su título el nombre en inglés de la disciplina de la gestión del conocimiento hasta 2003, indizándose ese año 452 documentos en la base de datos. Después de ese período, tras un año (2004) en el que apenas aumenta la producción anual, en 2005 se dobla el número de artículos indizados ese año, alcanzando los 800 documentos. Desde 2005, continúa un crecimiento continuado de la producción anual hasta alcanzar, en 2009 y con 1.325 documentos, el punto más alto en esa progresión. Pero es a partir de ese año, como ocurría en *Web of Science*, cuando el número anual de artículos comienza a desplomarse hasta alcanzar en 2015 (con sólo 863 documentos) valores cercanos a los de 2006 (812 documentos).

Obtenemos resultados muy similares si realizamos la misma búsqueda en esta plataforma, pero esta vez en el campo Palabra clave (Keyword) en lugar del campo Título. Tras esa nueva búsqueda, podemos generar la figura 10. En ella observamos cómo, a partir de 1995, aumenta de forma casi lineal la producción anual de artículos que versan sobre el tema de la gestión del conocimiento hasta 2009, año en el que se indizan 9.223 documentos en esta base de datos. Pero es también a partir de ese año cuando el número anual de artículos decae constantemente hasta alcanzar en 2015 (4.852 documentos producidos) valores similares a los obtenidos en 2006 (4.276 artículos indizados).

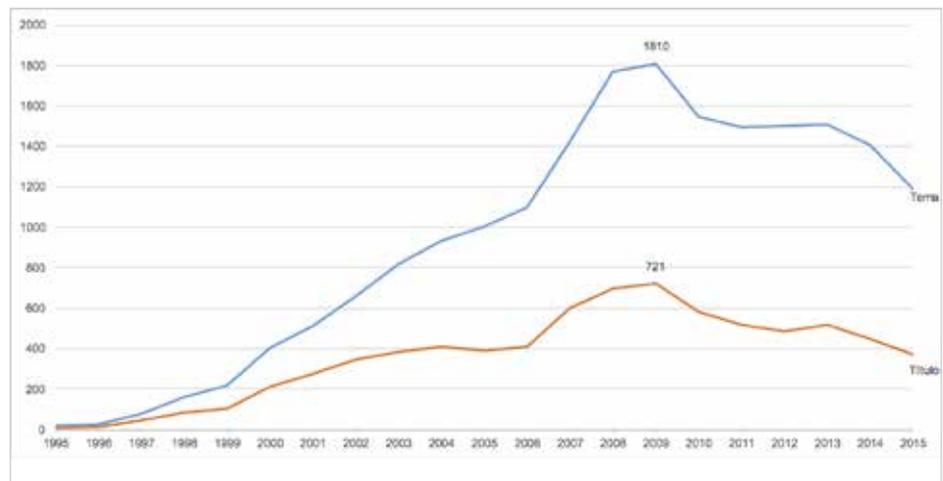


Figura 8. Gráfica obtenida con la búsqueda “knowledge management” (por título y tema) entre los años 1995 y 2015 dentro de la base de datos *Web of Science*

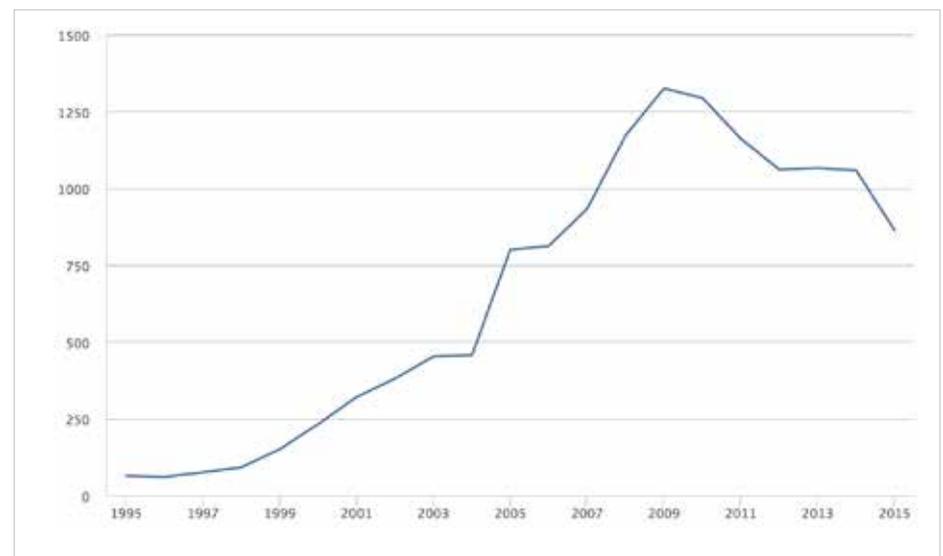


Figura 9. Gráfica obtenida con la búsqueda “knowledge management” (campo Title) entre los años 1995 y 2015 en la base de datos *Scopus*

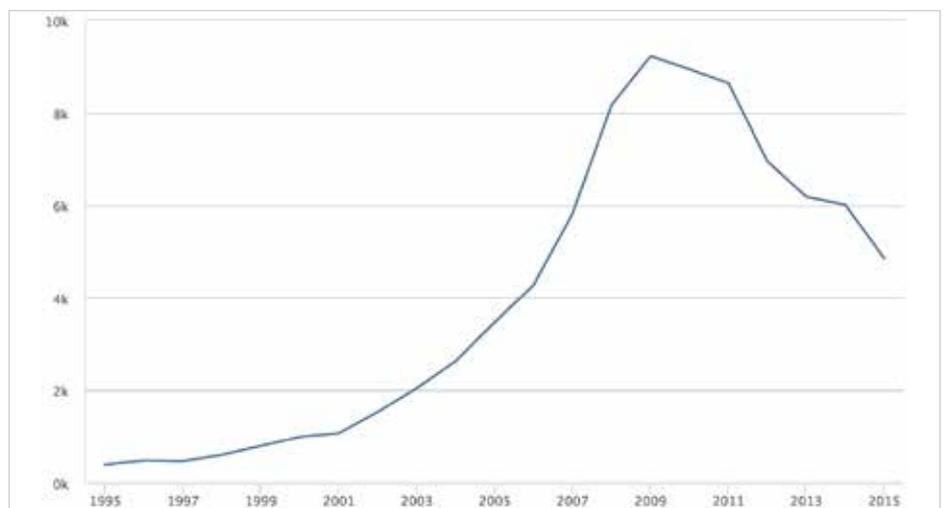


Figura 10. Gráfica obtenida con la búsqueda “knowledge management” (campo Keyword) entre los años 1995 y 2015 en la base de datos *Scopus*

### 3.3. Formación universitaria

Completemos el análisis del escenario actual de la disciplina de la gestión del conocimiento abordando también el tema de su formación universitaria.

El tema de la formación sobre gestión del conocimiento es muy extenso, y casi inabarcable en un modo exhaustivo. En el mundo existen actualmente miles de cursos universitarios (grados, masters, doctorados o extensión universitaria, entre otros) que versan directamente sobre la disciplina o que incluyen nociones sobre ese tipo de gestión en algunas de las competencias que adquirirán sus estudiantes una vez cubierto el curso.

Si nos centramos en los cursos o contenidos sobre gestión del conocimiento impartidos dentro de programas universitarios orientados a la formación de profesionales de la información, podemos comprobar que:

- el 47% de esos cursos se imparte en Norteamérica;
- el 24% en Oceanía (Australia y Nueva Zelanda);
- el 14% se imparten en el continente europeo (**Roknuzzaman; Umemoto, 2009; Pacios-Lozano; Rueda-Martínez, 2016**).

Este contraste de cifras se puede atribuir al mayor grado de control gubernamental existente en Europa respecto al diseño de programas formativos universitarios que suele limitar el catálogo y las competencias de los cursos acreditados.

Pero centrémonos ahora en el ámbito geográfico español. Podemos utilizar recursos online para identificar el volumen y distribución de cursos sobre gestión del conocimiento impartidos actualmente.

Así, por ejemplo, si realizamos en el portal *Universia* (financiado por el *Banco de Santander*) una búsqueda con la expresión “gestión del conocimiento” descubriremos que actualmente se imparten 25.497 cursos relacionados directa o indirectamente con esa disciplina (incluidos también algunos, menos del 1% del total, que forman parte de ciclos superiores de formación profesional). La mayoría se concentran en titulaciones de grado y master de empresariales, marketing, informática o ingeniería; aunque, en la mayoría de esos cursos recuperados, la relación con la disciplina sea muy escasa o tangencial.

<http://cursos.universia.es>

Algo muy parecido ocurre si realizamos la misma búsqueda en otro portal similar como *Educaweb*. En este caso, aunque recuperemos sólo 7.203 cursos, el panorama es similar. La mayoría se concentra en la oferta de master y grado en las áreas señaladas y muchos de ellos mantienen una relación indirecta o sólo testimonial con la disciplina.

<http://www.educaweb.com>

Pero si nos limitamos a identificar la oferta acreditada por el *Ministerio* español, el panorama cambia de forma radical. El portal de títulos de la *Aneca* (*Agencia Nacional de Evaluación de la Calidad y Acreditación*, organismo encargado de la acreditación de las titulaciones universitarias oficiales y reconocidas por el gobierno), tras la misma búsqueda, sólo ofrece 15 resultados (frente a 25.497 y 7.203 de los dos casos anteriores). La diferencia en las cifras de lo recuperado es sustancial. En este caso sólo aparecen 15 programas de master universitario (acreditados por la *Agencia*) centrados

en las mismas áreas destacadas, ninguno de grado, de los cuales 9 son públicos y 6 privados.

<http://www.aneca.es>

En el campo de la formación universitaria española relacionada con la ciencia de la información, el escenario de los cursos sobre gestión del conocimiento es muy escueto. Desde 2003, sólo se han diseñado 7 titulaciones (algunas ya desaparecidas) donde se han programado cursos sobre esta materia. En la mayoría de esos cursos la gestión del conocimiento ha sido una materia optativa dentro de sus programas respectivos. A fecha de hoy, sólo están programadas cuatro asignaturas sobre la disciplina, de las cuales 4 son de grado (2 optativas y 2 obligatorias), y una sola de máster (**Pacios-Lozano; Rueda-Martínez, 2016**).

El interés general por la gestión del conocimiento presenta un patrón similar al de otras propuestas que, tras un fuerte ascenso en popularidad, han ido perdiendo la atención general y diluyéndose poco a poco

### 4. Conclusiones

A partir de los datos expuestos, podemos extraer una serie de conclusiones, principalmente, sobre interés general, producción científica y formación. Apuntaremos también, algunas de las posibles causas que pueden explicar las conclusiones.

El interés general por la disciplina de la gestión del conocimiento presenta un patrón similar al de otras propuestas que, tras un fuerte ascenso en popularidad, han ido perdiendo la atención general y diluyéndose poco a poco. Desde mediados de la década de los 90, el término *knowledge management* se convirtió en una verdadera *buzz word* (palabra de moda) con la que se nombraban varias estrategias, algunas veces contradictorias entre sí, de innovación en la gestión organizacional. Sin embargo, ese interés comienza a decaer a partir de 2004 reduciéndose de forma paulatina la atención por esas propuestas hasta nuestros días.

En el campo de la producción científica ocurre algo muy similar al interés general mostrado por la disciplina. En la producción anual de monografías, el descenso comienza también en 2004 en el corpus de libros en inglés y dos años más tarde, en el 2006, en el fondo en lengua española. Centrados en el ámbito exclusivamente nacional, ese descenso no comienza hasta 2012. Sin embargo, en la producción de artículos científicos indizados en las dos principales bases de datos, el año de inflexión donde comienza el descenso de la producción es 2009. Este descenso en la producción científica no se corresponde necesariamente con un abandono total de la dimensión práctica de la disciplina. Frente a cierto agotamiento en las aportaciones científico-conceptuales, en algunos ámbitos profesionales, especialmente en empresas internacionales, las estrategias de gestión del conocimiento se han implantado con éxito e incorporado al propio ADN organizativo.

En muchas ocasiones, la evolución de la producción científica relacionada con una disciplina no se correlaciona directamente con la formación universitaria ofertada sobre ese tema. Los objetivos monetarios que en muchos casos persigue la formación y el carácter especialmente profesionalizador de sus salidas laborales pueden explicar la existencia de una elevada oferta formativa frente a un descenso paulatino de la producción científica. En todo caso, en términos cuantitativos, la oferta formativa sobre esta disciplina se concentra en territorio norteamericano; y en el caso español destaca la escasa oferta de titulaciones sobre el tema acreditadas por las agencias gubernamentales de evaluación.

No existe una causa única que pueda explicar el ascenso y declive de la disciplina de la gestión del conocimiento. Pero sí que es posible enumerar una serie de factores, relacionados con la propia dimensión cultural y tecnológica de su implementación en las organizaciones, que puedan justificar esa evolución mostrada.

En el ámbito cultural-organizativo destaca, por ejemplo:

- falta de apoyo en la dirección de las organizaciones respecto a los proyectos de este tipo. Sin este apoyo, se han visto abocados al fracaso;
- no se han desarrollado herramientas para la medición, evaluación y control del conocimiento que permitan realizar un análisis y una evaluación adecuada de los proyectos;
- la falta de una clara y transparente política de incentivos que puedan orientar e incentivar a los potenciales participantes y protagonistas de estos proyectos ha limitado también la consolidación de los mismos.

En el terreno tecnológico, las constricciones han sido también importantes. La expansión y consolidación de la disciplina se ha limitado por esa errónea percepción dominante de que la gestión del conocimiento se resume única y exclusivamente a una implantación tecnológica. Tampoco se ha desarrollado un software adecuado (y, por qué no, gratuito) que se haya convertido en un estándar tecnológico en el campo de la implementación. Sin este software, el futuro de la disciplina y sus potenciales beneficios se ha ensombrecido significativamente.

Por último, el hecho de no haber sabido exportar y aplicar decididamente esa disciplina en la administración pública también ha supuesto un freno importante. Esta exportación habría permitido obtener dos importantes beneficios:

- por el especial contexto y problemas que presentan estas organizaciones, esos proyectos se podrían haber convertido en el primer motor del desarrollo teórico-pragmático de la disciplina;
- habría aumentado la popularidad de los mismos al trasladarse directamente al ciudadano todos los importantes beneficios que se obtienen cuando se implementan proyectos de este tipo en las organizaciones encargadas de su administración.

## 5. Bibliografía

**Arrow, Kenneth** (1962). "The economic implication of learning by doing". *Review of economic studies*, v. 29, n. 3, p. 153-173.

**Davenport, Thomas** (1997). *Information ecology*. Oxford: Oxford University Press. ISBN: 978 0195111682

**Davenport, Thomas; Prusak, Laurence** (1998). *Working knowledge*. Boston: Harvard Business School Press. ISBN: 978 1578513017

**Gorey, Rian; Dobat, David R.** (1996). "Managing in the knowledge era". *The systems thinker*, v. 7, n. 8, pp. 1-5. <https://thesystemsthinker.com/managing-in-the-knowledge-era>

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

**Nonaka, Ikujiro** (1991). "The knowledge creating company". *Harvard business review on knowledge management*. Boston: Harvard Business School Publishing, pp. 21-46.

**Nonaka, Ikujiro; Konno, Noboru** (1998). "The concept of 'Ba': Building a foundation for knowledge creation". *California management review*, v. 40, n. 3, p. 40-54. <http://home.business.utah.edu/actme/7410/Nonaka%201998.pdf>

**Nonaka, Ikujiro; Takeuchi, Hirotaka** (1995). *The knowledge creating company*. Oxford: Oxford University Press. ISBN: 978 0195092691

**Pacios-Lozano, Ana-Reyes; Rueda-Martínez, Isabel** (2016). "Una mirada crítica a la formación en gestión del conocimiento en los estudios de información y documentación en España". *BiD: textos universitaris de biblioteconomia i documentació*, n. 36. <http://bid.ub.edu/36/pacios.htm>

**Prusak, Laurence** (2001). "Where did knowledge management come from?". *IBM systems journal*, v. 40, n. 4, pp. 1002-1007. [http://dx.doi.org/10.1016/S0957-4174\(97\)00018-3](http://dx.doi.org/10.1016/S0957-4174(97)00018-3)

**Roknuzzaman, Md; Umemoto, Katsuhiko** (2009). "Knowledge management education in library and information science schools: An exploratory study". En: *Preparing information professionals for international collaboration: Asia-Pacific conference on library & information education & practice (A-LIEP)*. Tsukuba, 6-8 March 2009, pp. 27-37. <http://www.slis.tsukuba.ac.jp/a-liep2009/proceedings/Papers/a48.pdf>

**Senge, Peter M.** (1990). *The fifth discipline: The age and practice of the learning organization*. London: Century Business. ISBN: 0385260954

**Stewart, Thomas A.** (1997). *Intellectual capital: The new wealth of organizations*. New York: Currency/Doubleday. ISBN: 978 0385483810

**Sveiby, Karl-Erik** (1997). *The new organizational wealth: Managing and measuring knowledge-based assets*. San Francisco: Berrett-Koehler. ISBN: 978 1576750148

**Sveiby, Karl-Erik** (2001). *What is knowledge management?* Brisbane: Sveiby Knowledge Associates. <http://www.sveiby.com/articles/KnowledgeManagement.html>

**World Wide Web Consortium** (2012). *Semantic web use cases and case studies*.

## CONSUMPTION OF PUBLIC INSTITUTIONS' OPENDATA BY SPANISH CITIZENS

Consumo de datos abiertos de las instituciones públicas por parte de los ciudadanos españoles

Manuel Gértrudix, María-Carmen Gertrudis-Casado and Sergio Álvarez-García

**Nota:** Este artículo puede leerse traducido al español en:  
[http://www.elprofesionaldelainformacion.com/contenidos/2016/jul/03\\_esp.pdf](http://www.elprofesionaldelainformacion.com/contenidos/2016/jul/03_esp.pdf)



**Manuel Gértrudix**, PhD in information sciences, is a professor of digital communication, and academic director of the *Center for Innovation in Digital Education*, at the *Universidad Rey Juan Carlos* (Spain). Specialist in the area of digital communication, his research analyzes the application of new media including topics such as communication, digital storytelling, e-learning, and competitive intelligence. He has been the technical director of new technologies at the *Ministry of Education and Science* of Spain. He has been a visiting researcher at the *University of Central Florida* (USA) and at the *University of Stirling* (UK).

<http://orcid.org/0000-0002-5869-3116>

[manuel.gertrudix@urjc.es](mailto:manuel.gertrudix@urjc.es)



**María-Carmen Gertrudis-Casado** holds a PhD in communication sciences from the *Universidad Rey Juan Carlos*, in Madrid. She is a researcher with the *Ciberimaginario* group. Since 2005 she has worked as a specialist in the application of information and communication technologies to education for the *National Institute of Educational Technologies and Teacher Training*, *Universidad Rey Juan Carlos*, *Learning 360 Soluciones*, *Universidad del Norte* (Colombia), and the foundation *Enlace Hispano Americano de Salud*. Her research focuses on immersive and interactive education and communication, and information and knowledge management.

<http://orcid.org/0000-0001-6770-203X>

Research Group *Ciberimaginario* (URJC-Icono14)  
[carmen.gertrudis@ciberimaginario.es](mailto:carmen.gertrudis@ciberimaginario.es)



**Sergio Álvarez-García** is an associate professor and researcher at the *Universidad Rey Juan Carlos*, in Madrid. He teaches undergraduate and graduate courses in the field of multimedia communication. He is the content projects coordinator in the *Center for Innovation in Digital Education*. He holds a PhD in information sciences from the *Complutense University of Madrid*. His research is related to the application of information and communication technologies to journalism, advertising, communication, and learning.

<http://orcid.org/0000-0001-9477-2148>

[sergio.alvarez@urjc.es](mailto:sergio.alvarez@urjc.es)

*Universidad Rey Juan Carlos, Faculty of Communication*  
Edif. Departamental, Camino del Molino, s/n. 28943 Fuenlabrada (Madrid), Spain

### Abstract

The consumption practices of Spanish citizens of open data published by public institutions are evaluated, within the context of the implementation of the *Transparency Portal* in Spain. Through an online survey, representative of the studied universe, and through an analysis of descriptive statistics, contingency tables and association measures for nominal variables, the degree of awareness about the availability of open data offered by public bodies are compared, their use is described, and reasons for their use are identified. The results reveal low penetration of the open data culture, and citizen distrust

of the actions carried out by public institutions for the access, use, and reuse of open data, which will require a change of orientation in communication and stimulation policies.

## Keywords

Open data; Public institutions; Transparency; Empowerment; Digital citizenship.

## Resumen

En el contexto de la puesta en marcha del *Portal de la transparencia* en España, se evalúan las prácticas de consumo de datos abiertos publicados por las instituciones públicas, que realizan los ciudadanos españoles. A través de una encuesta online, representativa del universo de estudio, y mediante el análisis de estadísticos descriptivos, tablas de contingencia y medidas de asociación para variables nominales, se contrasta el grado de conocimiento sobre la disponibilidad de datos abiertos que ofrecen los organismos públicos, se describe el uso que hacen de éstos, y se identifican las razones que les motivan a utilizarlos. Las principales conclusiones revelan una baja penetración de la cultura de datos abiertos, y cierta desconfianza de la ciudadanía sobre las acciones que realizan las instituciones públicas para favorecer el acceso, uso y reutilización de los datos abiertos, lo que exige un cambio de orientación en las políticas de comunicación y dinamización.

## Palabras clave

Datos abiertos; Organismos públicos; Transparencia; Empoderamiento; Ciudadanía Digital.

**Gértrudix, Manuel; Gertrudis-Casado, María-Carmen; Álvarez-García, Sergio (2016).** "Consumption of public institutions' open data by Spanish citizens". *El profesional de la información*, v. 25, n. 4, pp. 535-544.

<http://dx.doi.org/10.3145/epi.2016.jul.03>

## 1. Introduction

The implementation of the *Transparency Portal* in Spain is the end result of a lengthy legal battle aimed at providing access to public information. It gives a practical and enforceable specification to initiatives about access to open data from the public sector, thereby facilitating eGovernment, citizen participation, and co-innovation (Cobo, 2012).

“Open data may be considered an “industrial revolution of data”

The purpose of this process is to facilitate citizen access to open data, which is generated by public institutions to carry out their activities. Data that are accessible by any citizen for use, treatment, or re-use within the limits defined by the legislation, are presented in the open format in a direct and standardized way.

However, a culture of transparency and accountability requires multiple approaches to create an open Government; specifically, an improvement in citizen participation and a more involved civil society (Gertrudis-Casado; Gértrudix; Álvarez-García, 2016a).

To better understand this it is essential to discover what the consumption practices of Spanish citizens over 16 years of age are, in relation to open data published by public bodies, based on the citizen's level of knowledge and use, and the mode of access.

### 1.1. The Transparency portal

In accordance with *Law 19/2013 (España, 2013)*, the *Transparency portal* was launched on the 10<sup>th</sup> of December 2014, with over 59,000 visits in the first eight hours of its operation (EFE, 2014).

<http://transparencia.gob.es>

The portal gave access to 530,000 records related to Central administration and was organized into three areas:

- organizational information
- legally significant information, and
- economic information.

According to information published by the *Transparency portal*, on August 31<sup>st</sup> 2016, the number of requests was 4,275, of which 93.54% corresponded to completed files, 6.22% to files being processed, and 0.23% to files held in administrative silence. In terms of the resolution of requests, 68.77% had been resolved with the request being authorized. As for the rest of the requests, in accordance with *Law 19/2013*, 23.93% were not accepted (article 18), 3.23% were denied (article 14 and 15), and 4.08% were dismissed. 89.61% of requests were received through electronic access with a password, the rest on paper. Finally, in terms of the subject matter of requests, they were mostly related to

- statistical data;
- budgets/expenses/revenues;
- salaries and payments to personnel; and
- legislation (processing, acts, and reports).

The *Transparency Information Unit* of the *Ministry of Finance and Public Administration* receiving the most requests (*Ministerio de la Presidencia, 2016*).

### 1.2. Reuse of public sector information

Open data has been considered a kind of “industrial revolution of data” (Hellerstein, 2008). Linked with open government, it offers opportunities in which public sector information and its reuse become an asset for citizen empowerment and is recognized as a basic resource for the evolution of businesses that reuse said information and bring it added va-



Figura 1. Spanish Transparency Portal  
<http://transparencia.gob.es>

lue (Marcos-Martín; Soriano-Maldonado, 2011; European Parliament and Council, 2003).

In 2009 in the United States, transparency was defined for the first time in the *Memorandum on transparency and open government* (Obama, 2009). Transparency, participation and collaboration were established as pillars of open government, and the first international initiative was launched:

<https://www.data.gov>

Currently, sixty-four countries collaborate in the *Open Government Partnership* which facilitates governments in being more open, accountable, and responsive to citizens

Shortly afterwards, directories for New Zealand, Australia, and the United Kingdom were implemented (the last one with the Tim Berners-Lee support).

<https://data.gov.uk>

<http://data.gov.au>

<https://data.govt.nz>

The *Malmö Declaration*, in November 2009, established the aim of

“citizens and companies having to be empowered by electronic administration services [...] as well as by increasing access to public information, greater transparency and more effective methods for the involvement of participants” (Ministerio de la Presidencia, 2009, p. 2).

In 2011, the *Open Government Partnership* was created, initially with eight member countries, which Spain later joined in April 2012. Currently sixty-four countries collaborate on this international platform to facilitate governments in being

“more open, accountable and responsive to citizens” (Open Government Partnership, 2015).

Thus, new relationships have developed between public institutions and citizens: e-administration, e-democracy, e-government, e-governance (Eudel, 2011), based on transparency and access to information as a requirement for governance (Prats-Català, 2005). It implies the need for a strategic government in the public sector, capable of articulating these relationships (Muñoz, 2005) as interdependence-association-coproduction/co-responsibility between the government and the private and social associations (Aguilar-Villanueva, 2007).

The practice of governance has been supplemented by a new form of managing resources, *new public management* (García-Sánchez, 2007), which implies the incorporation of business criteria, allowing for valuable results to be obtained (accountability and value for money) by applying economic, efficiency, and effectiveness principles in public management. It has inherited the principles of total quality management (TQM) (Summers, 2006).

In November 2005 Spain started opening public sector information, and its promotion was entrusted to the public business entity *Red.es*. Two years later, *Law 37/2007 (España, 2007)* was implemented on the 16<sup>th</sup> of November, and the reuse of public sector information was approved, implementing *Directive 2003/98/EC* in national legislation,

and setting out the basic principles for reuse. In 2009 the *Aporta Project* was established by the *Spanish Ministry of Finance and Public Administration (Minhap)* and the *Ministry of Industry, Energy, and Tourism (Minetur)*. From this initiative, in 2010, the first *Public Sector Information Catalog* was created, accessed through the *datos.gob.es* portal. This catalog was developed from the directory of open access databases, sponsored by the administration of the State and created by *Group Publidoc-UCM* (2008).

“*Aporta* has established annual meetings to evaluate the impact of open data from public organizations”

Subsequently, the *Aporta* annual meetings (*Encuentros Aporta*), which began in 2014, established an agenda to evaluate the impact that open data has for both the citizen and institutional intelligence policies. It has been a long process (*Minhap; Minetur*, 2015, p. 4), and the results will be seen in the future, which will require the tendering of a large cross-section of society to extend the term of re-users to

“citizens, developers, sociologists and journalists, among other groups” (*Minetur; Minhap*, 2015, p. 5).

In the *Digital agenda* for Spain, the course of action that the plan develops for encouraging the digital economy and digital content incorporated in *Axis III*, a program for the reuse of public sector information, seeks to

“increase the publication of data sets by Public Administrations and increase business activity in the infomediary sector” (*Minetur; Minhap*, 2015).

On the 11<sup>th</sup> of July 2015, *Law 18/2015 (España, 2015)* of the 9<sup>th</sup> of July came into force, modifying *Law 37/2007* of the 16<sup>th</sup> of November (*España, 2013*), on the reuse of public sector information. This law fulfills the European Union mandate of correspondence with *Directive 2013/37/EU* and updates, in accordance with active legislation, the conditions for reuse of public sector information, especially in the regulation of formats, wherever possible promoting its availability in open, machine-readable formats, along with their metadata.

### 1.3. Social and economic effects of opening data

In 2013, according to *National Observatory for Telecommunications and the Information Society (Ontsi)*, there were over 500 companies dedicated to the reuse of public sector information and over 4,000 employees linked to said sector, producing products and/or services from the reuse of public data, with a turnover of between 450 and 500 million euros (*Ontsi*, 2015a).

In terms of the infomediary sector, according to the *Ontsi* study, there is a trend toward non-exclusive use of sources, using data coming from the public and private sectors to generate products and services for the market studies sector, the economic and financial sector, and others, such as support companies, database generators or price comparison creators (*Ontsi*, 2015b). It is also increasing the reuse of data

in smart city portals through apps (**Abella; Ortiz-de-Urbina-Criado; De-Pablos-Heredero**, 2015).

Some projects, such as *Meloda*, propose a model for measuring the reuse of open data based on four criteria: technical standards, form of access, legal framework, and data model (**Abella; Ortiz-de-Urbina-Criado; De-Pablos-Heredero**, 2014).

In terms of homogenization and standardization, the work of the *World Wide Web Consortium (W3C)* in Spain is notable, through the *Open Data Group* and *RISP (Reuse of Information in the Public Sector)*.

Likewise, the *UNE 178301 (Aenor, 2015)* regulation, created by the specific group of smartcities, defines a set of indicators divided into five pillars (political, organizational, technical, legal, and economic) and the measurements necessary to quantify the level of open data in Spanish cities. The regulation defines 10 sets of data, and its vocabulary aims to serve as a reference for governments to publish their data in a standardized, interoperable, and ultimately open way.

“There are over 500 companies dedicated to the reuse of public sector information with over 4,000 employees”

## 2. Methodology

### 2.1. Aims and hypothesis

The general aim of the research is to examine the consumption practices of Spanish citizens of open data published by public bodies. To do so:

- the extent of knowledge and use by citizens, of open data offered by public bodies, is compared;
- the media diet of individuals who use open data is identified;
- the sociodemographic profile of individuals who use open data is established;
- the reasons for their use are identified; and
- the value of public bodies opening data are revealed, as well as the quantity of open data currently available, the ease of access to them, and the ease of using, analyzing, and interpreting them.

The hypotheses are:

- The lacking culture of transparency and open data in Spain is directly related to the low levels of awareness and usage of open data by citizens.
- The citizens who are most aware and most often use open data are those who have a more diverse media diet with less focus on mass media.

### 2.2. Methodological strategies

To obtain information from citizens, a computer assisted online questionnaire was carried out, using the *LimeSurvey* system at a national level (including the Canary Islands, and the cities Ceuta and Melilla). The study universe is the internet-using Spanish population over 16 years of age.

The structured questionnaire had 16 questions with an estimated completion time of 12-15 minutes (Gértrudix; Gertrudis-Casado; Álvarez-García, 2016b).

The sample size was 1,183 interviews, with a random sample stratified by Autonomous Community (Spanish administrative regions) with proportional allocation, carried out by using the email address directory of the company *Toluna USA*, which is statistically representative and proportional to the Spanish population by size of home and province. The sampling error, for a 95.5% confidence level (two sigmas), and  $P = Q$ , is  $\pm 3.16\%$  for the whole sample and in the case of a simple random sample. The field work was carried out from the 6<sup>th</sup> to the 17<sup>th</sup> of October 2014.

In the data analysis, descriptive statistics were used to evaluate the profile, use, and evaluation of the open data, contingency tables to estimate whether the level of knowledge and use of the open data published by public bodies may be conditioned by the sociodemographic variables studied, and to check the degree of association between the variables a Person Chi-squared test was carried out, as well as other measures of association for nominal variables (based on the Chi-squared test) such as Phi, Pearson's C Contingency Coefficient, and Cramer's V Contingency Coefficient.

### 3. Results

#### 3.1. Profile and use of open data

To evaluate the profile and use of open data, the awareness and use, and reasons for use were analyzed, and subjects were classified using cluster analysis.

##### Knowledge and use of open data published by public bodies

The number of individuals who stated that they used public administration open data was 232, representing 20.6% of the sample. This figure rises, however, when they were asked whether they know said open data, making up 40% of the sample. From the data obtained, a profile of open public institution open data users was identified, taking into account the most revealing characteristics on sample distribution, from sociodemographic variables (Gértrudix; Gertrudis-Casado; Álvarez-García, 2016c).

##### Reasons for using open data published by public bodies

The main reasons for using open data published by public bodies was related to personal context, studies (22.5%), personal curiosity (22.3%), and personal issues (20.3%) being the main reasons, followed by actions in daily life (10.6%). It is notable that social

motivations obtained very low percentages: protests and complaints (8.4%), social demands (8.1%), common causes (6.3%), and collective actions and social movements (1.5%).

“A profile of public institutions' open data users can be constructed from sociodemographic variables”

#### Classification through clusters

In order to obtain a classification, in terms of the frequency of use of open data and their level of compliance with the type of dissemination carried out by public institutions on open data, a cluster analysis was carried out with two of the variables from the study:

- “Frequency of use of open data from public institutions” and
- “Sources of open data from public institutions”,

being the only variables that allowed different groups of respondents to be distinguished (Gértrudix; Gertrudis-Casado; Álvarez-García, 2016c).

Two types of individuals with different characteristics clearly appeared:

- Those individuals who are less linked with open data from public institutions: they rarely used the information based on open data from public institutions at all levels; and disagreed that public institutions provided sufficient open

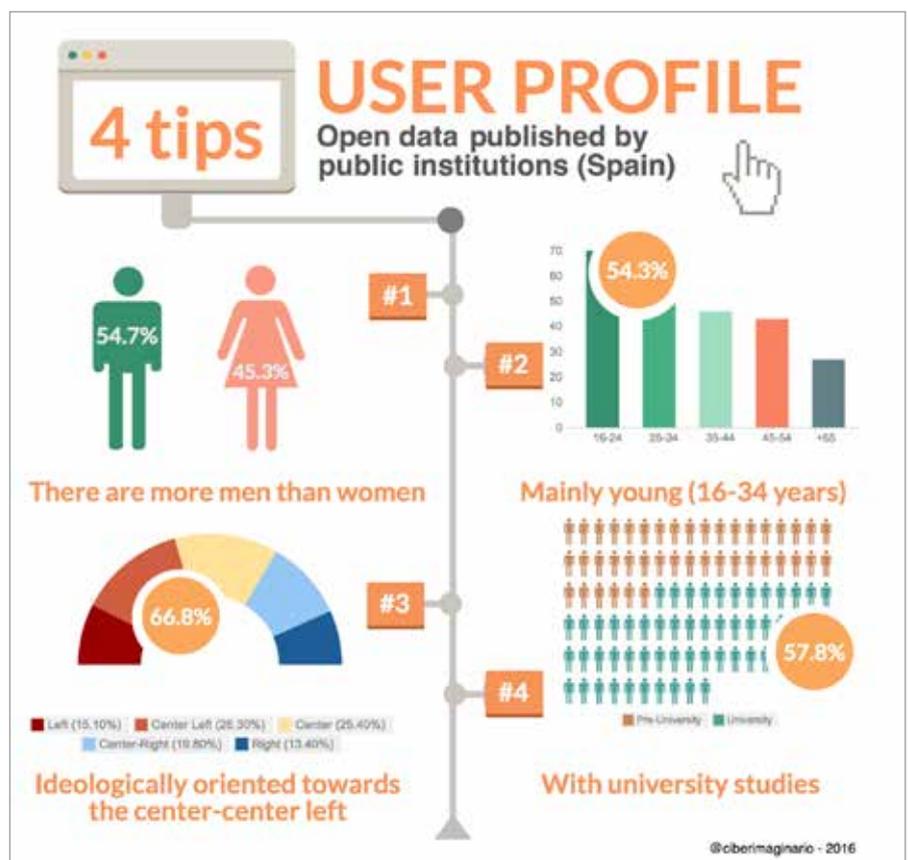


Figure 2. Open data user's profile

data on their activities, that these were offered in an accessible and easy to find manner, or offered citizens simple tools and applications to analyze and interpret the data provided.

- Those individuals who are more linked with open data from public institutions: defined by greater use of information from public institutions at all levels, notably using information that the European Union, Spanish government, autonomous communities, and municipalities disseminate on the internet about their activities. They had an intermediate evaluation of sources of open data. Despite this, they were clearly more positive toward open data than the first individuals.

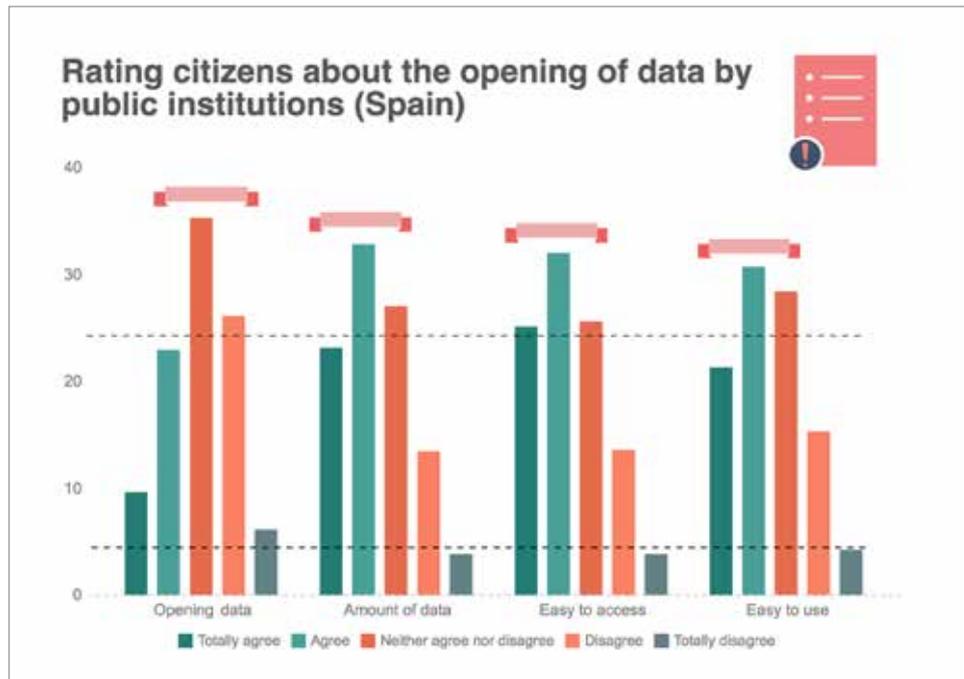


Figure 3. Rating citizens about the opening of data by public institutions of Spain

Once this cluster classification of the respondents had been obtained, a comparison was made with the variable “Access or use of open data published by public bodies”, showing interesting differences between both groups. The majority of the first group (51.5%) were unaware of the existence of open data, and the percentage of individuals who were aware of it had low usage (around 11%). However, a large proportion of the second group (46.8%) stated that they were aware of open data although they had never used them, and there was a rather high percentage of individuals who were aware of and had used open data (almost 30%).

Most respondents believe that public institutions do not offer enough open data about their activities

### 3.2. Evaluation of open data characteristics

There is not a distinct stance on whether public institutions publicize themselves on the internet through their open portals and banks, as 35% neither agree nor disagree. The other evaluations are mainly distributed in two positions on the scale: agree and disagree. Finally, it is notable that there are few cases placed in the most extreme categories.

Most respondents (55.8%) believed that public institutions do not offer sufficient open data about their activities.

Adding to this idea, most of the sample (57%) also disagreed that public institutions provided open data in an accessible and easy to find manner. The majority (50%) also believed that public institutions do not offer easy tools and applications in order to analyze and interpret the open data that they provide.

### 3.3. Level of awareness and use of open data published by public bodies depending on sociodemographic variables

For this analysis, three groups depending on the level of knowledge and use of open data made available by public institutions are considered: those who have used them; those who know their existence but have not used them; and those who do not know their existence.

Considering the gender variable, there were no great differences in the level of use of open data. A variable which presented a larger difference was the degree of awareness of this type of product, in this case women (58.3%) had lower awareness. In terms of the degree of association between both variables, the Chi-squared ( $\chi^2$ ) statistic value is 11.568 (> theoretical value of 5.99) and its level of significance (<0.001) which indicated that the null hypothesis of independence of variables may be rejected, and it can be established that the variables are statistically associated. However, the measures based on  $\chi^2$  (Phi, Pearson’s C Contingency Coefficient, and Cramer’s V Contingency Coefficient) take a value of 0.101 with a critical level of significance of 0.003, indicating that association between the variables was slight.

By age groups, it was observed that there was a greater awareness and use of open data by young people (between 16 and 34 years of age), and only 14.7% stated that they were unaware of it. As ages increased, the percentage of individuals who used open data decreases up to the 35 to 44 year-old age group. In terms of the group that was aware of it, but did not use it, there were not great differences between age groups, except at the extremes, where the youngest were most aware (23.3%) and the oldest were least aware (16.2%). Finally, in terms of those who were unaware of the existence of open data, it is notable

that it was the youngest group that had the highest percentage (21.6%).

Finally, the Chi-squared value ( $\chi^2=25.426 > \chi^2_t= 15.51$ ) and its level of significance (0.001) indicated that the variables were associated. However, the degree of association was not very high.

As there are few cases in many of the educational stages which preceded university, this variable was divided into two large groups: pre-university and university. It was observed that there was a larger percentage of individuals with a university education who were aware of and had used open data, compared with those who were not aware of it, where there were more pre-university individuals.

“The users who are most aware of and most often use open data are those who have a more diverse *media diet*”

However, the association of these variables was statistically significant. The Chi-squared value ( $\chi^2$ ) exceeds the theoretical level for two levels of freedom ( $16.699 > 5.99 \chi^2_t$ ) and, therefore, the null hypothesis of independence of variables can be rejected and it can be concluded that the level of study and access to open data are statistically related (Gérrudix; Gertrudis-Casado; Álvarez-García, 2016c).

In the group of individuals who work, the percentage of individuals who were aware of and had used open data was greater (59.5%) than those who were studying (24.1%). As far as their level of association, the Chi-squared statistic ( $\chi^2=41,852$ ) allows it to be concluded that access and use of open data and the labor situation are statistically linked.

In terms of level of awareness and use of open data published by public bodies, in terms of political ideology, there were not great differences. Furthermore, the statistics indicated that variables were not statistically associated.

Despite the statistics indicating that there was an association between the variables, there were not major differences between the three groups mentioned earlier in this section 3.3 in terms of personal economic circumstances, or civil status. Furthermore, both the value shown by statistics and their level of significance ( $>0.005$ ) indicated that these variables were independent.

### 3.4. Access or use of open data published by public bodies by sources used to search for information on the internet

One of the two hypotheses for this study stated that the users who are most aware of and most often use open data are those who have a more diverse media diet and are less focused on mass media. In order to test this hypothesis a comparative analysis was carried out between the dependent variable “Access or use of open data published by public bodies” and the independent variable “Sources for searching for information on the internet”.

The percentage distributions indicate that the group of those who had accessed and used open data from public institutions were notable for their use of digital communication and public institutions' and tertiary sector websites, and they made greater use of social networks and news aggregators. Likewise, despite the percentage being small, among those seeking information through alternative citizen journalism information channels and media, there was a greater percentage among those who used open data than those who did not use it.

However, the high value of the Pearson Chi-squared statistic (304,680) and its level of significance ( $< 0.001$ ) indicate that the variables are statistically linked (Gérrudix; Gertrudis-Casado; Álvarez-García, 2016c).

## 4. Conclusions and discussion

Despite the importance of expanding the culture of open data (García-Sánchez, 2007; Marcos-Martín; Soriano-Maldonado, 2011; European Parliament and Council, 2003), there is still low penetration in Spain, and the number of individuals in our study who stated that they had used open data from public institutions was only 20.6% of the sample.

Add to this low level of awareness a distrust by citizens towards public institutions that encourage access, use, and reuse of open data, and the long term effects demonstrated in the latest *Aporta meeting* are revealed (Minhap; Minetur, 2015; Minetur; Minhap, 2015). A significant number of citizens do not believe that public institutions offer sufficient information about their activities, or that they are easily accessible, or that they offer easy tools and applications to analyze and interpret open data. This may be explained partly by the lack of awareness shown in studies such as the *CIS* barometer (2012), but also by the lack of continuity which the administrations' evaluation of these issues has had.

“It is important to expand the culture of open data in Spain, because its penetration is currently low”

This may be considered as part of the institutions' lack of will to be transparent and apply the *Law of transparency* (AIE; CLD, 2015).

This shows the long path that open data still has ahead of it and the need for administrations to be aware that they must implement initiatives which is especially relevant when considering that *Law 19/2013* established, in its additional seventh provision, that to improve the impact and effectiveness of its application it would develop “an informative campaign aimed at citizens”.

Spain currently has an abundance of data with zero marginal cost (Rifkin, 2014, p. 333), but the benefits of the data have not reached citizens, possibly because citizens have not been considered (Minhap; Minetur, 2015; Minetur; Minhap, 2015).

Depending on the frequency of use and degree of acceptance of sources of open data we found there were two groups with different characteristics who possessed differentiated levels of awareness, behaviors, and attitudes relating to the phenomenon of open data. There is a relationship between interest, awareness, use, and trust which reveals the importance of public institutions maintaining a proactive attitude toward the design and implementation of initiatives and actions which give value to transparency and the data open-

There is an abundance of data at zero marginal cost for society, but citizens are not yet aware of the potential benefits

ning process, as the *Open Data Charter* (G8, 2013) and the *Estudio de la demanda y uso de gobierno abierto en España* (2013) from *Ontsi* indicate.

Contrary to expectations, social causes (protests and complaints, social demands, common causes, etc.) did not motivate the use of open data, as none of these reached 10%. It was instead personal motivations which were the main reasons for use, which is consistent with the results of studies such as *Gurstein* (2011), *Bates* (2012), *Janssen, Charalabidis and Zuiderwijk* (2012), *Tinati et al.* (2012) and *Zuiderwijk and Marijn* (2014).

The analysis of the relationship between level of awareness and use of open data published by public bodies and the sociodemographic variables show slight associations in terms of gender and age variables, although there was a greater awareness of open data in the case of men and a greater use in the age group between 16 and 34 years of age, falling as age increases. The level of education did have a statistically significant association and the higher the level of studies then the greater the awareness and use of open data, with the greatest number being individuals who had completed university studies. This is also the case with the occupation variable. Individuals who work were more likely to be aware of and use open data provided by public institutions. Significant relationships were not found in terms of personal economic circumstances, civil status, or geographic location by autonomous community in Spain.

The hypothesis that users who are more aware and make greater use of open data are those who have a more diverse media diet and are less focused on mass media was considered and showed that both variables are statistically related. It is therefore shown that access to data portals and the options which these provide help citizens to expand their media diet (*Hasebrink; Domeyer*, 2012; (*García-García; Gértrudix; Gértrudix*, 2014).

## Acknowledgement

This work was supported by the *Ministry of Economy and Competitiveness* of Spain under grant "*OpenDataCitizen. Digital citizenship and open data access: Citizen empowerment through social media in the digital environment*" (CSO2012-30756), and *Grupo de Excelencia investigadora URJC-Banco de Santander Anima+D-Ciberimaginario*.

## 5. References

- Abella, Alberto; Ortiz-de-Urbina-Criado, Marta; De-Pablos-Heredero, Carmen** (2014). "Meloda, métrica para evaluar la reutilización de datos abiertos". *El profesional de la información*, v. 23, n. 6, pp. 582-88.  
<http://dx.doi.org/10.31145/epi.2014.nov.04>
- Abella, Alberto; Ortiz-de-Urbina-Criado, Marta; De-Pablos-Heredero, Carmen** (2015). "The reuse of information in smart cities' ecosystems". *El profesional de la información*, v. 24, n. 6, pp. 838-844.  
<http://dx.doi.org/10.31145/epi.2015.nov.16>
- AIE; CLD* (2015). "Country data". *Global right to information rating*. Access Info; Centre for Law and Democracy.  
<http://www.rti-rating.org/country-data>
- Aenor* (2015). UNE 178301:2015. Ciudades inteligentes. Datos abiertos. 21 enero.  
<http://goo.gl/urFuTk>
- Aguilar-Villanueva, Luis-Fernando** (2007). *Gobernanza y gestión pública*. México D.F.: Fondo de Cultura Económica. ISBN: 978 9681681333
- Bates, Jo** (2012). "This is what modern deregulation looks like: co-optation and contestation in the shaping of the UK's Open Government Data Initiative". *The journal of community informatics*, v. 8, n. 2.  
<http://ci-journal.net/index.php/ciej/article/view/845/916>
- CIS* (2012). *Distribuciones marginales. Barómetro de mayo. Estudio nº 2.944*. Centro de Investigaciones Sociológicas.  
<http://goo.gl/LwzobW>
- Cobo, Cristóbal** (2012). "Gobierno abierto: de la transparencia a la inteligencia cívica". In: Hofmann, Andrés; Ramírez-Alujas, Álvaro; Bojórquez-Pereznieta, José-Antonio (coords.) *La promesa del gobierno abierto*. México D.F.: Itaip, InfoDF, pp. 101-118.  
<http://goo.gl/gmH7rO>
- EFE* (2014). "Más de 59.000 usuarios visitan el Portal de la transparencia el día de su estreno". *El mundo*, December 10.  
<http://goo.gl/xechMb>
- España (2007). "Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público". *BOE*, n. 276, 17 de noviembre.  
[https://www.boe.es/diario\\_boe/txt.php?id=BOE-A-2007-19814](https://www.boe.es/diario_boe/txt.php?id=BOE-A-2007-19814)
- España (2013). "Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno". *BOE*, n. 295, de 10 diciembre de 2013.  
<http://goo.gl/zqiB3o>
- España (2015). "Ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público". *BOE*, n. 164, 10 de julio.  
[https://www.boe.es/diario\\_boe/txt.php?id=BOE-A-2015-7731](https://www.boe.es/diario_boe/txt.php?id=BOE-A-2015-7731)
- España. *Minetur; Minhap* (2015). *Plan de impulso de la economía digital y los contenidos digitales*.  
<http://goo.gl/kx4u9b>
- España. *Minhap; Minetur* (2015). *Informe conclusiones me-*

- sas redondas. Encuentro aporta 2015: el dato público en una sociedad digital*. Madrid.  
<http://goo.gl/TWUJX6>
- España. *Ministerio de la Presidencia* (2009). *Declaración ministerial sobre la administración electrónica*.  
<http://goo.gl/e1GeHq>
- España. *Ministerio de la Presidencia* (2016). *Informe de situación. Portal de la transparencia*, febrero.  
<http://goo.gl/jZwJGC>
- Eudel (2011). *eGobernanza, participación ciudadana e innovación*. Asociación de Municipios Vascos.  
<http://www.eudel.eus/es/archivos/libro/eGobernanza.pdf>
- European Commission (2013). *DCAT application profile for data portals in Europe*.  
[https://joinup.ec.europa.eu/asset/dcat\\_application\\_profile/home](https://joinup.ec.europa.eu/asset/dcat_application_profile/home)
- European Parliament and Council (2003). "Directiva 2003/98/CE de 17 de noviembre de 2003 relativa a la reutilización de la información del sector público". *Diario oficial de la Unión Europea*, n. L 345 de 31/12/2003, pp. 90-96.  
<http://goo.gl/3nx2FN>
- G8 (2013). *Open data charter*. Lough Erne.  
<https://goo.gl/rHdJ9a>
- García-García, Francisco; Gértrudix, Manuel; Gértrudix, Felipe** (2014). "Analysis of the incidence of media digital repertoires in the utility and reliability of information on the internet in university students". *Communication & society*, v. 27, n. 1, 59-81.  
<http://hdl.handle.net/10171/36262>
- García-Sánchez, Isabel-María** (2007). "La nueva gestión pública: evolución y tendencias". *Presupuesto y gasto público*, n. 47, pp. 37-64.  
<http://goo.gl/KX2GEB>
- Gertrudis-Casado, María-Carmen; Gértrudix, Manuel; Álvarez-García, Sergio** (2016a). "Professional information skills and open data. Challenges for citizen empowerment and social change". *Comunicar*, n. 47, pp. 39-47.  
<http://dx.doi.org/10.3916/C47-2016-04>
- Gértrudix, Manuel; Gertrudis-Casado, María-Carmen; Álvarez-García, Sergio** (2016b). "Questionnaire open data project citizen". *Figshare*, 2 July.  
<https://dx.doi.org/10.6084/m9.figshare.3470123.v2>
- Gértrudix, Manuel; Gertrudis-Casado, María-Carmen; Álvarez-García, Sergio** (2016c). "Tables about use of open data in Spain". *Figshare*, 22 March.  
<https://dx.doi.org/10.6084/m9.figshare.3121927.v2>
- Gurstein, Michael B.** (2011). "Open data: Empowering the empowered or effective data use for everyone?". *First Monday*, v. 16, n. 2.  
<http://dx.doi.org/10.5210/fm.v16i2.3316>
- Hasebrink, Uwe; Domeyer, Hanna** (2012). "Media repertoires as patterns of behaviour and as meaningful practices: A multimethod approach to media use in converging media environments". *Participations. Journal of audience & reception studies*, v. 9, n. 2, pp. 757-779.  
<http://goo.gl/5oTYc2>
- Hellerstein, Joseph** (2008). "The commoditization of massive data analysis". *O'Reilly radar*, November 19.  
<http://radar.oreilly.com/2008/11/the-commoditization-of-massive.html>
- Janssen, Marijn; Charalabidis, Yannis; Zuiderwijk, Anneke** (2012). "Benefits, adoption barriers and myths of open data and open government". *Information systems management (ISM)*, v. 29, n. 4, pp. 258-268.  
<http://dx.doi.org/10.1080/10580530.2012.716740>
- Marcos-Martín, Carlos; Soriano-Maldonado, Salvador-Luis** (2011). "Reutilización de la información del sector público y open data en el contexto español y europeo. Proyecto Aporta". *El profesional de la información*, v. 20, n. 3, pp. 291-297.  
<http://dx.doi.org/10.3145/epi.2011.may.07>
- Muñoz, Emilio** (2005). "Gobernanza, ciencia, tecnología y política: trayectoria y evolución". *Arbor. Ciencia, pensamiento y cultura*, v. 715, pp. 287-300.  
<http://dx.doi.org/10.3989/arbor.2005.i715.413>
- Obama, Barack** (2009). "Transparency and open government". *The White House*.  
<https://goo.gl/iRdOZo>
- Ontsi** (2012). *Estudio de caracterización del sector infomediario en España*. Madrid.  
[http://datos.gob.es/sites/default/files/Info\\_sector%20infomediario\\_2012\\_vfr.pdf](http://datos.gob.es/sites/default/files/Info_sector%20infomediario_2012_vfr.pdf)
- Ontsi** (2013). *Estudio de la demanda y uso de gobierno abierto en España*. Madrid.  
<http://www.ontsi.red.es/ontsi/ca/node/6896>
- Ontsi** (2015a). *Estudio de caracterización del sector infomediario en España 2014. Parte I. Sector infomediario público*. Madrid.  
<http://www.ontsi.red.es/ontsi/en/node/7032>
- Ontsi** (2015b). *Estudio de caracterización del sector infomediario en España 2014. Parte II. Reutilización de la información del sector privado*. Madrid.  
<http://www.ontsi.red.es/ontsi/en/node/7036>
- Open Government Partnership (2015). *Open Government Partnership*.  
<http://www.opengovpartnership.org/es>
- Prats-Català, Joan** (2005). "Modos de gobernación de las sociedades globales". In: Cerrillo-Martínez, Agustí (coord.). *La gobernanza hoy: 10 textos de referencia*. Madrid: Instituto Nacional de Administración Pública. ISBN: 8473512391
- Publidoc-UCM** (2008). *Bases de datos de libre acceso difundidas por la Administración General del Estado*. Madrid: Editorial Complutense. ISBN: 978 8474919400
- Rifkin, Jeremy** (2014). *La sociedad de coste marginal cero: El internet de las cosas, el procomún colaborativo y el eclipse del capitalismo*. Barcelona: Paidós. ISBN: 978 8449330513

**Sánchez-de-Diego, Manuel** (2010). "De la libertad de expresión al derecho de acceso a la información pública". *Comunicación y pluralismo*, v. 9, pp. 231-266. <http://goo.gl/Dehe3X>

**Summers, Donna C. S.** (2006). *Administración de la calidad*. México: Prentice-Hall. ISBN: 978 9702608134

**Tinati, Ramine; Carr, Les; Halford, Susan; Pope, Catherine** (2012). "Exploring the impact of adopting open data in

the UK government". In: *Digital futures 2012*, Aberdeen, GB. <http://eprints.soton.ac.uk/id/eprint/344808>

**Zuiderwijk, Anneke; Marijn Janssen** (2014). "Open data policies, their implementation and impact: A framework for comparison". *Government information quarterly*, v. 31, n. 1, pp. 17-29. <http://dx.doi.org/10.1016/j.giq.2013.04.003>

ANUARIO

# Think

EPI

**ISSN: 1886-6344**  
**ISBN: 978 84 9116 439 5**

## ANUARIO THINKEPI 2016

**PRECIOS ANUARIO THINKEPI**

**Suscripción online (2007-2016)**

Instituciones ..... 80 €

Individuos (particulares) ..... 48 €

**Números sueltos**

**Instituciones**

Anuario ThinkEPI 2016 (pdf) ..... 55 €

Anuario de años anteriores\* ..... 30 €

**Individuos (particulares)**

Anuario ThinkEPI 2016 (pdf)..... 30 €

Anuario de años anteriores\* ..... 22 €

\*Años 2007 a 2013 disponibles en papel + pdf.  
A partir de 2014 sólo disponible en pdf

Desde 2014 es posible el acceso mediante suscripción a todos los **Anuarios ThinkEPI** publicados hasta el momento desde el Recyt de la Fecyt

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

**Más información:**  
Isabel Olea  
[epi.iolea@gmail.com](mailto:epi.iolea@gmail.com)



# SOCIAL SCIENCES AND HUMANITIES RESEARCH DATA AND METADATA: A PERSPECTIVE FROM THEMATIC DATA REPOSITORIES

Datos y metadatos de investigación en ciencias  
sociales y humanidades: una aproximación desde los  
repositorios temáticos de datos



**Nancy-Diana Gómez, Eva Méndez and Tony Hernández-Pérez**

**Nota:** Este artículo se puede leer en español en:

[http://www.elprofesionaldelainformacion.com/contenidos/2016/jul/04\\_esp.pdf](http://www.elprofesionaldelainformacion.com/contenidos/2016/jul/04_esp.pdf)



**Nancy-Diana Gómez**, librarian and graduated in arts from the *University of Buenos Aires*, currently is a PhD student with the *Archives and libraries in the digital environment* program at the *Carlos III University of Madrid (UC3M)*. She has taught in the *Department of Library and Information Science* at the UC3M (2009-2013) and at the *University of Buenos Aires*, where she was also director of the *Central Library of the Faculty of Natural Sciences* (1994-2005). She is co-coordinator of the *Latin American list on Open Access Repositories (Llaar)*, and participates in national and international research projects.

<http://orcid.org/0000-0002-6218-6248>

[ndgomez@bib.uc3m.es](mailto:ndgomez@bib.uc3m.es)

**Eva Méndez** is an associate professor with the *Department of Library and Information Science* at the *Carlos III University of Madrid*, where she is currently vice provost for *Strategy and digital education*. Doctor of documentation, her teaching and research deal with metadata, semantic web, digital libraries, open access, information policies, and social web. She is a member of the *Dublin Core (DCMI)* advisory board. Since 2015 she has also belonged to the *OpenAIRE* advisory committee and the *Rebiun* executive committee. She has participated as an independent expert for the *European Commission* on digital libraries and open science.

<http://orcid.org/0000-0002-5337-4722>

[emendez@bib.uc3m.es](mailto:emendez@bib.uc3m.es)

Manuscript received on 18-04-2016

Approved on: 08-06-2016

**Tony Hernández-Pérez** holds a PhD in information science and is a professor with the *Department of Library and Information Science* at the *Carlos III University of Madrid* where he is the director of the doctoral program in documentation. His teaching and research are linked to the *TecnoDoc* group matters: social web, web content management, metadata, information retrieval, e-learning, and journalistic and audiovisual documentation.  
<http://orcid.org/0000-0001-8404-9247>

[tony@bib.uc3m.es](mailto:tony@bib.uc3m.es)

*Universidad Carlos III de Madrid*  
*Facultad de Humanidades, Comunicación y Documentación*  
C/ Madrid, 128. 28903 Getafe (Madrid), Spain

## Abstract

This paper studies research data repositories in the social sciences and humanities (SSH), from the *Registry of Research Data Repositories (re3data)*, paying particular attention to metadata models used to describe the datasets included in them. 397 repositories are reviewed at the general level, including those of a multidisciplinary nature. We discuss and reflect on the special features of research data in these disciplines, and on coverage and information collected by *re3data*. The metadata schemas and standards most commonly used in SSH repositories are analyzed, with special emphasis on the six main repositories.

## Keywords

Repositories; Research data; Metadata; Social sciences; Humanities; *Re3data*.

## Resumen

Se estudian los repositorios de datos de investigación en ciencias sociales y humanidades (CSH), recogidos en el *Registro de repositorios de datos de investigación (re3data)*, prestando especial atención a los modelos de metadatos que utilizan para describir los datasets incluidos en ellos. Se revisan a nivel global los 397 repositorios que, según *re3data*, recogen datos de investigación sobre esas disciplinas, incluidos, los de carácter multidisciplinar. Se discute y reflexiona sobre las particularidades de los datos de investigación en estas disciplinas y sobre la cobertura e información que recoge *re3data*. Se analizan los esquemas y estándares de metadatos más utilizados en los repositorios de CSH, con un análisis más pormenorizado de los seis repositorios de datos especializados más importantes.

## Palabras clave

Repositorios; Datos de investigación; Metadatos; Ciencias sociales; Humanidades; *Re3data*.

**Gómez, Nancy-Diana; Méndez, Eva; Hernández-Pérez, Tony (2016).** "Social sciences and humanities research data and metadata: A perspective from thematic data repositories". *El profesional de la información*, v. 25, n. 4, pp. 545-555.

<http://dx.doi.org/10.3145/epi.2016.jul.04>

## 1. Introduction

Research data management is becoming increasingly important in all scientific fields. A logical and necessary evolution due, on the one hand, to the technological development that increasingly allows a science based on data, and secondly, to the political impetus of the idea of open science that includes, besides open access to publications, the opening of the data used in the research process.

Sharing research data has become standard practice in disciplines where there is a collaborative scientific culture, such as physics, astronomy (Pepe et al., 2014), and genetics (Paltoo et al., 2014). This disciplinary culture is further compounded by the fact that publicly funded research institutions are beginning to require researchers to publish the results, not only in the form of publications, but also by opening the underlying data used. Opening research data is recommended by OECD (2015) and required by the US government and various funding agencies such as the *National Science Foundation (NSF, 2014)* and *National Institutes of*

*Health (NIH, 2015)*. In Europe, open access to research data has been, so far, only a pilot (*ORD Pilot*) for nine areas of projects funded under *Horizon 2020* with other areas and programs invited to voluntarily participate (*European Commission, 2016*). However, on April 19, 2016, the *Commission* stated that by 2017 research data will be open by default for all new *H2020* funded projects (COM 2016, p. 8).

The trend towards open data is growing within all institutions involved in research, both by the agencies that fund, and the organizations that carry out research (e.g. *League of European Research Universities, LERU, 2013*), and by journal editors who publish research results (e.g. *PLoS, 2014*). Although this trend varies from one discipline to another and between individual researchers, there are many motivations for sharing data (Kim; Stanton, 2016) and benefits that transcend trends or mandates (Lyon, 2016):

- increases the possibility of research having more impact and visibility;
- favors the reproducibility of science;

- saves costs when creating data;
- promotes collaboration;
- contributes to increased credibility in the system.

Of course, there are also many researchers reluctant to share “their” data. A study carried out by *Wiley* surveyed 2,886 researchers (**Ferguson**, 2014) and revealed some of their concerns:

- afraid of the negative consequences of sharing data (misuse, legal or commercial consequences, etc.);
- lack of recognition;
- amount of work involved in preparing the data for publication;
- lack of knowledge about how and where to share data.

For each discipline or scientific domain there is a unique interpretation of datasets or datasets’ research, their nature, data collection, and metadata description

### 1.1. Research data: a discipline problem as seen from the social sciences and humanities (SSH)

For each discipline or scientific domain there is a particular interpretation of datasets and research data, their nature, and collection procedures. And of course, variations in the way that data are described with metadata and the problems associated with sharing. **Christine Borgman**, who has extensively dealt with this (**Borgman**, 2008; **Borgman**; **Wallis**; **Mayernik**, 2012) refers to the concept of data as:

“Facts, numbers, letters and symbols that describe an object, idea, condition, situation or other factors” and also “digital manifestations of literature (including text, sound, still images, moving images, models, games or simulations)”.

Moreover, the *NSF* in the USA distinguishes between observational data, computer data, and experimental data, but all are considered digital (**Borgman**; **Wallis**; **Mayernik**, 2012).

However, in the social sciences and humanities (SSH) not all data are collected digitally and data may take many other forms and formats. For example, in sociology the data from surveys and interviews can easily be captured digitally; however, in archeology the results of observational data can be more closely linked to the object and to the background information about the object [geographical coordinates, samples and drawings of the object (on paper), photographs, or videos (digital)] (**Frank**; **Yakel**; **Faniel**, 2015).

Another key issue in SSH is the source of the data, because many investigations are based on data that were not originally produced by or for the researchers. For example, government data and corporate documents which are used to generate new data, that is, data used “for” research to generate other data “from” research. Humanities scholars are much more dependent on external data sources than researchers from other disciplines. Almost every record of human activity can be considered “data” (**Borgman**, 2008).

Compared with those of pure sciences, SSH researchers generate much less data through observations, since generally they tend to use data from all kinds of sources, which may include sounds for linguistic studies and films for object, dress, or speech analysis. They also use historic materials, such as books, maps, newspapers, journals, photographs, and administrative records —as a result research data and publications may be confused or intermingled.

The *National Endowment for the Humanities (NEH)* in the USA defines data as materials generated or collected in the course of an investigation, for example, citations, software code, databases, geospatial coordinates, reports, and articles. However, the *NEH* expressly excludes article drafts and communications with colleagues (*NEH*, 2015). Furthermore, within the broad spectrum of subjects and disciplines covering the humanities, there can be different definitions of data, which can further complicate the outlook for their management and recovery.

Perhaps the unique characteristics of SSH researchers helps explain why only 46% of them share data in repositories (**Meadows**, 2014). Or, perhaps it is a lack of knowledge about where and how to share, the fuzzy boundaries between data and publications, and between data “from” research and “for” research.

### 1.2. Metadata or how to make data useful for research

Unlike what happens with publications, where despite different disciplinary styles there is a common core of formal properties, scientific data show a heterogeneity that varies radically across disciplines, thematic areas, and even between research groups and researchers.

The *NSF* in the US requests that the *data management plan* includes the metadata standards that are used (**Bischoff**; **Johnston**, 2015). The pilot open data (*ORD Pilot*) of the *European Commission* (2016) further requests that the metadata associated with data —ultimately what makes the data useful - be included. In the world of digital libraries, metadata have always contributed to making data useful by describing publications and other digital or digitized objects or assets. And in the world of data, metadata makes data useful: describing, dimensioning, and contextualizing so that they can be found, regardless of the silo discipline in which are situated, enabling reuse across other domains. Without metadata and descriptions of research methods and context, data are just collections of numbers, codebooks, pretty pictures, or boxes of stones (**Borgman**, 2008).

Funding agencies are raising awareness and putting pressure on researchers to manage their data, share data in a reusable way, facilitate the recovery and preservation of data, and ensure that data are FAIR (findable, accessible, interoperable, and reusable). The creation of FAIR data and science highlights the need to improve the e-infrastructure for scientific information reuse (**Wilkinson et al.**, 2016), but also the need to promote interoperability from the metadata.

When researchers share their data and metadata in a data repository, they should translate the meta-information they

use in their VREs (virtual research environments), on their servers, and on their personal computers –what **Tenopir et al.** (2015) called *laboratory metadata* or *institution specific metadata*– into the standard metadata schema used in the repository. **Tenopir** and his research team surveyed more than 1,000 researchers in each of its two studies, conducted in 2011 and 2015, on how to manage their data (**Tenopir et al.**, 2011); more than 50% said they did not use any metadata standard, 14% said they used some standard within their institution, and 20% used their laboratory standard (in the 2011 study); the 2015 study found similar results (47.9% none and 16.7% laboratory standard). In our study we analyzed the metadata schemes used by repositories, or at least those schemes that repository administrators claim to use to describe the data deposited in *re3data* by SSH researchers.

‘ The creation of data and science FAIR underscores the need to improve the e-infrastructure for the reuse of scientific information and the need to promote interoperability from metadata ’

## 2. Objectives and methodology

According to the context that we provided in the previous section, this article focuses on two domains (social sciences and humanities) where there has not been a historic tradition of collaboration, managing research data, standardized metadata schemes (with some exceptions), virtual research environments, or other e-infrastructures that require the use of metadata. We address the problem of scientific data management in SSH, through a study of data repositories of those disciplines, included in *re3data* (a repository funded by the *German Research Foundation*), to answer the following research questions:

- What kind of data are stored and managed by specific SSH repositories?
- How is the distribution of research data repositories among the various areas of knowledge within SSH?
- What thematic areas are most represented?
- What metadata schemes are used in these repositories to identify and describe the different types of data?
- Is there a predominant scheme or model in each case?

### 2.1. Objectives

- To identify SSH research data repositories.
- To study what types of data result from research in these disciplines by analyzing data stored in major repositories.
- To present the metadata schemes most used in these repositories.

This is an exploratory study to identify the most representative SSH specialized repositories, to investigate their practices, and to verify the type of stored data and metadata schemes that they use or claim to use.

### 2.2. Methodology

For the analysis we used the aforementioned *re3data* (*Registry of Research Data Repositories*) as a source, because it is a reference registry for data repositories recommended by both the *European Commission* (2016), and various publishers (*PeerJ*, *Springer*, *Nature's Scientific Data*, etc.). This registry enables easy identification of data repositories by subject or discipline.

Initially a quantitative and analytical methodology to analyze the 397 repositories included in the SSH category of *re3data* was considered. However, during the initial phase of the research the course was changed to carry out a detailed study of a small sample of the most representative SSH specialized repositories, three in social sciences and three in humanities, to verify the declared metadata schemes. Thus, the work was carried out in three phases:

#### a) Extraction and treatment of *re3data* records

In this phase, several tasks were carried out:

a.1. Retrieval and extraction, through the API provided by *re3data*, of a total of 1,457 registered repositories (at the time of data collection, February, 18th 2016). Please note that in April 2016, *re3data* announced that it has already reached 1,500 data repository records. Although the latest version of the descriptive scheme (metadata) of *re3data* is 3.0 (**Rücknagel et al.**, 2015), the API responds to the first version of the scheme, which is much more limited than the latest version.

a.2. From the repositories list, 1,457 records describing them were downloaded in xml format using scraping techniques with *R*<sup>1</sup>.

a.3. The records were treated through *xslt*<sup>2</sup> to process the information for this study, mainly: data and metadata types used by the repositories and their classification and identification schemes.

#### b) Sample selection and quantitative analysis of the extracted data

The objective of this phase was to filter the data repositories to which we wanted to focus the study, those with SSH content. We selected those that contained some thematic classification scheme on humanities or social sciences in their description, according to the classification used by *re3data* that can be seen in table 1.

In the case of thematic classification, it should be noted that to classify a repository according to its metadata schema, *re3data* provides the property *SubjectScheme* as a mandatory attribute that allows researchers to enter an unlimited number of values, always bearing in mind that the only allowed values are those from the thematic classification of the *German Research Foundation* (*DFG Classification of subject area*). This classification covers four broad areas:

- humanities and social sciences
- life sciences

- natural sciences
- engineering sciences.

It should be taken into account that each repository can be described with as many subjects as it covers, so that a single repository may appear in more than one subject area and even simultaneously in the four thematic areas, as happens in multidisciplinary cases.

After filtering, we obtained 397 records related to SSH, which constituted our sample size to analyze the types of data and metadata schemas of each repository.

To identify the type of data, the *ContentType* property (not mandatory) of the *re3data* scheme was used, which allows specification of all types of content available in a repository. The values allowed in this field are restricted to the types of content recognized and identified in the *Parse.insight (Permanent Access to the Records of Science in Europe)* project. The *Parse* classification has 15 options:

- Archived data
- Audiovisual data
- Configuration data
- Databases
- Images
- Network based data
- Plain text
- Raw data
- Scientific and statistical data formats
- Software applications
- Source code
- Standard office documents
- Structured graphics
- Structured text
- Other.

Table 1. Number of SSH repositories, including multidisciplinary repositories, according to the *DFG* classification

<i>SubjectScheme de DFG</i>	Number of repositories
1 Humanities and social sciences	397
101 Ancient cultures	15
102 History	34
103 Fine arts, music, theatre and media studies	26
104 Linguistics	47
105 Literary studies	10
106 Non-European languages and cultures, social and cultural anthropology, Jewish studies and religious studies	18
107 Theology	4
108 Philosophy	3
109 Education sciences	146
110 Psychology	14
111 Social sciences	155
112 Economics	114
113 Jurisprudence	27

However, it is important to note that it is not mandatory to select one of them when completing the registration on the repository.

Finally, to identify the metadata schema, we used the *MetadataStandardName* scheme property, which again is not mandatory.

### c) Identification of a subset of key data repositories in SSH

Once we identified the subset, a qualitative and individual analysis of the metadata schemes was conducted.

This last phase of the methodology was included because we identified two limitations of *re3data*:

- to complete / declare the metadata schema used by a repository is not mandatory;
- the information about the standard used is the one at the time when the registration was completed and it might have changed over time.

The idiosyncrasies of social sciences and humanities researchers may lead many of them to withhold their data, but this withholding may also be the result of ignorance about where and how to share

When accessing the repositories other difficulties were revealed:

- corroborating the metadata schemes declared in *re3data*;
- restricting access to authorized users, in some cases;
- missing manual or bibliography, etc., in some cases.

So, to continue the study, three repositories in social sciences and three in the humanities were selected based on:

- coverage or number of datasets stored;
- level of use made by their respective communities;
- representation for this study, covering various topics and countries.

In the case of humanities, a repository of linguistics (*Clarín*), one of archeology, and another of history and art (*Prometheus*) were selected, because these subdisciplines were represented by the data repositories in *re3data*. Selected repositories are shown in table 2.

## 3. Results and discussion

### 3.1. Research data repositories in SSH

A first overview of the existing repositories registered in *re3data* SSH, can be seen in figure 1: A *treemap* representing the number / volume of repositories of the areas studied, according to the sub-classification of SSH in table 1.

In order to give a more accurate picture, a table with the number of repositories according to the *DFG* classification and the *SubjectScheme* used by *re3data*, which includes four levels, is provided. In table 1 it is indicated to the third level.

It is to be noted that a multidisciplinary repository may be in more than one category, so the sum of the parts exceeds the

total. According to the thematic classification of the DFG, social sciences (codes 109 to 113) have a higher representation: 456 versus 157 in the humanities.

### 3.2. SSH research data

The types of content available in *re3data* were represented according to the types recognized and identified in the *Parse.insight* project. And, as noted, the type of scientific and statistical data (formats such as *spss*, *fits*, *gis*, etc.) along with documents (*Word*, *Excel* or similar *OpenOffice* formats) and images (*jpeg*, *jpeg2000*, *gif*, *tif*, *png*, *svg*, etc.) are the most commonly used in digitization projects in the humanities. Figure 2 shows that the proportion of content types is relatively balanced in all areas of science, in contrast with the use made in humanities and social sciences. For each type of data (scientific data, images, plain text, raw data, etc.) the use that is done in SSH is usually about 27% (a minimum

Table 2. Selection of representative repositories (SSH)

Social sciences	
<i>Inter university Consortium for Political and Social Research (ICPSR, EUA)</i>	<a href="http://www.icpsr.umich.edu">http://www.icpsr.umich.edu</a>
<i>UK Data Service (Reino Unido)</i>	<a href="https://www.ukdataservice.ac.uk">https://www.ukdataservice.ac.uk</a>
<i>Gesis Zacat (Alemania)</i>	<a href="http://zacat.gesis.org/webview">http://zacat.gesis.org/webview</a>
Humanities	
<i>Common Language Resources and Technology Infrastructure (Clarín, EU):</i>	<a href="http://www.clarin.eu">http://www.clarin.eu</a>
<i>Archaeology Data Service (Reino Unido)</i>	<a href="http://archaeologydataservice.ac.uk">http://archaeologydataservice.ac.uk</a>
<i>Prometheus (Alemania)</i>	<a href="http://www.prometheus-bildarchiv.de">http://www.prometheus-bildarchiv.de</a>

of 20% and a maximum of 32%). It is surprising that there was not a higher percentage of “standard office documents” as compared to “scientific and statistical data formats” or “raw data” type. All document types are present with more or less the same proportions, (73% in other disciplines and 27% in SSH). Even in “audiovisual data”, near the end of the graph, there were fewer “standard office documents”, and the proportion (69.3% in other disciplines and 30.7% in SSH) was maintained.

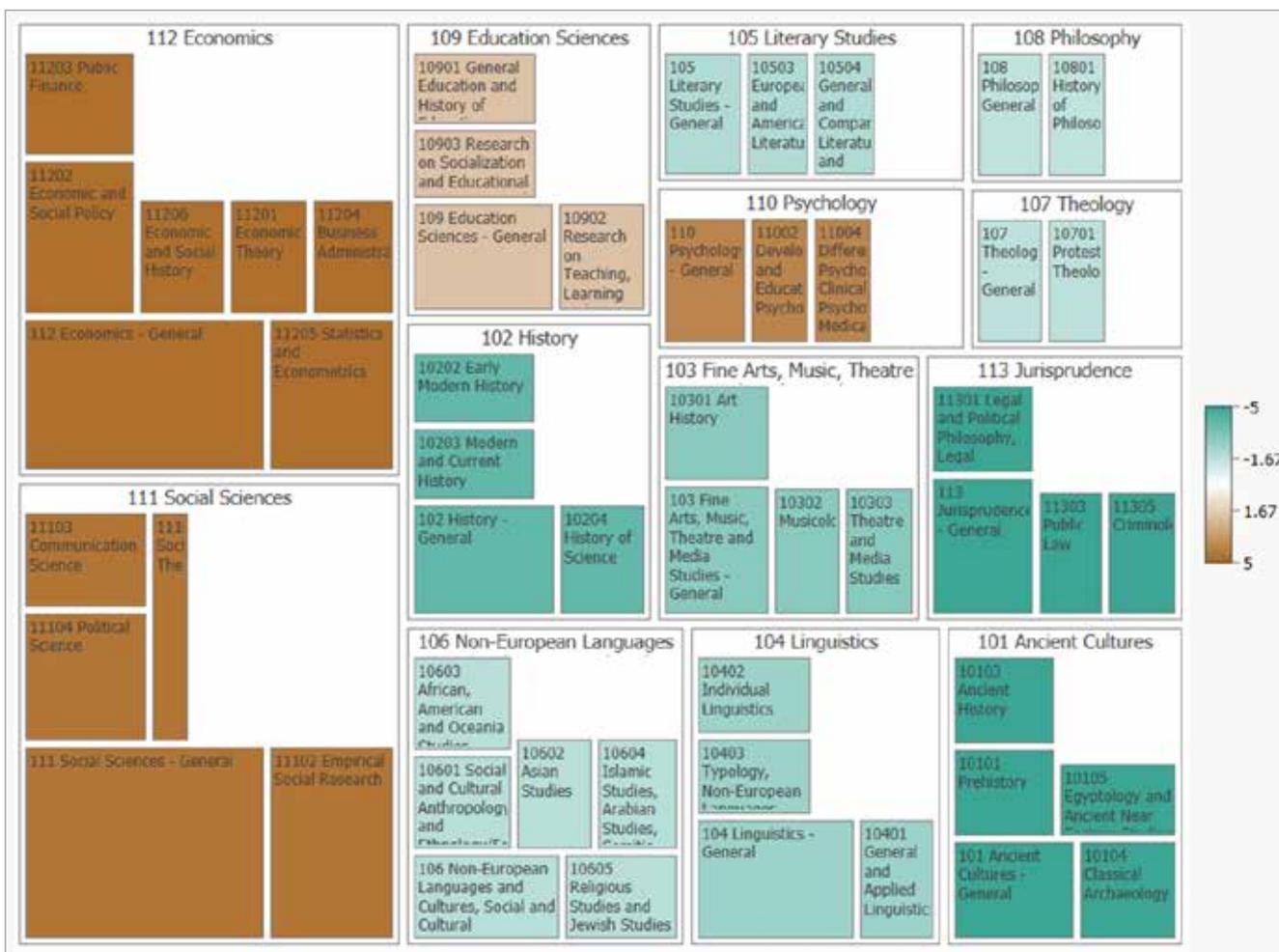


Figure 1. Proportional representation of the SSH repositories in *re3data*, including multidisciplinary repositories

Table 3. Metadata schemas in representative repositories of SSH

Repository	Metadata schema
<b>Social sciences</b>	
<i>Inter university Consortium for Political and Social Research ICPSR (EUA)</i> <a href="http://www.icpsr.umich.edu">http://www.icpsr.umich.edu</a>	DDI DC
<i>UK Data Service (Reino Unido)</i> <a href="https://www.ukdataservice.ac.uk">https://www.ukdataservice.ac.uk</a>	DDI, DC, ISO 19115, METS ( <i>Metadata encoding and transmission standard</i> ), ISAD ( <i>International standard archival description</i> )
<i>Gesis Zacat (Alemania)</i> <a href="http://zacat.gesis.org/webview">http://zacat.gesis.org/webview</a>	DDI DC
<b>Humanities</b>	
<i>Common Language Resources and Technology Infrastructure (Clarin, EU)</i> <a href="http://www.clarin.eu">http://www.clarin.eu</a>	IMDI ( <i>ISLE meta data initiative</i> ), TEI headers, DC, DCTerms, DC-OLAC ( <i>Open language archive community</i> ) ( <b>Van-Uytvanck; Stehouwer; Lampen, 2012</b> )
<i>Archaeology Data Service (Reino Unido)</i> <a href="http://archaeologydataservice.ac.uk">http://archaeologydataservice.ac.uk</a>	ADS Schema DC MIDAS
<i>Prometheus (Alemania)</i> <a href="http://www.prometheus-bildarchiv.de">http://www.prometheus-bildarchiv.de</a>	EDM ( <i>Europeana data model</i> ) METS DC

### 3.3. Metadata schemes used in SSH research data repositories

22.8% (332) of all *re3data* repositories specify the metadata scheme/s they use. As seen in figure 3, in the field of SSH (in lighter color) *Dublin Core* and DDI (*Data documentation initiative*) are by far the most used. The reason for "other" being the highest value is that it is a non-mandatory field in all versions of *re3data* scheme, and indicates the wide variety of metadata used in all disciplines, with a few dominant schemes in certain areas, and many specific variations in those disciplines in which no scheme stands out as dominant.

Moreover, 25.2% of SSH repositories declare some type of metadata schema. Of these, 45% use *Dublin Core*, the most common metadata model in 45 repositories. Both DDI and "other", are second with 37% each, used in 37 repositories.

It should be noted that "other" refers to homegrown metadata schemes (of the institution or of the laboratory). Both the graphic representation of the situation and the metadata schema name and number repositories that use them can be seen in figure 4. It is noteworthy that 74.8% of the repositories do not provide this information, and that multidisciplinary repositories may use more than one scheme, so it is possible to find schemes from other scientific areas.

In order to review the metadata schemes used in SSH, six representative repositories (table 2) were selected. We studied them identifying the metadata schema used, either by analyzing the repository, or looking at the repository site guidance or instructions for the deposit, or looking for papers on the repositories studied where this information was declared.

The detailed analysis of the selected repositories confirms

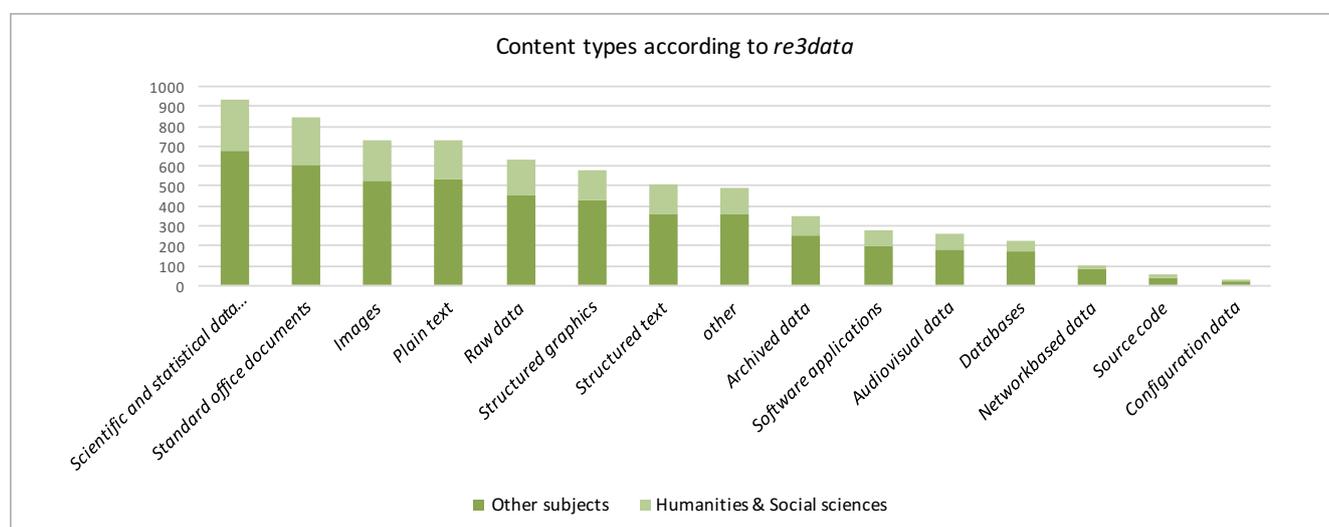


Figure 2. Types of content declared in *re3data*. In the vertical axis there is the number of repositories where each type of content was found.

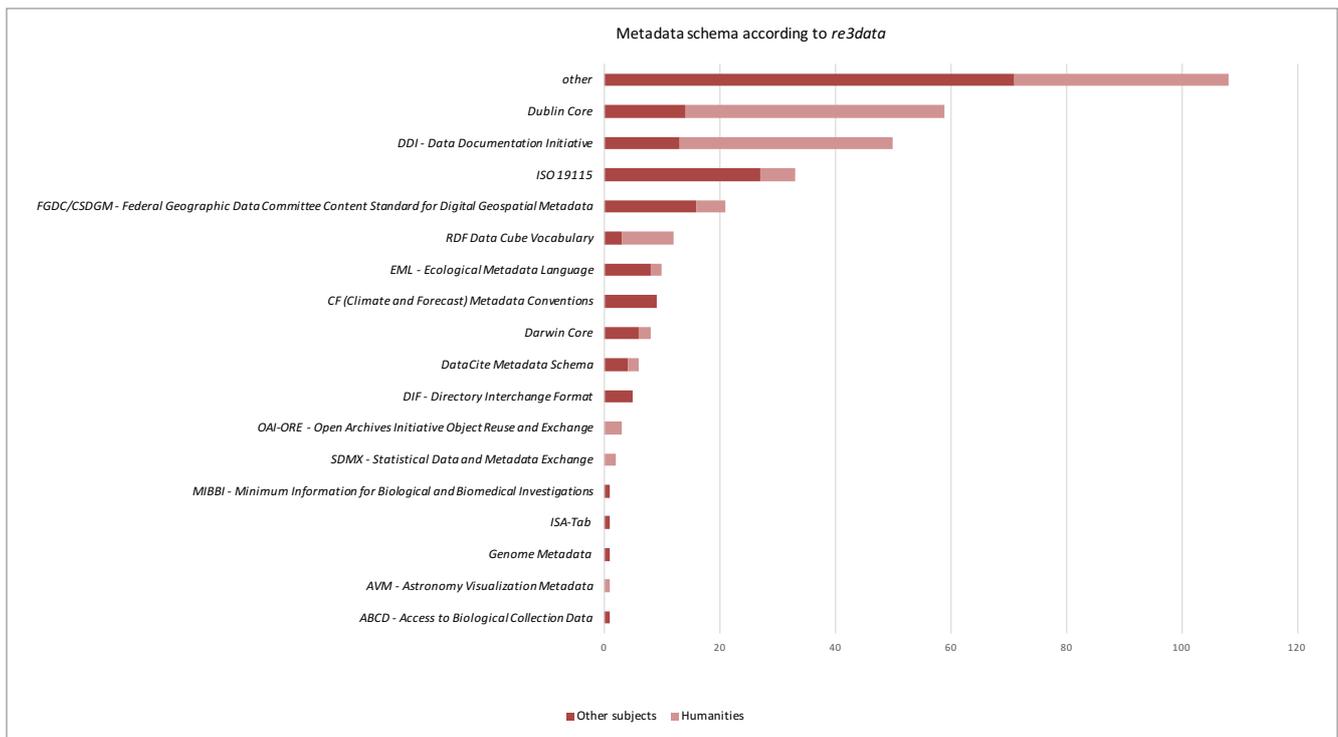


Figure 3. Metadata schemes declared by the data repositories included in re3data

the trend that re3data offers in social science: the dominant metadata scheme is DDI (*Data documentation initiative*), an international standard for describing statistical data and social science data with great tradition. DDI describes the data resulting from observation methods in social, behavioral, economic, and health sciences. It takes into account the data collection processes, the varying levels of description, and methods. It is a scheme that could be called classic, since it was originated in 1995, when the *Dublin Core* appeared within the social science community, and with the objective of describing data. Since then it has evolved steadily, maintained by the *DDI Alliance* (Vardigan, 2013). <http://www.ddialliance.org>

“The heterogeneity and complexity of research data repositories is manifested in the metadata schemes that are chosen to describe them, which is even more evident in the humanities”

In the case of the humanities, the metadata schemes used are more diverse and particular, as shown by the selected repositories analyzed in this article (table 3). However, most schemes are not found within the repositories registered in re3data. This explains the high percentage for the category “other” (37%) in the humanities repositories, because until version 3.0 of the re3data scheme, only metadata standards collected by the *Digital Curation Centre* were recognized as allowable values. <http://www.dcc.ac.uk/resources/metadata-standards>

This diversity of metadata, or lack of common or regular hu-

manities standards, is justified in the heterogeneity of data repositories and to what is considered as “data” in humanities, as discussed in the introduction.

Within SSH the use of *Dublin Core* (DC) is extensive (figure 4). This predominance is due to:

- a linkage with document / publications repositories; and a lack of distinction between these and data repositories, and
- the level of standardization that DC has attained and its interoperability OAI-PMH between repositories.

We agree with the argument given by Willis, Greenberg and White (2012) that creators of metadata schemes are more likely to change and adapt or enhance an existing scheme than to create a new one. Once the DC has been installed, it is easier to adapt than it is to adopt a new schema.

#### 4. Conclusions

The main conclusion we draw from this study is the corroboration of the heterogeneity and complexity of research data repositories, which is glaring within the humanities. This heterogeneity is manifested in the metadata schemes that researchers choose for description. We have reached several conclusions in the course of this work:

- 1) Following the merger between *Databib* and *re3data* in the same registry at the end of 2015, *re3data* has become the registry par excellence for finding research data repositories in all disciplines; which we used to identify and analyzed 397 repositories in SSH. Its greatest weakness, for now, is that it lacks mechanisms to know when a data repository record has been modified or how to change the characteristics initially declared. The information about the repositories cannot be updated online. Since February 2016 this

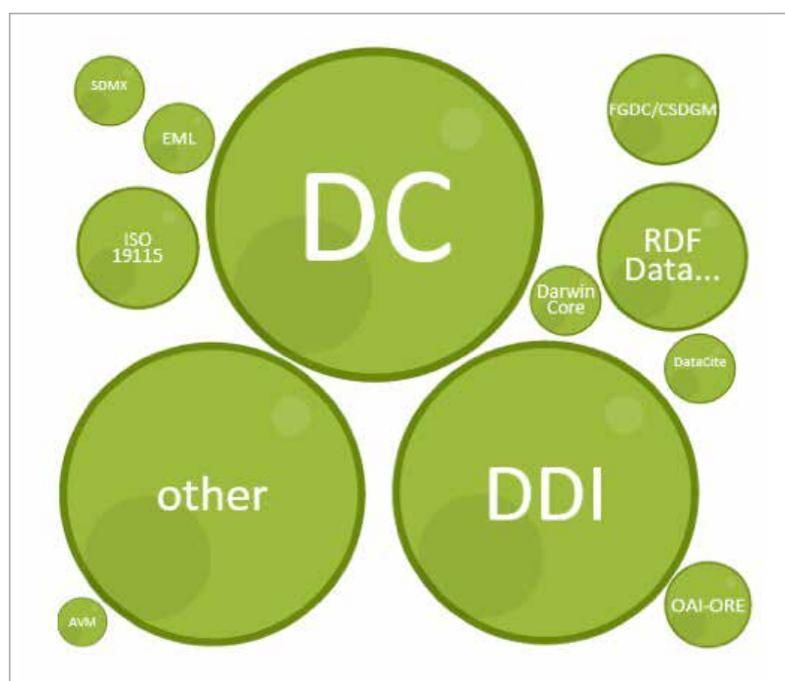


Figure 4. Metadata schemes used in SSH according to *re3data*

Scheme name	N.
Dublin Core	45
DDI	37
Other	37
RDF Data cube vocabulary	9
ISO 19115	6
FGDC/CSDGM	5
OAI-ORE	3
EML	2
Darwin Core	2
DataCite	2
SDMX	2
AVM	1

problem has been alleviated by sending a form to *re3data* requesting the needed changes. A manual mechanism that is, hopefully, temporary.

The metadata scheme used by *re3data* in its current version (v. 3.0) describes repositories and incorporates some characteristics about reuse, metrics and policies. This model seems to be evolving in the right direction if it does not include a large increase the existing set of characteristics. Automation mechanisms and online editing, as now happens with publications repositories and aggregators, should be implemented.

*Re3data (Registry of Research Data Repositories)* is the source of reference for identifying repositories to deposit research data classified by subject or discipline

The *DFG* thematic classification used by *re3data* is too generic. Therefore, it is not easy to narrow the theme of each repository because the vast majority are declared multidisciplinary, but often they are not, or they are multidisciplinary in a very small way.

2) Taking into account the limitations of *re3data* to describe the repositories, we can say that data and metadata schemas are less homogeneous in humanities than in social sciences. Despite the small number of data repositories that declare the metadata standard used, *re3data* confirms the trend of use of DDI metadata schema in social sciences. This may be due to the maturity of the standard, its amount of implementations, and that it was a scheme that was originally created to describe data, not documents. It is something

similar to the case of digital geospatial information systems, where, since the mid-90s, *FGDC (Federal Geographic Data Committee)* and *ISO 19115* standards have been used to describe geospatial data infrastructures.

The adoption of DDI by some of the most important repositories such as *Icpsr*, *Gesis*, and the *Dataverse* network of data repositories bode well for the future of metadata standards in social sciences, where “from” (and “for”) research data support statistics, surveys, opinion polls, etc., to which the DDI standard has been addressed from its inception.

3) In humanities the situation is more complex and diverse. *Dublin Core (DC)* seems to be widely used according to the generic data extracted from *re3data*, but if we drill down to the details of data repositories on specific fields such as linguistics or archeology, we see that they are using their own schemes or adapting DC to a greater or lesser extent. It should also be noted that many humanities projects, especially on text digitization, use *TEI Header* linked to the *TEI (Text encoding initiative)* standard, while in other cases they lack description schemes. The exposure of their research data is done simply through content managers with little use of metadata. Also it is not unusual to see metadata schemas used to describe humanities data in their data repositories, standards used for library creation, or for the description of textual publications, images, or audiovisuals (not only DC, but also EDM, METS, and MIDAS). This happens because of the tenuous differentiation in some of these disciplines, between data and documents, and between data “from” and “for” research.

4) *Dublin Core (DC)* is the default standard for publications’ repositories, and this trend includes data repositories, at least in the first instance or approach. Although DC has well established mechanisms to create application profiles that fit the description of any type of information or private co-

lection, it is still too early to confirm whether this standard can be adapted to the idiosyncrasies of all disciplinary research data.

## Notes

1. Web scraping (web harvesting or web data extraction) is a computer software technique for extracting information from websites.

*R* is a programming language and software for statistical computing and graphics supported by the *R Foundation for Statistical Computing*. It is widely used among statisticians and data miners for developing statistical software and data analysis.

2. Xslt (extensible stylesheet language transformations) is a language for transforming xml documents into other xml documents, or other formats such as html for web pages, plain text or into xsl formatting objects, which may subsequently be converted to other formats, such as pdf, postscript and png.

## Acknowledgements

This work is part of the project *Curator-e: custody and digital management for the reuse of open data research in the humanities and social sciences*, funded by the Spanish State Program of Research, Development and Innovation Facing Society Challenges, Ministerio de Economía y Competitividad (Mineco), Spain (CSO2013-46754-R).

## 5. References

**Bishoff, Carolyn; Johnston, Lisa** (2015). "Approaches to data sharing: An analysis of NSF data management plans from a large research university". *Journal of librarianship and scholarly communication*, v. 3, n. 2, p. eP1231. <http://dx.doi.org/10.7710/2162-3309.1231>

**Borgman, Christine L.** (2008). "Data, disciplines, and scholarly publishing". *Learned publishing*, v. 21, n. 1, pp. 29-38. <http://dx.doi.org/10.1087/095315108X254476>

**Borgman, Christine L.; Wallis, Jillian C.; Mayernik, Matthew S.** (2012). "Who's got the data? Interdependencies in science and technology collaborations". *Computer supported cooperative work (CSCW)*, v. 21, n. 6, pp. 485-523. <http://nldr.library.ucar.edu/repository/assets/osgc/OSGC-000-000-012-014.pdf> <http://dx.doi.org/10.1007/s10606-012-9169-z>

COM (2016) 178 final. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: European cloud initiative - Building a competitive data and knowledge economy in Europe. [http://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=15266](http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=15266)

European Commission (2016). *Guidelines on open access to scientific publications and research data in Horizon 2020*, v. 2.1. European Commission. Directorate General for Research and Innovation. [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

**Ferguson, Liz** (2014). "How and why researchers share data (and why they don't)". *Wiley Exchanges. Discover the future of research*, 3 November. <https://hub.wiley.com/community/exchanges/discover/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont?referrer=exchanges>

**Frank, Rebecca D.; Yakel, Elizabeth; Faniel, Ixchel M.** (2015). "Destruction/reconstruction: preservation of archaeological and zoological research data". *Archival science*, v. 15, n. 2, pp. 141-167. <http://dx.doi.org/10.1007/s10502-014-9238-9>

**Kim, Youngseek; Stanton, Jeffrey M.** (2016). "Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis". *Journal of the Association for Information Science and Technology*, v. 67, n. 4, pp. 776-799. <https://www.asis.org/asist2013/proceedings/submissions/papers/123paper.pdf> <http://dx.doi.org/10.1002/asi.23424>

LERU (2013). *LERU roadmap for research data*. Advice paper n. 14. [http://www.leru.org/files/publications/AP14\\_LERU\\_Roadmap\\_for\\_Research\\_data\\_final.pdf](http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf)

**Lyon, Liz** (2016). "Transparency: the emerging third dimension of open science and open data". *Liber quarterly*, v. 25, n. 4. <http://dx.doi.org/10.18352/lq.10113>

**Meadows, Alice** (2014). "To share or not to share? That is the (research data) question...". *The scholarly kitchen*, 11 November. <http://scholarlykitchen.sspnet.org/2014/11/11/to-share-or-not-to-share-that-is-the-research-data-question>

NEH (2015). *Data management plans for NEH Office of Digital Humanities. Proposals and awards*. [http://www.neh.gov/files/grants/data\\_management\\_plans\\_2015.pdf](http://www.neh.gov/files/grants/data_management_plans_2015.pdf)

NIH (2015). "NIH sharing policies and related guidance on NIH-funded research resources". *National Institutes of Health*. <https://grants.nih.gov/policy/sharing.htm>

NSF (2014). "Chapter II. Proposal preparation instructions". *Grant proposal guide*. National Science Foundation. Where discoveries begin. [http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg\\_2.jsp#dmp](http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#dmp)

OECD (2015). *Making open science a reality*. Organisation for Economic Co-operation and Development. [https://www.innovationpolicyplatform.org/sites/default/files/DSTI-STP-TIP%282014%299-REV2\\_0\\_0\\_0.pdf](https://www.innovationpolicyplatform.org/sites/default/files/DSTI-STP-TIP%282014%299-REV2_0_0_0.pdf)

**Paltoo, Dina N.; Rodriguez, Laura-Lyman; Feolo, Michael; Gillanders, Elizabeth; Ramos, Erin M.; Rutter, Joni L.; Sherry, Stephen; Wang, Vivian-Ota; Bailey, Alice; Baker, Rebecca; Caulder, Mark; Harris, Emily L.; Langlais, Kristofor; Leeds, Hilary; Luetkemeier, Erin; Paine, Taunton; Roomian, Tamar; Tryka, Kimberly; Patterson, Amy; Green, Eric D.** (2014). "Data use under the NIH GWAS data sharing policy and future directions". *Nature genetics*, v. 46, n. 9, pp. 934-938. <http://dx.doi.org/10.1038/ng.3062>

**Pepe, Alberto; Goodman, Alyssa; Muench, August; Crosas, Merce; Erdmann, Christopher** (2014). "How do astronomers share data? Reliability and persistence of datasets linked in AAS publications and a qualitative study of data practices among US astronomers". *PLoS one*, v. 9, n. 8, p. e104798.  
<http://dx.doi.org/10.1371/journal.pone.0104798>

PLoS (2014). "PLoS data policy prior to March 3, 2014". *PLoS*.  
<http://goo.gl/QIRlab>

**Rücknagel, Jessika; Vierkant, Paul; Ulrich, Robert; Kloska, Gabriele; Schnepf, Edeltraud; Fichtmüller, David; Reuter, Evelyn; Semrau, Angelika; Kindling, Maxi; Pampel, H.; Witt, Michael; Fritze, Florian; Van-de-Sandt, Stephanie; Klump, Jens; Goebelbecker, Hans-Jürgen; Skarupianski, Michael; Bertelmann, Roland; Schirmbacher, Peter; Scholze, Frank; Kramer, Claudia; Fuchs, Claudio; Spier, Shaked; Kirchhoff, Agnes** (2015). *Metadata schema for the description of research data repositories*, v. 3.0.  
<http://dx.doi.org/10.2312/re3.008>

**Tenopir, Carol; Allard, Suzie; Douglass, Kimberly; Aydinoglu, Arsev-Umur; Wu, Lei; Read, Eleanor; Manoff, Maribeth; Frame, Mike** (2011). "Data sharing by scientists: Practices and perceptions". *PLoS one*, v. 6, n. 6.  
<http://dx.doi.org/10.1371/journal.pone.0021101>

**Tenopir, Carol; Dalton, Elizabeth D.; Allard, Suzie; Frame, Mike; Pjesivac, Ivanka; Birch, Ben; Pollock, Danielle; Dorsett, Kristina** (2015). "Changes in data sharing and data reuse practices and perceptions among scientists worldwide". *PLoS one*, v. 10, n. 8, p. e0134826.  
<http://dx.doi.org/10.1371/journal.pone.0134826>

**Van-Uytvanck, Dieter; Stehouwer, Herman; Lampen, Lari** (2012). "Semantic metadata mapping in practice: the vir-

tual language observatory". En: *LREC 2012: 8<sup>th</sup> Intl conf. on language resources and evaluation. European Language Resources Association (ELRA)*, pp. 1029-1034.  
<http://goo.gl/IMgP4c>

**Vardigan, Mary** (2013). "Timeline DDI". *Iassist quarterly*, v. 37, pp. 51-55.  
[http://www.iassistdata.org/sites/default/files/iq/iqvol371\\_4\\_vardigan2.pdf](http://www.iassistdata.org/sites/default/files/iq/iqvol371_4_vardigan2.pdf)

**Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, Ij-sbrand-Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie; Blomberg, Niklas; Boiten, Jan-Willem; Da-Silva-Santos, Luiz-Bonino; Bourne, Philip E.; Bouwman, Jildau; Brookes, Anthony J.; Clark, Tim; Crosas, Mercè; Dillo, Ingrid; Dumon, Olivier; Edmunds, Scott; Evelo, Chris T.; Finkers, Richard; González-Beltrán, Alejandra; Gray, Alasdair J.G.; Groth, Paul; Goble, Carole; Grethe, Jeffrey S.; Heringa, Jaap; Hoen, Peter A.C't; Hooft, Rob; Kuhn, Tobias; Kok, Ruben; Kok, Joost; Lusher, Scott J.; Martone, Maryann E.; Mons, Albert; Packer, Abel L.; Person, Bengt; Rocca-Serra, Philippe; Roos, Marco; Van-Schaik, Rene; Sansone, Susanna-Assunta; Schultes, Erik; Sengstag, Thierry; Slater, Ted; Strawn, George; Swertz, Morris A.; Thompson, Mark; Van-der-Lei, Johan; Van-Mulligen, Erik; Velterop, Jan; Waagmeester, Andra; Wittenburg, Peter; Wolstencroft, Katherine; Zhao, Jun; Mons, Barend** (2016). "The FAIR guiding principles for scientific data management and stewardship". *Scientific data*, v. 3, p. 160018.  
<http://dx.doi.org/10.1038/sdata.2016.18>

**Willis, Craig; Greenberg, Jane; White, Hollie** (2012). "Analysis and synthesis of metadata goals for scientific data". *Journal of the American Society for Information Science and Technology*, v. 63, n. 8, pp. 1505-1520.  
<http://dx.doi.org/10.1002/asi.22683>



RecBib  
Recursos Bibliotecarios

Información sobre empleos, cursos, oposiciones, noticias, eventos y más cosas del mundo de las bibliotecas, archivos, documentación e información  
¡Descúbreanos!

[www.recbib.es](http://www.recbib.es)  [fb.me/recbib](https://fb.me/recbib)  [@recbib](https://twitter.com/recbib)



**Digitalización** enriquecida  
de fondo antiguo y patrimonial

**Software** de gestión para  
Bibliotecas, Archivos y Museos

DIGIBIB\*

DIGIARCH\*

DIGIMUS\*

**Recolector OAI-PMH** DIGIHUB\*  
de metadatos de diversos proveedores



DIGIBÍS transforma  
y enriquece las bibliotecas  
con aplicaciones de gestión  
digital de nueva generación.



El usuario accede  
fácilmente a los recursos  
de la biblioteca desde cualquier  
lugar, en cualquier momento,  
con cualquier dispositivo.



La biblioteca es  
recolectada automáticamente  
por Hispana, Europea  
y otros agregadores  
internacionales.



¡Con estándares internacionales  
para un **mundo enlazado y abierto!**



# GRADO DE CUMPLIMIENTO DE LAS LEYES DE TRANSPARENCIA, ACCESO Y BUEN GOBIERNO Y DE REUTILIZACIÓN DE LOS DATOS DE CONTRATACIÓN DE LA ADMINISTRACIÓN CENTRAL ESPAÑOLA

Degree of compliance with the *Laws of Transparency, access, and good governance and of the Reuse of procurement data* from the Spanish central government



**Pilar Beltrán-Orenes y Esther Martínez-Pastor**



**Pilar Beltrán-Orenes** es licenciada en filosofía y doctora en filosofía y metodología de la ciencia. Trabaja como profesora de documentación informativa en la *Universidad Rey Juan Carlos*, en Madrid. Ha sido profesora de documentación en las universidades de *La Coruña*, *Salamanca*, *Carlos III* y *Complutense de Madrid*. En los últimos años ha trabajado en proyectos de representación y recuperación de la información, así como en diseño y desarrollo de herramientas documentales de aprendizaje virtual.

<http://orcid.org/0000-0002-1408-6365>

[pilar.beltran@urjc.es](mailto:pilar.beltran@urjc.es)



**Esther Martínez-Pastor** es profesor contratado doctor de la *Facultad de Ciencias de la Comunicación* de la *Universidad Rey Juan Carlos* en Madrid. Es doctora y licenciada en ciencias de la información por la *Universidad Complutense de Madrid* y licenciada en derecho por la *Universidad Nacional de Educación a Distancia*. Sus líneas de investigación se centran en el derecho de la publicidad, en la publicidad institucional de las administraciones públicas y en menores.

<http://orcid.org/0000-0002-2861-750X>

[esther.martinez.pastor@urjc.es](mailto:esther.martinez.pastor@urjc.es)

*Universidad Rey Juan Carlos, Facultad de Ciencias de la Comunicación  
28943 Fuenlabrada (Madrid), España*

## Resumen

Comprobación del grado de cumplimiento de las leyes de transparencia, acceso y buen gobierno, y la reutilización de los datos de contratación en el sector público de la Administración central española. Se toma como punto de partida la descripción de las normas jurídicas más relevantes en relación a este tema. Se identifican los sujetos y la información relativa a la contratación que están obligados a mostrar de una forma accesible para poder ser reutilizada por cualquier interesado, así como, las plataformas en las que se alojan dichos datos públicos. Para ello, se han tomado como muestra los datos la *Plataforma de la Contratación del Sector Público del Ministerio de Hacienda y Administraciones Públicas* relativa a la contratación de la Administración central durante el mes de marzo de 2016. Los resultados más relevantes son que se aplica la ley en cuanto a la reutilización de la información pero, por el contrario, el acceso a la información es difícil y casi la mitad de los sujetos obligados no ofrecen los datos de contratación.

## Palabras clave

Administración pública; *Open government*; Transparencia administrativa; Información del sector público; Difusión de la información; Acceso a la información; Reutilización de la información; *Open data*; *Big data*; Contratación pública.

## Abstract

This study looks at how and if the Spanish central government is currently meeting the laws of transparency and good governance by providing access to public data. The first part is a description of the most relevant legal standards relating to this topic. Subjects, information on recruitment, formats available to reuse information, and platforms on which these public data are published are identified. This study was performed using March 2016 data from the *Public Platform for Hiring* which

is provided by the *Ministry of Finance and Public Administration* of Spain. The most important finding may be that the law has been correctly applied by the Spanish government for the reuse of information. On the other hand, accessing information is difficult and nearly half of the subjects do not offer hiring data despite this requirement.

## Keywords

Public administration; Open government; Administrative transparency; Public information; Dissemination of information, Access to the information; Reuse of information; Open data; Big data; Contracts public.

**Beltrán-Orenes, Pilar; Martínez-Pastor, Esther (2016).** "Grado de cumplimiento de las *Leyes de transparencia, acceso y buen gobierno*, y de reutilización de los datos de contratación de la Administración central española". *El profesional de la información*, v. 25, n. 4, pp. 557-567.

<http://dx.doi.org/10.3145/epi.2016.jul.05>

## 1. Introducción

Dos años después de la entrada en vigor de *Ley 19/2013*, de 9 de diciembre, de *transparencia, acceso a la información pública y buen gobierno (LTBG)*, la estructuración del acceso a la información a la que obliga sigue siendo una asignatura pendiente.

A la falta de concreción de la norma jurídica en lo relativo a aspectos tan importantes como las sanciones por su incumplimiento, o al propio estatuto o naturaleza de la misma, se suman la inoperancia de los mecanismos habilitados para su cumplimiento y su dispersión, cuando los hay.

En este artículo se ha combinado el análisis jurídico con un estudio de los datos ofrecidos por la administración central/ estatal a día de hoy, a los que obliga la *LTGB*, del que se extraen los múltiples fallos de la norma jurídica, tanto en el plano estrictamente jurídico como en el práctico de acceso y análisis de datos.

## 2. Antecedentes: posiciones doctrinales, normas jurídicas y jurisprudencia

El 9 de diciembre de 2013 entró en vigor la *Ley 19/2013*, *Ley de transparencia, acceso a la información pública y buen gobierno (LTBG)*. Es una norma jurídica consecuencia de la inercia "histórica, política, social y, por supuesto, jurídica de la emergencia de la transparencia y el acceso a la información" (Cotino-Hueso, 2014, p. 242) que ha permitido considerar el derecho de acceso a la información pública como un derecho fundamental.

Sin embargo, tal como señala Rollnert-Liern (2015) hay tres posiciones doctrinales al respecto:

- la primera considera que el derecho de acceso es un derecho constitucional de configuración legal ya que está recogido en la *Constitución española* sin ser un derecho fundamental (Fernández-Ramos, 2012; Chinchilla-Martín, 1988);
- la segunda mantiene una interpretación intermedia porque considera que es tanto un derecho autónomo como un derecho fundamental asociado a la libertad de información (Guichot-Reina, 2014);
- la postura de los que defienden la naturaleza del derecho de acceso como la facultad de investigar y de buscar que se integra en el artículo 20 de la *Constitución*, además de

ser recogido por el artículo 105b cuando se trata de archivos y registros administrativos y de estar aceptado a nivel internacional (Sánchez-de-Diego, 2015; Cotino-Hueso, 2015; Olmedo-Palacios, 2014).

En cualquier postura, lo que está claro es que en toda sociedad democrática es necesario e imprescindible el derecho de acceso a la información para que se pueda configurar una opinión pública (Rollnert-Liern, 2015).

La última posición doctrinal, se justifica tanto en la jurisprudencia internacional, iberoamericana y europea (Cotino-Hueso, 2014; Sánchez-de-Diego, 2014) como por la *Carta de los derechos fundamentales de la Unión Europea* (artículo 40).

Respecto a la jurisprudencia internacional que reconoce el derecho a acceder a la información pública como un derecho fundamental cabe destacar el caso *Claude Reyes vs Chile*, de la *Corte Iberoamericana* de 2006, en que el que se reconoció el acceso a la información pública como derecho incluido en la libertad de expresión. En la misma línea se encuentran las sentencias del *Tribunal Europeo de Derechos Humanos* de los casos:

- *Társaság a Szabadságjogokért vs Hungría* (sentencia de 14 de abril de 2009);
- *Kenedi vs Hungría* (sentencia de 26 de agosto de 2009);
- *Youth initiative for human rights vs Serbia* (de 25 de junio de 2013).

En los tres casos se alude a la libertad de recibir información y el derecho de acceso a la información. De hecho, la *Carta de los derechos fundamentales de la Unión Europea*, adquiere un valor jurídico con el *Tratado de Lisboa*, en vigor desde 2009, considerando que el derecho de acceso a la información pública es de naturaleza *iusfundamental* tanto en el derecho a una buena administración como en el derecho de acceso a los documentos (Cotino-Hueso, 2014).

En España, hasta el año 2013 no se disponía de una ley *ad hoc* sobre transparencia o derecho de acceso. No se ha regulado como ley orgánica por no entenderse que es un derecho fundamental (Cotino-Hueso, 2014; Rollnert-Liern, 2014), a pesar de las opiniones doctrinales como anteriormente se ha indicado. Hasta ese momento, quedaba recogido en la *Constitución*, aunque fuera de los derechos fundamentales (artículo 105), limitando el acceso de los ciudadanos a los

archivos y registros administrativos y siempre y cuando no afecte la información de los mismos a la seguridad y defensa del Estado, la averiguación de los delitos y la intimidad de las personas (**Moretón-Toquero**, 2014). Este acceso estaba recogido en la *Ley 30/1992, del régimen jurídico de las administraciones públicas y del procedimiento administrativo común (LRJPAC)* y de leyes sectoriales sobre determinadas materias (**Ares-González**, 2014; **Piñas-Mañas**, 2014). Además, se contaba con la *Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público*, cuyo objeto es la regulación jurídica aplicable a la reutilización de los documentos elaborados, que fue modificada en 2015 (*España*, 2007a; 2015).

No obstante, el concepto real de transparencia consiste en entender la información pública como materia prima y su valor fundamental para ser utilizado por otras organizaciones que no sean solamente las administraciones (**Cotino-Hueso**, 2015; **Ares-González**, 2014). Es decir, la información es un valor en sí mismo y su reutilización por parte de otros es su valía real. Ahí es donde debe de incluirse el derecho de acceso y la transparencia del buen gobierno. Por este motivo nuestro ordenamiento jurídico, tanto por su contenido como por ser una normativa desflecada y fragmentada, se quedaba corto frente a las legislaciones europeas de transparencia que incorporaban en sus textos legales el derecho de acceso a la información y su reutilización (**Cotino-Hueso**, 2014; **Ares-González**, 2014). Por ello, se propuso y entró en vigor la citada *Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno (LTBG)* (*España*, 2013) y la posterior reforma de 2015 de la *Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público* (*España*, 2015).

La información es un valor en sí mismo y su reutilización por parte de otros es su valía real. Ahí es donde debe de incluirse el derecho de acceso y la transparencia del buen gobierno

### 3. LTBG: open government, transparencia, acceso e información pública

La aprobación de la *LTBG* respondía a los compromisos adquiridos con la *Alianza por el gobierno abierto (open government)* y con la *Alianza por el gobierno abierto de los documentos oficiales del Consejo de Europa* (**Cotino-Hueso**, 2015 y **Sánchez-de-Diego**, 2014). Este término fue incluido en el discurso de Obama en 2009, en el que afirmaba que su administración sería transparente con sus ciudadanos (**Ferrer-Sapena**; **Peset**; **Aleixandre-Benavent**, 2011). Este hecho generó un gran impulso mediático en cuanto a la forma de entender la gestión de los gobiernos y, en 2011, se creó la *Alianza para el gobierno abierto*.  
<http://www.opengovpartnership.org/es>

En esta plataforma internacional se promueve que los gobiernos sean más abiertos y mejoren su capacidad de respuesta hacia sus ciudadanos. Cuenta con 64 países para llevar a cabo

e implementar reformas ambiciosas en torno al gobierno abierto, entre los que está España, que ha participado con un informe sobre el gobierno abierto en España en 2014.  
<http://www.opengovpartnership.org/country/spain>

Sus objetivos principales son:

- rendición de cuentas
- tecnología e innovación
- participación ciudadana
- transparencia.

No se trata sólo de poner en marcha sistemas de administración electrónica que la acerque a los ciudadanos sino que se generen procedimientos para que éstos interactúen con sus gestores públicos y que se pueda reutilizar la información pública y “explorar las posibilidades de los *big data* o del *data mining* respecto de todos esos datos” (**Boix**, 2015, p. 242). Sobre esta filosofía del *open government* y los *open data* se fundamenta nuestra *Ley 19/2013*. Pero antes de analizar esta norma jurídica conviene reflexionar sobre estos términos: transparencia, acceso e información pública.

La transparencia incluye el derecho de acceso a la información pública. Transparencia no es lo mismo que derecho de acceso a la información. El derecho de acceso a la información pública, configurado como derecho fundamental, garantiza la transparencia, pero ésta por sí sola no amplía los derechos sino que necesita del acceso (**Sánchez-de-Diego**, 2013).

Por su parte, el derecho de acceso está vinculado directamente con el derecho a la información, donde se entiende implícitamente la facultad de buscar e investigar, ya que para informar previamente se ha tenido que acceder a una información. Se puede entender que el derecho de acceso es un derecho fundamental por la *Declaración universal de los derechos humanos* y el *Pacto internacional de derechos civiles y políticos* que reconocen la facultad de investigar o buscar (**Sánchez-de-Diego**, 2014; **Olmedo-Palacios**, 2014).

Por documento público, que se aplica por extensión a la información pública, se entiende que se refiere a los documentos, contenidos o informaciones registradas, redactadas o recibidas en poder de las autoridades públicas (*Convenio del Consejo de Europa sobre el acceso a los documentos oficiales*).

### 4. LTBG: sujetos obligados, información y formas de acceso a la información pública (Portal de la transparencia)

Esta norma jurídica ha querido responder a sus objetivos principales de incrementar y reforzar la transparencia, garantizar el acceso a la información pública y establecer las obligaciones de un buen gobierno, tal y como establece el preámbulo de la presente *Ley*. De hecho, el anteproyecto de la *LTBG* fue presentado en una página *ad hoc* con el objetivo de que desde el principio fuera accesible, transparente y participativa con los ciudadanos ya que éstos podían remitir cualquier comentario o sugerencia sobre la misma en un plazo de 15 días y hubo unas 3.600 contribuciones (**Cotino-Hueso**, 2014).

<http://transparencia.gob.es>

Estos comentarios se sometieron a una comisión de expertos en el *Centro de Estudios Políticos y Constitucionales (CEPC)* y posteriormente se presentó el proyecto al *Congreso de los Diputados*. Finalmente, se aprobó la *Ley 19/2013*. Hay que destacar que, además de esta norma jurídica, se hizo el desarrollo normativo del *Real decreto 919/2014*, de 31 de octubre, por el que se aprueba el *Estatuto del Consejo de Transparencia y Buen Gobierno*.

“Sobre la filosofía del *open government* y los *open data* se fundamenta nuestra *Ley 19/2013*, de 9 de diciembre, de *transparencia, acceso a la información pública y buen gobierno*”

En su primer artículo, la *Ley* presenta su ámbito de aplicación y dice así:

“La *Ley* implica la regulación general de la transparencia de la actividad pública y el acceso a la información pública”.

Este precepto es claro y conciso, mientras que el ámbito subjetivo de aplicación es muy amplio y ambicioso porque incluye a sujetos diversos:

- sujetos *di natura* de la administración pública;
- sujetos que están sujetos a derecho administrativo;
- empresas privadas con una participación pública de más del 50% y partidos políticos;
- organizaciones sindicales y empresariales que reciban subvenciones públicas.

### 1) Sujetos que son *di natura* de la administración pública

- Administración General del Estado, administraciones de las comunidades autónomas y de las ciudades de Ceuta y Melilla, y las entidades que integran la administración local;
- organismos autónomos, agencias estatales, entidades públicas empresariales y entidades de derecho público, incluidas las universidades públicas;
- entidades gestoras y servicios comunes de la *Seguridad Social*, así como las mutuas de accidentes de trabajo y enfermedades profesionales colaboradoras de la *Seguridad Social*;
- corporaciones de derecho público, en lo relativo a sus actividades sujetas a derecho administrativo;
- organismos autónomos, agencias estatales, entidades públicas empresariales y entidades de derecho público;
- fundaciones del sector público previstas en la legislación en materia de fundaciones;
- asociaciones constituidas por las administraciones, organismos y entidades previstos en este artículo.

### 2) Sujetos que están sujetos a derecho administrativo

- la *Casa de su Majestad el Rey*, el *Congreso de los Diputados*, el *Senado*, el *Tribunal Constitucional* y el *Consejo General del Poder Judicial*, así como el *Banco de España*, el *Consejo de Estado*, el *Defensor del Pueblo*, el *Tribunal de Cuentas*, el *Consejo Económico y Social* y las instituciones

autónomas análogas, en relación con sus actividades sujetas a derecho administrativo.

### 3) Empresas privadas con una participación pública del más del 50%

- empresas privadas como sociedades mercantiles, fundaciones o en cuyo capital social la participación, directa o indirecta, de las entidades públicas previstas en este artículo sea superior al 50%.

### 4) Partidos políticos, organizaciones sindicales y empresariales que reciban subvenciones públicas

- partidos políticos, organizaciones sindicales y organizaciones empresariales;
- entidades privadas que perciban durante el período de 1 año ayudas o subvenciones públicas en una cuantía superior a 100.000 euros o cuando al menos el 40% del total de sus ingresos anuales tengan carácter de ayuda o subvención pública, siempre que alcancen como mínimo la cantidad de 5.000 euros.

El presente estudio se centra en los sujetos obligados a disponer la información accesible en materia de contratación, y que aparecen recogidos en el inventario de entidades de la administración central.

“Es imposible saber si los organismos públicos estatales están haciendo publicidad de todos los procesos de contratación a los que están obligados”

En cuanto a las formas de acceso a la información pública, la publicidad activa, es importante indicar que la *Ley* dispone el cómo y el dónde se accede y qué información es susceptible de esta norma jurídica (Cotino-Hueso, 2014; Ares-González, 2014).

La información debe publicitarse de forma periódica y actualizada (artículo 5.1), de forma clara, estructurada y entendible para los interesados y en formatos reutilizables (artículo 5.4), y su acceso debe ser fácil y gratuito (artículo 5.5) a través del *Portal de la transparencia* (artículos 11 y 10).

Se hace alusión a cómo debe ofrecerse la información pública y de qué manera puede el ciudadano reutilizarla a través de formatos que lo permitan, para un posible tratamiento e interpretación con otros datos.

En cuanto al dónde, la información debe ser accesible tanto en las webs de los organismos sujetos a la *Ley* como en el *Portal* (artículo 10 y 11).

En relación al qué, se refiere al tipo de información que está sujeta a este derecho de acceso y, por ende, a la transparencia. De hecho el artículo 13 entiende que es información pública todos los contenidos o documentos, cualquiera que sea su formato o soporte, que hayan sido elaborados o adquiridos en el ejercicio de las funciones de los sujetos obligados por esta norma jurídica. Los preceptos 6 y 7 expresan que deben publicarse: la información institucional, organizativa y de planificación; la relativa a las funciones,

normativa, estructura, organigrama, planes y los programas, actividades, medios y tiempo para acometerlas, así como grado de cumplimiento y resultados (artículo 6).

También son susceptibles de ser publicadas las directrices, instrucciones, acuerdos, circulares o respuestas a consultas planteadas por los particulares u otros órganos en la medida en que supongan anteproyectos de ley, etc. (artículo 7).

Del mismo modo debe publicarse la información en materia económica (artículo 8). Este precepto dispone que debe hacerse pública toda la información relativa a: contratos, convenios suscritos, subvenciones y ayudas públicas concedidas, presupuestos, cuentas anuales, retribuciones percibidas anualmente por los altos cargos y máximos responsables de las entidades, resoluciones de autorización o reconocimiento de compatibilidad que afecten a los empleados públicos, declaraciones anuales de bienes y actividades de los representantes locales, información estadística necesaria para valorar el grado de cumplimiento y calidad de los servicios públicos, contratos o convenios y subvenciones celebrados con una administración pública y la relación de los bienes inmuebles que sean de su propiedad o sobre los que ostenten algún derecho real.

En concreto, en materia de contratación (licitaciones), motivo de este análisis, los datos que deben publicarse son, según el artículo 8.1.a) de la *LTBG* y el artículo 26 del *Real decreto legislativo 3/2011*, de 14 de noviembre, por el que se aprueba el texto refundido de la *Ley de contratos del sector público*: indicación de las partes, objeto, duración, importe de licitación y de adjudicación del contrato, así como, el procedimiento utilizado para su adjudicación, modificaciones del contrato, acciones de desistimiento y renuncia, si las hubiese. Sin embargo, todos estos datos en las licitaciones no siempre se cumplen en el *Portal de la transparencia* ya que en unos casos se incluyen unos y se excluyen otros como sucede con los datos de publicación o la duración de las licitaciones, que no suele constar.

## 5. El Portal de la transparencia del Gobierno de España y la Plataforma de contratación del sector público (PCSP)

La web que ha creado el Estado español para garantizar el cumplimiento de la *Ley de transparencia* en el ámbito estatal es el *Portal de la transparencia* del Gobierno de España. El apartado de contrataciones, objeto de este estudio, está estructurado como un directorio, y sólo tiene un filtro por fechas. Permite navegar por los 13 ministerios, más Jefatura del Estado y Presidencia del Gobierno, pero no permite la búsqueda por tipo de entidad, nombre de la entidad, etc. Aunque ofrece la información requerida por ley en la publicidad de los procesos de contratación, no permite filtrar la información de las entidades que se encuentran adscritas a cada ministerio, las presenta todas juntas, sin posibilidad de filtrar los resultados más que por el rango de fecha. Por otra parte, no permite la descarga de los resultados para filtrarlos por otros criterios. Es imposible, por tanto, realizar un análisis del grado de seguimiento que está teniendo la *Ley de transparencia* en contratación desde el *Portal de la transparencia*. Además, éste ha recibido quejas por su efec-

to disuasorio sobre el derecho a la información, que ha reconocido el *Defensor del Pueblo* (*Tu derecho a saber*, 2015).

En materia de contratación existe un servicio creado por el *Ministerio de Hacienda y Administraciones Públicas* en cumplimiento del artículo 309 de la *Ley 30/2007*, de 30 octubre, de *contratos del sector público (LCSP)* (España, 2007b): la *Plataforma de contratación del sector público (PCSP)*. En dicha plataforma, y según reza en su página:

“podrá encontrar la información necesaria para contratar con cualquier organismo del sector público mediante los servicios que ésta pone a su disposición. La información sobre las licitaciones publicadas en cualquier perfil del contratante del sector público puede consultarse desde esta *Plataforma*” (*MHAP*, 2016).

El propio *Portal*, al entrar en las licitaciones (contratos), remite al perfil contratante de la *Plataforma* desde el apartado “Para más información” de sus anuncios de contratación.

En ella, además, pueden publicitar sus contrataciones las entidades u organismos autónomos y locales obligados y que no puedan o no quieran hacerlo, según recoge el texto de la *Ley*, por sus propios medios.

Esta plataforma recoge 19 tipos de entidades que están obligadas a cumplir con la transparencia, en los que, no obstante, no se encuentran todos los enumerados en dicha *Ley* (tabla 1).

Esa tipología es genérica e incluye a los 3 tipos de administración: estatal/central, autonómica y local. Sin embargo, no están desglosadas todos los tipos específicos de la administración central, objeto de este estudio.

Tabla 1. Entidades obligadas a cumplir con la transparencia

	Tipo de administración, entidad u organismo
1	Administración General del Estado: Ministerios
2	Administraciones de comunidades autónomas
3	Ciudades autónomas
4	Ayuntamientos
5	Diputaciones provinciales
6	Cabildos y <i>Consell Insular</i>
7	Comarcas
8	Mancomunidades (983)
9	Agrupaciones de municipios
10	Áreas metropolitanas
11	Entidades locales menores (eatimes)
12	Entidades gestoras y servicios comunes de la <i>Seguridad Social</i>
13	Organismos autónomos
14	Entidades públicas empresariales
15	Universidades públicas
16	Agencias estatales
17	Otras entidades de derecho público
18	Sociedades mercantiles
19	Consortios

Elaborado a partir de la *Plataforma de contratación del sector público* del *Ministerio de Hacienda y Administraciones Públicas*

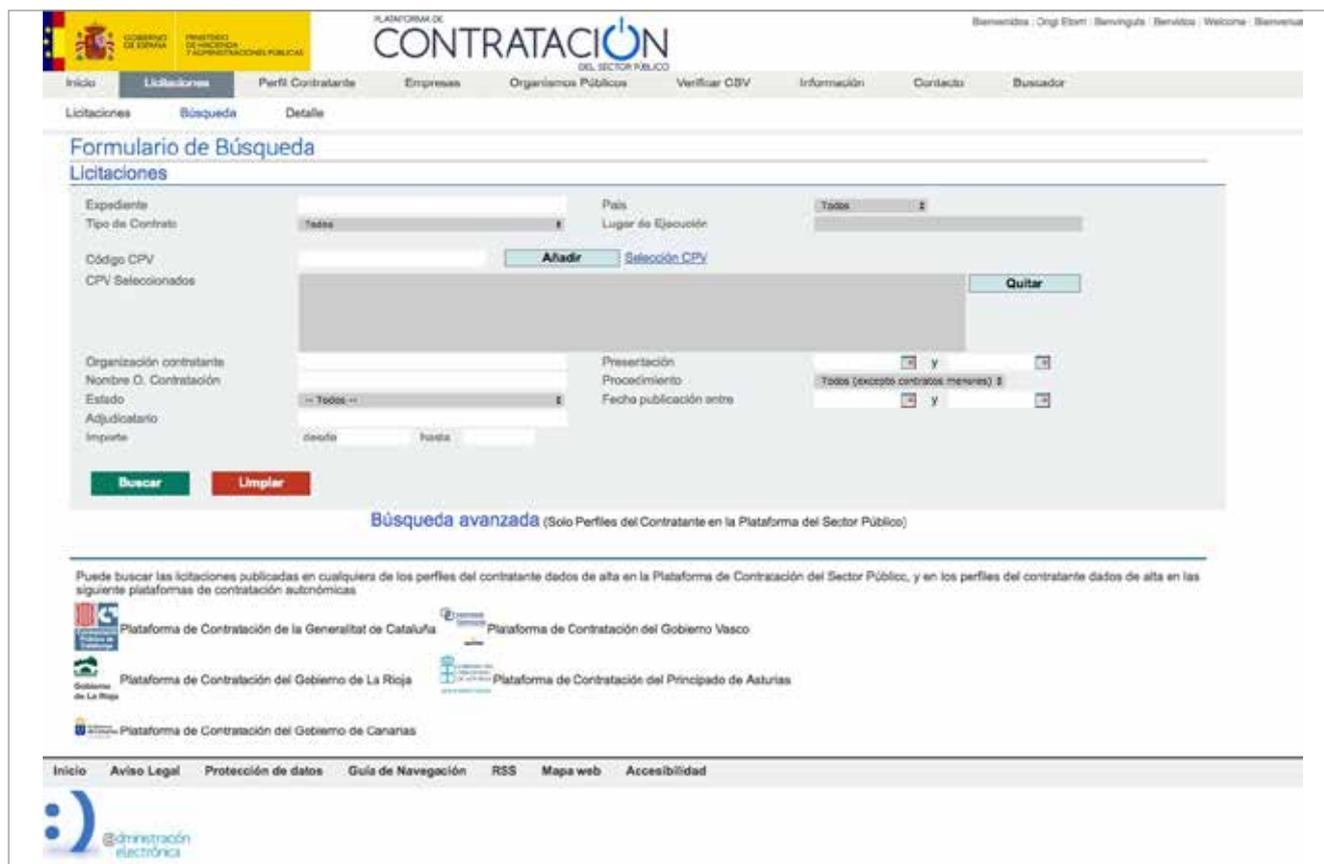


Imagen 1. Plataforma de contratación del sector público <https://contrataciondelestado.es/wps/portal/plataforma>

La opción de búsqueda ofrece un directorio guiado por los criterios de “Estado” (de la licitación) y “Tipo de contrato”, y un formulario de búsqueda con filtros y combinación de los mismos (imagen 1).

En todos los casos, la información ofrecida a través de la *Plataforma* se ofrece en formato html y permite la descarga de los resultados, es reutilizable.

Dada la imposibilidad de hacer el estudio desde el *Portal* y puesto que la *Plataforma* sí permite un mayor filtrado de los resultados, está publicando constantemente licitaciones (está actualizada), y el primero remite a la segunda para ampliar información, se ha elegido la segunda para llevar a cabo el análisis propuesto.

## 6. Selección de la muestra<sup>1</sup>

El primer problema ha sido realizar el inventario de las entidades centrales/estatales desde la *Plataforma*. Además de aparecer entidades de los tres tipos de administración (central, autonómica y local), no se ofrece dicha información en formato reutilizable. Por esta razón se ha hecho la recopilación de entes públicos desde los inventarios *Entes del sector público* y *Entes del sector público estatal* con los datos recogidos en ambos de las consultas hechas entre el 14 y el 18 de marzo de 2016.

Elaborados ambos por la *Intervención General de la Administración del Estado* del *Ministerio de Hacienda y Administraciones Públicas*, la información se ofrece a través de los buscadores:

- *Invente* (*Inventario de entes del sector público*);
- *Invespe* (*Inventario de entes del sector público estatal*).

Tabla 2. Tipos de entidades estatales

	Administración estatal/central
1	Administración General del Estado: ministerios
2	Agencias estatales
3	Consortios estatales
4	Entidades gestoras y servicios comunes de la <i>Seguridad Social</i>
5	Entidades públicas empresariales estatales
6	Fondos carentes de personalidad jurídica estatales
7	Fundaciones
8	Mutuas de accidentes de trabajo y enfermedades profesionales colaboradoras de la <i>Seguridad Social</i>
9	Organismos autónomos estatales (todos dependen de algún ministerio)
10	Organismos que no pertenecen a la Administración del Estado, diferenciados en los <i>Presupuestos generales del Estado (PGE)</i> <sup>3</sup>
11	Otras entidades de derecho público
12	Otros entes participados por el Estado: consorcios del ámbito autonómico o local
13	Otros entes participados por el Estado: consorcios del sector privado
14	Sociedades mercantiles estatales y asimiladas

Elaborado a partir de los listados obtenidos de *Invente* e *Invespe* de la *Intervención General de la Administración del Estado* (IGAE, 2016a; 2016b)

Tabla 3. Muestra estratificada y resultados del análisis. En todos los casos el formato obtenido fue html

Administración estatal/central	Universo de discurso	Tamaño de la muestra	Nº entidades que publican en la PCSP	Seguimiento %
Administración General del Estado: ministerios	13	9	9	100,0
Agencias estatales	10	7	4	57,0
Consortios estatales	18	12	8	66,6
Entidades gestoras y servicios comunes de la <i>Seguridad Social</i>	5	3	3	100,0
Entidades públicas empresariales estatales	13	9	9	100,0
Fondos carentes de personalidad jurídica estatales	25	17	0	0,0
Fundaciones	40	28	12	42,9
Mutuas de accidentes de trabajo y enfermedades profesionales colaboradoras de la <i>Seguridad Social</i>	23	16	16	100,0
Organismos autónomos estatales	59	41	35	85,4
Organismos que no pertenecen a la Administración del Estado, diferenciados en los <i>Presupuestos generales del Estado (PGE)</i>	6	4	1	25,0
Otras entidades estatales de derecho público	57	39	28	71,8
Otros entes participados por el estado: consorcios del ámbito autonómico o local	63	43	1	2,3
Otros entes participados por el estado: consorcios del sector privado	3	2	0	0,0
Sociedades mercantiles estatales y asimiladas	148	102	53	52,0
Total	483	332	179	53,9

Elaborado a partir de los inventarios *Invente e Invespe* de la *Intervención General de la Administración del Estado (IGAE, 2016a; 2016b)* y de la *Plataforma de contratación del sector público (PCSP)*

Para hacer el universo de discurso completo de la administración estatal/central se han usado los dos, puesto que en *Invente* no se encontraban todos los tipos de entes estatales/centrales desglosados por tipo y en *Invespe* se encontraban también los entes con participación estatal minoritaria, que se han excluido puesto que la *Ley* no obliga a la transparencia en estos casos (sólo a partir del 50% de participación estatal).

Desde *Invente* se pueden extraer los datos en formato *Excel*, mientras que desde *Invespe*, aunque ofrece la opción, no descarga la información que contiene en ese formato, por lo que ha sido extraída directamente desde html que sí lo permite.

La información obtenida de estas dos herramientas arroja la tipología de entidades estatales que se presenta en la tabla 2.

Una vez recopilados los listados con todos los entes inventariados en ambos servicios se ha realizado una muestra estratificada con un alfa: 1,96 y error: 0,03 (tabla 3).

Una vez hecho el listado de los entes de cada tipología, se han numerado y escogido al azar las entidades (*Alazar.info*, 2016). Del análisis de la muestra estratificada se han obtenido los resultados que se muestran en las columnas 4 y 5 de la tabla 3.

El desglose de entidades según ministerios (excluyendo a los propios ministerios) y el grado de cumplimiento por ministerio es el recogido en la tabla 4<sup>5</sup>.

## 8. Discusión

De esta muestra seleccionada para conocer el grado de cumplimiento de las leyes de *Transparencia y buen gobierno y Reutilización de los datos del sector público de la Administración central* se observa que la publicación de los datos aparece centralizada en el *Portal de la transparencia*, tal y como dispone la norma jurídica, lo que facilita, además, la localización de la información de acuerdo con la *Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno (LTBG)*. Sin embargo, al no permitir el filtrado y la reutilización de los datos en realidad no cumple su cometido legal en materia de contratación.

Los resultados de las búsquedas de la *Plataforma de contratación del sector público (PCSP)* sí que se ofrecen en formato html extraíble y permiten filtrados por distintos conceptos. No ocurre lo mismo con el listado de entidades que están publicitando su contratación en él. Por consiguiente, para encontrar las licitaciones de una entidad es preciso conocer su nombre exacto, siendo en algunos casos difícil de localizar pues se usan acrónimos o nombres completos sin un criterio fijo.

La otra opción es pasar por el listado, en menú desplegable y no extraíble, de todas las entidades, una a una, hasta localizar la buscada, lo que implica saber de antemano el organigrama completo de la administración a la que pertenece. Lo anterior dificulta aplicar lo establecido en la *LTBG* respecto a cómo debe ofrecerse la información pública de manera

Tabla 4. Desglose de entidades por ministerios

		Muestra	MAGRAMA	MAEC	MD	MINECO	MECD	MEYSS	MFOM	MINHAP	MINETUR	MIR	MJU	Presi- dencia	MSSSI
Agencias estatales	Sí	4	0	—	—	1	1	—	1	0	—	—	—	—	1
	No	3	1	—	—	1	0	—	0	1	—	—	—	—	0
Consortios estatales	Sí	8	—	2	—	4	1	—	—	1	—	—	—	—	—
	No	4	—	1	—	2	1	—	—	0	—	—	—	—	—
Entidades gestoras SS	Sí	3	—	—	—	—	—	2	—	—	—	—	—	—	1
	No	0	—	—	—	—	—	0	—	—	—	—	—	—	0
Entidades públicas empresariales	Sí	9	—	—	—	2	—	—	5	—	2	—	—	—	—
	No	0	—	—	—	0	—	—	0	—	0	—	—	—	—
Fondos carentes de personalidad jurídica	Sí	0	0	0	—	0	—	—	0	0	0	—	—	—	0
	No	17	2	3	—	2	—	—	2	4	3	—	—	—	1
Fundaciones	Sí	12	1	—	0	3	3	1	0	2	0	1	0	—	1
	No	16	0	—	1	2	4	1	2	2	2	1	1	—	0
Mutuas de accidentes de trabajo SS	Sí	16	—	—	—	—	—	16	—	—	—	—	—	—	—
	No	0	—	—	—	—	—	0	—	—	—	—	—	—	—
Organismos autónomos	Sí	35	8	—	3	4	4	2	0	3	3	0	2	3	3
	No	6	1	—	0	0	0	1	1	1	0	1	0	0	1
Organismos diferenciados en los PGE	Sí	1	—	—	—	—	—	—	—	1	—	—	—	—	—
	No	3	—	—	—	—	—	—	—	3	—	—	—	—	—
Otras entidades de derecho público	Sí	28	—	—	2	2	3	—	14	4	1	1	1	0	—
	No	11	—	—	1	0	0	—	2	6	0	1	0	1	—
Otros entes: consorcios autoómicos o locales	Sí	1	0	0	—	0	1	—	0	0	0	—	—	—	—
	No	42	2	1	—	3	26	—	3	6	1	—	—	—	—
Otros entes participados: consorcios del sector privado	Sí	0	—	—	—	0	0	—	—	—	—	—	—	—	—
	No	2	—	—	—	1	1	—	—	—	—	—	—	—	—
Sociedades mercantiles	Sí	53	3	—	0	2	—	—	15	30	3	—	—	—	—
	No	49	1	—	2	13	—	—	19	10	4	—	—	—	—
Ministerios	Sí	9	1	1	1	1	—	1	1	1	1	—	1	—	—
	No	0	0	0	0	0	—	0	0	0	0	—	0	—	—
Totales	Sí	179	13	3	6	19	13	22	36	42	10	2	4	3	6
	No	153	7	5	4	24	32	2	29	33	10	3	1	1	2
%	Sí	53,9	65,0	37,5	60,0	44,2	28,9	91,7	54,4	56,0	50,0	40,0	80,0	75,0	75,0

Elaborado a partir de la *Plataforma de contratación del sector público (PCSP)*



Imagen 2. Solicitudes de derecho de acceso y reclamaciones ante el *Consejo de Transparencia y Buen Gobierno*

que permita al ciudadano localizarla y reutilizarla a través de formatos destinados a tal fin, para su posible tratamiento e interpretación con otros datos.

Por otra parte, de todas las entidades obligadas a publicitar su contratación pública por la *Ley de transparencia*, sólo el 53,9% lo están haciendo en la *Plataforma*. Si lo están haciendo en el *Portal* es imposible determinarlo de forma directa pues es necesario saber a qué ministerio, subdirección general, etc., se encuentra adscrita cada una. No existe, por tanto, una centralización accesible *de facto*.

El seguimiento según tipo de entes es muy dispar. Mientras hay tipos en los que el 100% de las entidades están publicitando sus contrataciones en la *Plataforma* (ministerios, entidades gestoras y servicios comunes de la *Seguridad Social*, entidades públicas empresariales estatales y mutuas de accidentes de trabajo y enfermedades profesionales colaboradoras de la *Seguridad Social*), otros como las agencias, fundaciones, organismos separados de los *Presupuestos generales del Estado (PGE)*, consorcios participados mayoritariamente por el Estado (autonómicos, locales y privados) o sociedades mercantiles, no llegan a la media.

Puede ser que en algunos casos, como ocurre a los organismos diferenciados en los *Presupuestos*, tengan sus propios portales, pero la *LTBG* los obliga a hacerlo, además, por los medios estatales destinados a ese fin. Caso similar es el de los consorcios de ámbito autonómico y local, que pueden estar haciéndolo por los portales autonómicos y locales habilitados; sin embargo, si la participación estatal es mayoritaria (más del 50%), deben hacerlo también por los mecanismos estatales.

El desglose por ministerios arroja que 8 de ellos superan la media de cumplimiento de entidades obligadas. Sin embar-

go, los de *Asuntos Exteriores y de Cooperación (MAEC)*, y *Educación, Cultura y Deportes (MECD)* quedan lejos de la media con un 37,5% y 28,9%, respectivamente.

Es imposible saber si los organismos públicos estatales están haciendo publicidad de todos los procesos de contratación a los que están obligados, o sólo de una parte de ellos. Esta información habría que contrastarla con las partidas económicas al finalizar cada ejercicio; es imposible saber si se está cumpliendo la *LTBG* en tiempo real en este punto. De hecho, se han detectado casos de entidades de las que se han encontrado licitaciones sólo anteriores a 2015, cuando no han sido eliminadas en la reestructuración de la administración pública.

Los resultados del análisis parecen avalar la denominación de "ley gatopardesca" que se ha dado en entornos jurídicos a la *LTBG* al entender que respondía "a la necesidad de cambiar todo, para que nada cambie" (**De-la-Nuez-Sánchez-Cascado, 2014**). Por una parte, el *Portal de la transparencia* no responde a los requerimientos de la norma jurídica en cuanto a la accesibilidad y reutilización de la información, y en la *PCSP* se sobrepasa escasamente el 50% de cumplimiento, y eso en el mejor de los casos y haciendo un acto de fe al considerar que se está publicitando toda la contratación a la que obliga la ley. Pero es que la norma jurídica no incluye garantías sancionadoras por los incumplimientos de la ley en relación a la transparencia (**Cotino-Hueso, 2014**).

## 9. Conclusiones

La recopilación de la información es muy difícil debido a su dispersión. Sólo para la elaboración de los listados de entidades obligadas a la transparencia en su contratación se han tenido que usar dos herramientas (*Invente* e *Invespe*)

a las que ni el *Portal de la transparencia* ni la *Plataforma de contratación del sector público (PCSP)* remiten desde sus páginas.

Aunque el *Portal* estuviera cumpliendo la *Ley 19/2013*, de 9 de diciembre, de *transparencia, acceso a la información pública y buen gobierno (LTBG)* sería imposible medir el grado de cumplimiento. Las graves deficiencias en sus opciones de búsqueda (no permite filtrar los documentos más que por ministerio y fecha, haciendo imposible el acceso a un sólo tipo de las entidades enumeradas por la norma jurídica, o a las entidades individualizadas), o la imposibilidad de extracción de datos para su reutilización, la hacen una herramienta inoperante. Para que el *Portal de la transparencia* cumpliera con los objetivos de la *LTBG* debería ofrecer, al menos, la posibilidad de búsqueda por el listado de entidades que se enumeran en la norma jurídica, o remitir a los enlaces donde se puede encontrar esa información, ya que así lo recoge la *Ley*, y permitir la extracción de la información para su reutilización.

La *Plataforma*, aunque es manifiestamente mejor que el *Portal*, presenta deficiencias como la imposibilidad de la descarga de los listados de las entidades que hacen su publicidad en ella, o fallos en el control de nombres de las entidades recogidas por el deficiente control de autoridades.

La *LTBG* no se está cumpliendo en sentido estricto en cuanto a contratación porque no es posible acceder directamente a la información de licitaciones por entidades desde el *Portal de la transparencia* creado para ese fin, y, además, no ofrece enlaces que suplan la dispersión de fuentes.

La dispersión en los mecanismos de publicidad junto con los fallos de presentación y control en los mismos, además de hacer muy difícil el acceso a la información, dificultan el control del cumplimiento de las entidades de forma conjunta e individualizada.

El seguimiento por parte de la Administración central es insuficiente a casi 3 años de la entrada en vigor de la *LTBG*, presentando grandes diferencias en el grado de cumplimiento, tanto entre los tipos de entidades como entre los ministerios, a lo que está contribuyendo seguramente que no existen sanciones por incumplimiento.

## Notas

1. Aunque algunas entidades de los listados y de las muestras escogidas al azar han sido suprimidas, tal como aparece en el *Informe CORA (Comisión para la Reforma de las Administraciones públicas)* de febrero de 2016, de muchas de ellas aparecen licitaciones en la *Plataforma de contratación* hasta 2015. Puesto que siguen apareciendo en el *Inventario de entes del sector público y público estatal*, se han mantenido para este trabajo.

<http://transparencia.gob.es/transparencia/dam/jcr:7b44c3a3-9e15-4b50-98d0-47b8744c04c7/CORA-informe-anual-progreso-2016-02.pdf>

2. Los “Fondos sin personalidad jurídica” no están obligados por la *LTBG* a publicitar sus licitaciones; no obstante, son mencionados en ella y aparecen en la relación de la *Plataforma*, por lo que han sido incluidos en el estudio.

3. Los “Organismos diferenciados en los *Presupuestos generales del Estado*” están obligados a hacer publicidad y no se contempla en la *Ley* que puedan hacerlo en sitios propios. Sus licitaciones deberían aparecer en los medios estatales habilitados a tal fin. Se han asimilado todos al *Ministerio de Hacienda y Administraciones Públicas*, que es de quien reciben sus fondos directamente.

4. Los consorcios autonómicos o locales incluidos en la muestra han sido extraídos del listado de consorcios participados mayoritariamente por el Estado. En algunos casos se ha comprobado que ha cambiado la participación mayoritaria estatal en favor de la autonómica o la local, pero al no haber sido cambiada su clasificación oficial, se han mantenido en este apartado.

5. En la tabla 4 se han usado estos acrónimos de los ministerios:

- *Ministerio de Agricultura, Alimentación y Medio Ambiente (MAGRAMA)*;
- *Ministerio de Asuntos Exteriores y de Cooperación (MAEC)*;
- *Ministerio de Defensa (MD)*;
- *Ministerio de Economía y Competitividad (MINECO)*;
- *Ministerio de Educación, Cultura y Deportes (MECD)*;
- *Ministerio de Empleo y Seguridad Social (MEYSS)*;
- *Ministerio de Fomento (MFOM)*;
- *Ministerio de Hacienda y Administraciones Públicas (MIN-HAP)*;
- *Ministerio de Industria, Energía y Turismo (MINETUR)*;
- *Ministerio de Interior (MIR)*;
- *Ministerio de Justicia (MJU)*;
- *Ministerio de Presidencia (Presidencia)*;
- *Ministerio de Sanidad, Servicios Sociales e Igualdad (MSSSI)*.

## Bibliografía

Alazar.info (2016). *Generador de números enteros sin repetición*. <http://www.alazar.info/generador-de-numeros-aleatorios-sin-repeticion>

Ares-González, Valle (2014). “Los ámbitos subjetivos y objetivos de la transparencia de la actividad pública”. *Revista jurídica de Castilla y León*, n. 33, pp. 1-27. <http://goo.gl/9URUwa>

Boix, Andrés (2015). “Transparencia, participación y procedimiento en la elaboración de disposiciones reglamentarias para un modelo de *open government*”. En: Cotino-Hueso, Lorenzo; Sahuquillo-Orozco, José-Luis; Corredoira-Alfonso, Loreta (eds.). *El paradigma del gobierno abierto. Retos y oportunidades de la participación, transparencia y colaboración*. Madrid: Universidad Complutense de Madrid, pp. 123-129. ISBN: 978 84 606 9678 0 <http://eprints.ucm.es/35859>

Chinchilla-Martín, Carmen (1988). *La radiotelevisión como servicio público esencial*. Madrid: Tecnos. ISBN: 84 309 1574 5

Cotino-Hueso, Lorenzo (2014). “La nueva Ley de transparencia y acceso a la información”. *Anuario de la Facultad de Derecho*, n. 7, pp. 241-256. Universidad de Alcalá. <http://dspace.uah.es/dspace/handle/10017/22075>

- Cotino-Hueso, Lorenzo** (2015). "Derechos humanos, internet y TICs". En: Rey-Martínez, Fernando (dir.). *Los derechos humanos en España: un balance crítico*. Valencia: Tirant Lo Blanch, pp. 449-513. ISBN: 978 84 370 9729 9
- De-la-Nuez-Sánchez-Cascado, Elisa** (2014). "El proyecto de Ley de transparencia, acceso a la información pública y buen gobierno ¿Una ley gatopardesca?". ¿Hay Derecho?, 24 de septiembre.  
<http://hayderecho.com/2012/09/24/el-proyecto-de-ley-de-transparencia-acceso-a-la-informacion-publica-y-buen-gobierno-una-ley-gatopardesca>
- España (2007a). "Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público". *BOE*, n. 276, 17 de noviembre.  
[https://www.boe.es/diario\\_boe/txt.php?id=BOE-A-2007-19814](https://www.boe.es/diario_boe/txt.php?id=BOE-A-2007-19814)
- España (2007b). "Ley 30/2007, de 30 de octubre, de Contratos del sector público". *BOE*, n. 261, 31 de octubre. Texto consolidado 16/11/2011, en vigor a partir de 16/12/2011.  
<https://www.boe.es/buscar/act.php?id=BOE-A-2007-18874>
- España (2013). "Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno". *BOE*, n. 295, 10 de diciembre. Texto consolidado.  
<https://www.boe.es/buscar/act.php?id=BOE-A-2013-12887>
- España (2015). "Ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público". *BOE*, n. 164, 10 de julio.  
[https://www.boe.es/diario\\_boe/txt.php?id=BOE-A-2015-7731](https://www.boe.es/diario_boe/txt.php?id=BOE-A-2015-7731)
- Fernández-Ramos, Severiano** (2012). "Comentario de urgencia al anteproyecto de Ley de transparencia, acceso a la información pública y buen gobierno". *Actualidad administrativa*, n. 12.
- Ferrer-Sapena, Antonia; Peset, Fernanda; Aleixandre-Benavent, Rafael** (2011). "Acceso a los datos públicos y su reutilización: *open data* y *open government*". *El profesional de la información*, v. 20, n. 3, pp. 260-269.  
<http://dx.doi.org/10.3145/epi.2011.may.03>
- Guichot-Reina, Emilio** (coord.) (2014). *Transparencia, acceso a la información pública y buen gobierno. Estudio de la Ley 19/2013, de 9 de diciembre*. Madrid: Tecnos. ISBN: 978 84 309 6166 5
- IGAE (2016a). *Inventario de entes del sector público (Invente)*. Secretaría de Estado de Presupuestos y Gastos-Ministerio de Hacienda y Administraciones Públicas-Gobierno de España.  
<http://www.igae.pap.minhap.gob.es/sitios/igae/es-ES/invente/Paginas/inicio.aspx>
- IGAE (2016b). *Inventario de entes del sector público estatal (Invespe)*. Secretaría de Estado de Presupuestos y Gastos, Ministerio de Hacienda y Administraciones Públicas, Gobierno de España.  
<http://www.igae.pap.minhap.gob.es/sitios/igae/es-ES/ClnInvespe/Paginas/invespe.aspx>
- Ministerio de Asuntos Exteriores y de Cooperación (2014). *Informe sobre el gobierno abierto en España*. Gobierno de España.  
<http://www.opengovpartnership.org/country/spain>
- Ministerio de Hacienda y Administraciones Públicas (2016). *Plataforma para la contratación del sector público. Gobierno de España*.  
<https://contrataciondelestado.es/wps/portal/plataforma>
- Moretón-Toquero, Arancha** (2014). "Los límites del derecho de acceso a la información pública". *Revista jurídica de Castilla y León*, n. 33, pp. 1-24.  
<http://goo.gl/SXyCal>
- Olmedo-Palacios, Manuel** (2014). "La Ley 10/2013, de 9 de diciembre, de transparencia, acceso a la información y buen gobierno". *Diario la ley*, n. 8237.
- Piñas-Mañas, José-Luis** (2015). "Transparencia y derecho de acceso a la información pública. Algunas reflexiones en torno al derecho de acceso en la Ley 19/2013, de transparencia, acceso a la información y buen gobierno". *Revista catalana de dret públic*, n. 49, (diciembre 2014), pp. 1-19,  
<http://dx.doi.org/10.2436/20.8030.01.29>
- Rollnert-Liern, Göran** (2014). "El derecho de acceso a la información pública como derecho fundamental: una valoración del debate doctrinal a propósito de la Ley de transparencia". *Teoría y realidad constitucional*, n. 34, pp. 349-368.  
<http://revistas.uned.es/index.php/TRC/article/view/14097>
- Rollnert-Liern, Göran** (2015). "Ley de transparencia y el derecho de acceso: la discusión doctrinal y parlamentaria en torno a su naturaleza de derecho fundamental". En: Cotino-Hueso, Lorenzo; Sahuquillo-Orozco, José-Luis; Corredoira-Alfonso, Loreto (eds.). *El paradigma del gobierno abierto. Retos y oportunidades de la participación, transparencia y colaboración*. Madrid: Universidad Complutense de Madrid, pp. 105-114. ISBN: 978 84 606 9678 0  
<http://eprints.ucm.es/35859>
- Sánchez-de-Diego, Manuel** (2013). "Transparencia y acceso a la información pública ¿son lo mismo? Por un derecho fundamental a acceder a la información pública". En: Corredoira-Alfonso, Loreto; Cotino-Hueso, Lorenzo (dirs.). *Libertad de expresión e información en internet. Amenazas y protección de los derechos personales*. Madrid: Centro de Estudios Constitucionales, pp. 285-324. ISBN: 978 8425915611
- Sánchez-de-Diego, Manuel** (2014). "El 'día después' de la Ley de transparencia". *Revista jurídica de Castilla y León*, n. 33, pp. 1-27.  
<http://goo.gl/T4mJXK>
- Sánchez-de-Diego, Manuel** (2015). "Derecho a la información e información de tribunales. Los secretos oficiales". En: Bel-Mallén, Ignacio; Corredoria-Alfonso, Loreto (eds.). *El ejercicio del derecho a la información y su jurisprudencia*. Madrid: Centro de estudios Políticos y Constitucionales, pp. 451-488. ISBN: 978 84 259 1622 9
- Tu derecho a saber* (2015). "El Defensor del Pueblo reconoce que la complejidad del Portal de transparencia puede tener un efecto disuasorio sobre el derecho de acceso a la información". *El blog de tu derecho a saber.es*, 4 de septiembre.  
<http://goo.gl/KEauB8>

# DATA MANAGEMENT IN AUDIOVISUAL BUSINESS: NETFLIX AS A CASE STUDY

## Gestión de datos en el negocio audiovisual: *Netflix* como caso de estudio

Eva-Patricia Fernández-Manzano, Elena Neira and Judith Clares-Gavilán

**Nota:** Este artículo puede leerse traducido al español en:  
[http://www.elprofesionalde lainformacion.com/contenidos/2016/jul/06\\_esp.pdf](http://www.elprofesionalde lainformacion.com/contenidos/2016/jul/06_esp.pdf)



**Eva-Patricia Fernández-Manzano** graduated in audiovisual communication and obtained her PhD degree (2005) from *Complutense University of Madrid*. She has a Master's in audiovisual business management from *Instituto de Empresa*, and a Master's in big data and business intelligence management from *Escuela de Organización Industrial (EOI)*, both in Madrid. She is a professor at the *Rey Juan Carlos University* on media management. Her research focuses on how media companies work with technology. She has published several papers and books about media and big data, the last one being *Big data, eje estratégico de la industria audiovisual* (UOC, 2016).  
<http://orcid.org/0000-0001-7655-872X>

*Universidad Rey Juan Carlos*  
Edificio Departamental, Despacho 207  
Camino del Molino, s/n. 28943 Fuenlabrada (Madrid), Spain  
[evapatriciafer@gmail.com](mailto:evapatriciafer@gmail.com)



**Elena Neira** has a bachelor of law and media studies (audiovisuals). She has worked as a market researcher in *Grupo Zeta*, and marketing manager in the movie distribution business (*Wide Pictures*). Currently, she is working as a consultant for several media companies. She specializes in online marketing, social media, VoD, and new audiovisual distribution models. She is a teacher and lecturer. Member of *Innovación Audiovisual*. Author of *El espectador social* (2013) and *La otra pantalla* (2015).  
<http://orcid.org/0000-0002-8565-2889>

*La otra pantalla*  
Rosselló, 421 - 4, 1. 08025 Barcelona, Spain  
[elena@laotrapantalla.com](mailto:elena@laotrapantalla.com)



**Judith Clares-Gavilán** is a lecturer and researcher at the *Faculty of Information Sciences and Communication* at the *Open University of Catalonia (UOC)*. She holds a degree in journalism and a PhD in communication from the *Ramon Llull University*. Her main research interests are: structure and policies of the audiovisual sector and especially the changes in the industry with the advent of audiovisual distribution through internet, video on demand (VoD), and online TV. Member of the *GAME* group, she is a researcher of the R+D project "Cultura lúdica, competencia digital y aprendizajes" funded by the *Ministry of Economy and Competitiveness* of Spain.  
<http://orcid.org/0000-0002-7462-9712>

*Universitat Oberta de Catalunya, Estudios de Ciencias de la Información y de la Comunicación*  
Rambla Poblenou, 156. 08018 Barcelona, Spain  
[jclares@uoc.edu](mailto:jclares@uoc.edu)

### Abstract

Big data has become an enormous asset for on-demand content distribution services, helping information supply and decision-making, regarding both the content of the database and subscribers to the database. In this article we describe and define big data and data management in a media company devoted to on-demand audiovisual content distribution: *Netflix*. This article suggests that big data is a prime strategy in media business and outlines the upcoming challenges that follow its global expansion.

Manuscript received on 31-03-2016

Approved on: 10-04-2016

## Keywords

Audience; Big data; Business intelligence; Business models; Data analysis; Data mining; *Netflix*; On demand audiovisual content distribution; Video on demand (VOD).

## Resumen

El negocio de la distribución de contenidos audiovisuales bajo demanda ha encontrado en los *big data* un aliado estratégico para la obtención de información y la toma de decisiones en torno al *core content* de su negocio: el contenido y sus suscriptores. A través de este artículo analizamos qué son los *big data* y cuál es su flujo de gestión de datos en una empresa audiovisual orientada a la distribución de contenido audiovisual bajo demanda. Para ello, seleccionamos como caso de estudio la compañía *Netflix*. A partir de aquí veremos por qué los *big data* son considerados eje estratégico del negocio y cuáles son los retos que se deben afrontar para la expansión internacional.

## Palabras clave

Audiencias; Análisis de datos; *Big data*; Inteligencia de negocio; Distribución audiovisual bajo demanda; Minería de datos; Modelos de negocio; *Netflix*; Vídeo bajo demanda (VOD).

**Fernández-Manzano, Eva-Patricia; Neira, Elena; Clares-Gavilán, Judith** (2016). "Data management in audiovisual business: *Netflix* as a case study". *El profesional de la información*, v. 25, n. 4, pp. 568-576.

<http://dx.doi.org/10.3145/epi.2016.jul.06>

## 1. Introduction

The application of new technologies in the audiovisual field has led to a series of systemic changes in business models. Certain sectors and companies have been quick to adapt to the new trends and workflows, but for the great majority the transition has not been completed. The adaptation roadmap is marked by two fundamental aspects:

- utility
- time slot.

This is because the new digital methods for accessing and using audiovisual products require new modes of action on the user's part. We notice that, in most cases, consumers do not integrate these technological changes in their day-to-day use until they realize their usefulness, which can range from improved image quality, satisfaction that the recommendation system understands the individual consumer's tastes, or the possibility of viewing content on demand rather than pre-recorded. As **Fernández-Manzano** (2016) points out, in the case of big data or mass data management in audiovisual fields, "the data is of no value in itself, its importance stems from the use that is made of it". This is, therefore, the utility. Furthermore, a temporary nuance appears which affects both data management and users, it entails the immediate alignment of people's technological demands with what the tools and algorithms can supply. That is, the amount of data new technologies can handle does not always offer the speed that users demand. This is the so-called time slot and is linked to such prolific fields as artificial intelligence and machine learning.

Both terms, utility and temporality, are intrinsically linked to the fact that the viewing of online audiovisual content leads to a change in attitude amongst the users, favoring the emergence of transmedia formats (**Scolari**, 2013) and the *prosumer* (**Jenkins**, 2008; **Napoli**, 2008), a user who takes part in the *storytelling*, adopting an active role (**Gubbins**, 2012).

Therefore, viewing audiovisual content on devices such as *tablets* or *smartphones* makes the viewer a data generator. This happens whenever a user is identified, content is shared, or something is "liked". Today's spectator generates information in the internet ecosystem.

The audiovisual, on demand, distribution industry is aware that its competitive advantage comes from obtaining information from its own users. Thus, *Netflix* has become the exemplar of internal data management and use of metadata, demonstrating a use of information that is flexible and adaptable to its environment and results in decision making based on big data. In under two decades, *Netflix* has established a business model that places the user at the core of its decisions. In practice, this translates into a major contribution, in terms of TV viewing innovation, with technology as its main ally. Owing, to a certain degree, to this philosophy, *Netflix* is currently a global brand with over 81 million subscribers, reproducing over 125 million hours of content per day, according to official company figures, corresponding to the first quarter of 2016.

The strategies applied to new media models, such as video on demand (VOD), are of significant academic interest. As **Izquierdo-Castillo** (2015) points out, the key factors that make up *Netflix's* strategy can be summed up in three inter-linked elements:

"a wide and varied catalogue that includes *premium* content in its basic package; a permanence-free economical rate that provides unlimited access to the listings; and a service geared towards image quality, simple browsing, similar content tips and accessibility through multiple devices."

In this paper we aim to contribute to this research path, outlining why big data are considered a strategic axis of the most recent media enterprises such as Video on Demand (VoD), which provides the user personalized content by

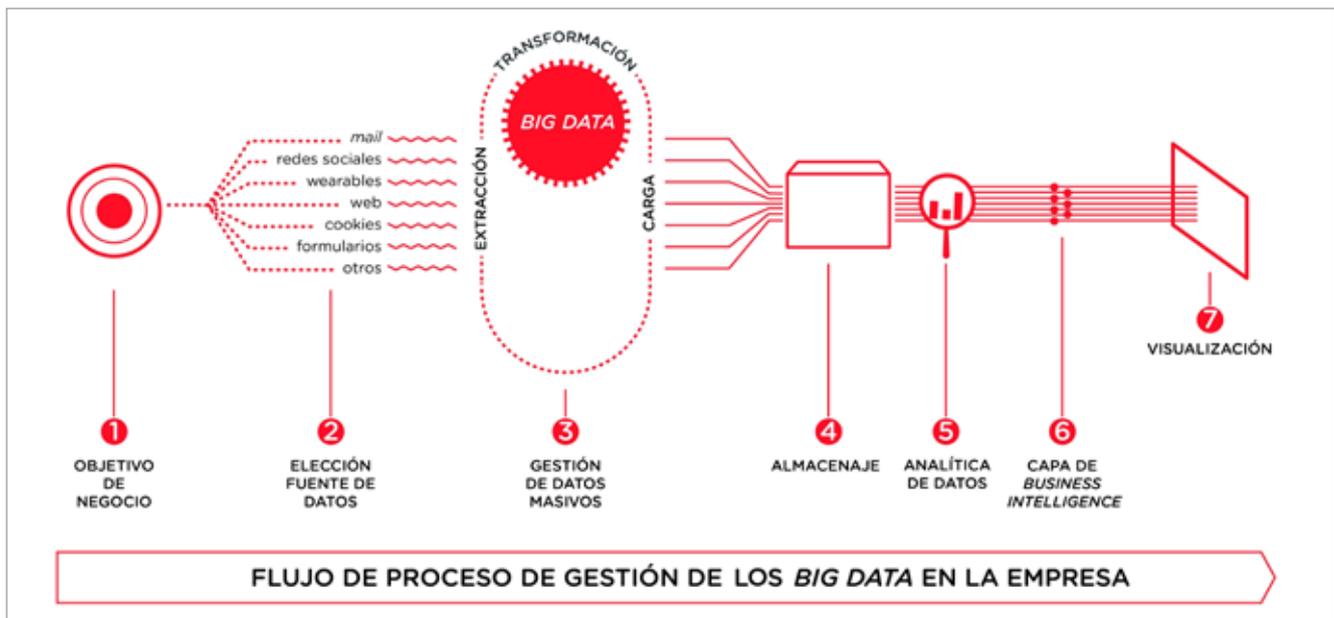


Figure 1. Process flow of big data management in the company (Fernández-Manzano, 2016)

means of an intelligent combination of suggested viewing algorithms, own content production, or new catalog acquisitions in line with client demand.

Big data are key to decision making for both the service’s technical quality as well as those pertinent to the catalog. It provides guidance on what content to offer, how to present it, and in what quantity. It also helps to make better decisions related to in-house productions (*Netflix originals*). This generation of personalized content has also led to promoting an alternative distribution strategy, adapted to recent demands by a new type of spectator who calls for total freedom when it comes to viewing audiovisual content, where, how, and whenever he wishes (Neira, 2015). Netflix took a gamble on releasing the entire season of original television shows at once, which ultimately led to what is now known as *binge watching* (Jenner, 2014). As pioneers in this field, Netflix has become a reference point for the changing model in the pay-TV industry.

“The audiovisual, on demand, distribution industry is aware that its competitive advantage comes from obtaining information from its own users”

## 2. Objectives and methodology

In this article we describe and define big data and data management. We will provide new arguments to the academic debate, and explain why mass data management is thought to be a strategic axis for the on demand audiovisual content distribution industry.

To do this we focus on *Netflix* as our case study. Its business targets will be identified, source selection, data gathering, transformation, and loading criteria will be

analyzed, as well as the monitoring and application of their own algorithms related to the application of the business decision-making layer, also known as *Business intelligence*. Finally, we will draw conclusions, focusing on the targets set.

To carry out this paper, we follow the methodology of the Yin (2009) study case, working with primary and secondary sources related to the firm’s own data management: bibliographic reviews, analysis of the data published by the firm itself (*Netflix Tech Blog* and the corporate area of *Netflix.com*), information provided by its staff in various discussion forums, and analysis of data provided by the specialized press.

## 3. Data management in the audiovisual business

New technologies are allowing companies to store and generate large volumes of information, and even demanding highly defined technological action, at a business level. In fact, users and businesses generate data in their interaction with new technology. The created data is so prolific that it has been termed big data. This is the term used to refer to the vast amount of data, which owing to its characteristics cannot always be processed by existing computing systems. This could be due to a volume issue, which will make it necessary to work with physical mass data storage (datacenters) or in the cloud. We can also take into account details such as time scales and velocity, enabling prediction making.

Based on the theoretical framework presented in Fernández-Manzano (2016), concerning big data management flow in business, we incorporate it into this paper as a guide to place and analyze the various technical aspects

Currently, the origin of the data included in big data terminology includes such dispirit sources as the web and the social media, biometrics, banking transactions, and machine to machine (M2M). As Serrano-Cobos (2014) suggests,

years ago firms only had access to their own data (in-house or through consultancies), but today, owing to the Internet, these businesses can contrast their information with market data, consumer feedback, or even scientific publications. Handling this information provides a crucial opportunity to come up with clever business questions, in addition to making the best decisions.

The main features of big data include volume, variety, veracity, and velocity (**Martínez-Martínez; Lara-Navarra, 2015**):

- volume, an enormous amount of available or accessible data;
- variety characteristically comes from diverse media sources, images, texts, music, figures, spreadsheets, different language types, and so on;
- data quality implies the need for veracity, given that its usefulness depends upon it;
- once again, the importance of velocity comes into play: the larger the data load, the greater the chance of speeding up decision making. This is a clear allusion to the business intelligence sector.

#### 4. The *Netflix* strategy in data analysis as a case study

The process flow of a company devoted to on demand audiovisual content distribution, aware of the usefulness of data, begins with a strategic business approach in which objectives are set; this will establish their parameters and performance indicators, also known as KPIs (key performance indicators).

Applying this logic to *Netflix* involves, as previously mentioned, understanding the user as a data generator. The 81 million subscribers worldwide, according to 2016 1T data, represent 125 million viewing hours per day. They acquire such magnitude within the company that it goes beyond the merely quantitative. In practice, each of these hours consumed by means of a device connected to the internet (streaming) is linked to individual information batches which, when controlled and analyzed, helps *Netflix* to understand its niche and identify its goals. All this allows them to design a flow of data management processes in accordance with their aspirations.

##### 4.1. Big data in *Netflix's* business goals

*Netflix*, self-proclaimed “world’s leading internet TV service”, bases its business model on subscription rather than on the advertising typical of open television, whether it be public or commercial (**Clares-Gavilán; Ripoll-Vaquero; Tognazzi-Drake, 2013**). It uses data collected from its users, not to negotiate with advertisers or media agen-

cies, in order to improve members’ retention, reduce cancellations, achieve long term fidelity, and obtain positive satisfaction ratings for their product (**Gómez-Urbe; Hunt, 2015**).

Without neglecting the main objective, it is paramount to list a complete set of sub-objectives, closely linked to this main one, in order to guarantee its compliance. Thus, data serves other functions, such as those listed below:

- Quality control of acquired rights (**Govind, 2014**).
- Optimal size of their catalog. As Jenny McCabe explains in *the Netflix quick guide for programming* (**McCabe, 2013**), *Netflix* does not aspire to have it all, but guarantees that their content is interesting and consumable. Hence, their catalog rotation and the non-renewal titles less popular amongst its users (not as profitable in terms of audience volume and licensing costs). Monthly pages containing information about upcoming titles are highly popular.
- The chance to offer subscribers personalized recommendations about content and help them choose without being overwhelmed by the size of the catalog.
- Produce content with their clients’ preferences in mind.

“*Netflix* has established a business model that places the user at the core of its decisions”

All these elements combined help to guarantee compliance with the first business objective mentioned: attract and avoid subscriber cancellation. Above all, the recommender system is at the core for keeping subscribers and long-term customer fidelity. Service cancellations, as **Gómez-Urbe; Hunt (2015)** point out, are more frequent during the free trial month because

“the recommendation service does a satisfactory job helping long-standing clients, but is not as effective with new subscribers, of whom little is known.”

Data analysis which allows them to know their clients is,



Figure 2. User watching *Netflix* on an internet connected device.  
<https://media.netflix.com/es/company-assets>

therefore, one of the strategic keys that guarantee the success of this VOD service. By not being advertising dependent, *Netflix's* profit is directly proportional to the number of paying subscribers. As McCabe points out in *the Netflix quick guide for programming* (McCabe, 2013), for the content license payment to be profitable, optimum viewing rates must be ensured.

Therefore, knowledge of their clients is the cornerstone of their *modus operandi*; allowing them to meet their aforementioned targets thanks to improved viewer experience and the constant search for efficiency in the type of content they offer.

Big data are key to decision making for both the service's technical quality as well as those pertinent to the catalog. It provides guidance on what content to offer, how to present it, and in what quantity. They also helps to make better decisions related to in-house productions (*Netflix originals*)

#### 4.2. Netflix's choice of data sources

In data source selection lies the implicit business goals. In the case of *Netflix*, it would be defined by the information provided by the users themselves. We are talking about a service requiring an internet-connected device, which allows for a significant number of data gathering, interaction, and convergence possibilities.

The average user spends 60-90 seconds deciding what to watch. After this time the chances of viewing any content are greatly reduced (Gómez-Uribe; Hunt, 2015). This window is key to understanding the monitoring that *Netflix* carries out on its consumers' habits. As can be understood from Mohammed Sabah's statements in *Hadoop Summit* (Sabah, 2012) and Yellin's speech on *How Netflix uses big data* (Yellin, 2015), aspects such as:

- viewed or discarded content type;
- playback characteristics (if it plays at normal speed, if forwarded or rewind, paused or abandoned);
- playback intensity (frequency and number of *streaming* hours per session);
- rating the user gives the content (keeping in mind the ratings given to other content);
- the sort of device used to gain access (connected TV, tablet, mobile phone, or decoder).

It also keeps track of the paths taken by the user: searches, geographic location, time and day, where on the page the content was found when selected and even the discarded recommendations).

*Netflix* has opted to focus its business model on the demand for technical viewing quality, as well as on the mobile and accessibility capacity of content (Izquierdo-Castillo, 2015). This is why it evaluates the technological state of each playback (Govind, 2014):

- how the content storage rate and transmission speed (*re-buffer* and *bitrate*) affect user behavior (hours spent on the service);
- how personalized experiences can be created (depending on the technology used or the quality of their bandwidth);
- how to improve content delivery (how content should be placed on the servers so they can be played more quickly, technical audio improvements, subtitles, and so forth).

This data is crossed with the metadata (information in the data itself) that *Netflix* assigns to all the content it adds to the platform (Madrigal, 2014): technical and artistic information, genre attributes or semantic markers (tags related to the narrative of the program or film), amongst others. From here, it is possible to clarify to what extent knowing the subscribers' preferences is linked to a service's high rate of use.

The data that the users actively provide (when rating content, filling out forms during the enrollment phase, logging on as specific users from a household with several members, or deciding to register with their social profiles) are testimonial in comparison with information generated (not supplied) by use, in an involuntarily manner (what is played and what is discarded, at what time it is viewed, on what connected device, how often, and so on). This monitoring mode, carried out within the user's own environment and not perceived by any connected device, produces valuable unconditioned answers that enable the gathering of knowledge about the users' motivation to watch some content or other (Brogan, 2015). This type of research, of an ethnographic nature, has become the engine of audiovisual structure reshaping as far as production and distribution is concerned (Izquierdo-Castillo, 2012), promoted in part by this new agent that is *Netflix*.

#### 4.3. The extract, transform and load process and Netflix's data storage systems

The data handling technical layer is known as ETL (extract, transform, and load). It involves the mass data (big data) management process in which criteria relevant to technical characteristics, such as quality (veracity characteristic), are applied. This allows the use of information that is useful, both because it does not contain errors (duplicate, erroneous, or invalid data) and because it corresponds with business goals. The loading process is framed within the data architecture and defines such aspects as data integration models or where it will be stored. There are usually two options:

- a huge data repository (warehouse or store), known generically as data warehouse, or
- storage in the cloud.

*Netflix* stores data on the *Amazon Web Service (AWS)* cloud infrastructure. This large scale management is what allows monthly streaming volumes of close to one billion hours per month (Tse, 2015).

#### 4.4. The analytical layer at Netflix

Once the data has been prepared, that is, once the big data required for the business' needs are available, the next step to be taken is known as the analytical layer. It offers informa-

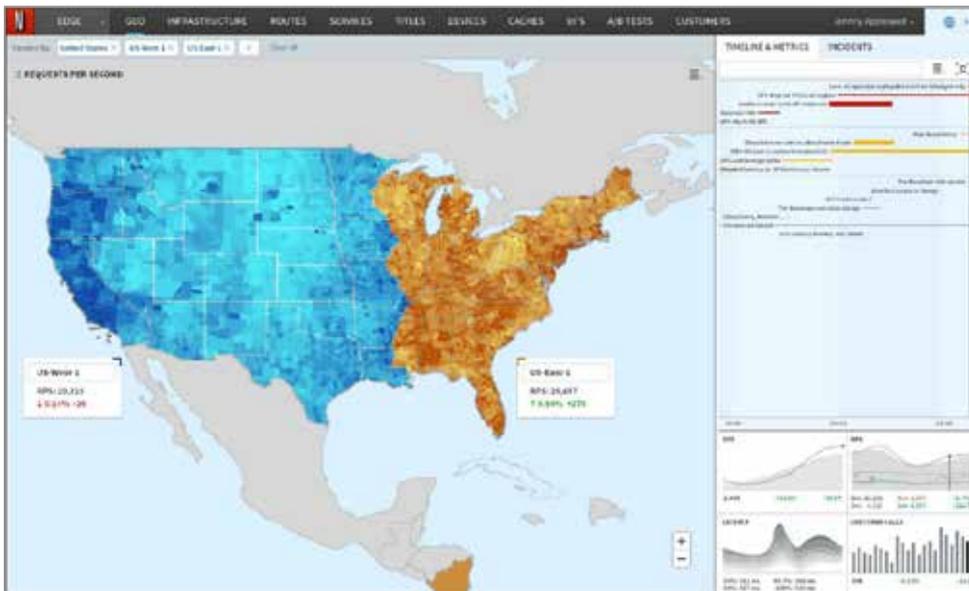


Figure 3. View of the front-end where graphics and dynamic data are presented and updated in real time

tion on metrics and previously established key performance indicators (KPIs). Data analysis shows information intended to become future strategies, but can also deal with economic aspects such as *return on investment* (ROI). Bearing in mind that *Netflix* has a ROI based on member subscription rather than on advertising (Clares-Gavilán; Ripoll-Vaquer; Tognazzi-Drake, 2013), it is the ideal setting to achieve strategies based on profiling information through segmentation. Achieving this return by means of monitored data involves starting a process to find out their value. To this end, an iterative process with large volumes of data is carried out in order to find a valid model. A common example in *online* consumption could be the identification of a useful algorithm that can detect either the desired information (in which episode of a series do the viewers become loyal followers?), or a conclusion (which country has the highest business model conversion rate?).

*Netflix's* catalog relies upon a complex web of algorithms that, nourished by monitoring its subscribers online use, aims to contribute towards formalizing satisfactory viewing decisions. As Gómez-Uribe and Hunt (2015) point out, depending on the characteristics of the device, *Netflix* can show up to forty selection rows (which tend to correspond to a specific algorithm). Two out of three hours of played *Netflix* content, explain the authors, are “discovered” at this stage. In practice, there are as many personalized models appearing under

recommended viewing as users subscribing to the platform.

Amongst the better-known algorithms, Gómez-Uribe and Hunt (2015) highlight the *Personal video ranker*, a sophisticated upgrade of traditional genre classification, hyperspecialized due to the task of assigning metadata to all the content uploaded onto the platform. *Netflix* counts on a team of people (known as *taggers*) whose job it is to watch programs and assign tags and categories. This content indexing has given rise to almost 80,000 microgenres within its audiovisual ecosystem (Madri-

gal, 2014). It allows them, furthermore, to find similitudes among films and breathe new categories (or subcategories) based on such patterns as production details, year, thematic fields, key roles of artistic or technical staff, adjectives ordered by popularity, and so on.

The authors also highlight *Top N video ranker*, which includes everything the viewer watches under the label *Top Picks*, a set of recommendations for the main categories generated by the previous algorithm, filtered on the basis of their popularity, but personalized for each client.

*Trending now* is the algorithm responsible for generating category trends, the logical result of the digital and social environment in which the new spectator is placed. Under this umbrella fall content choices relevant to consumption peaks with a limited life cycle, such as Christmas, Saint Valentine's

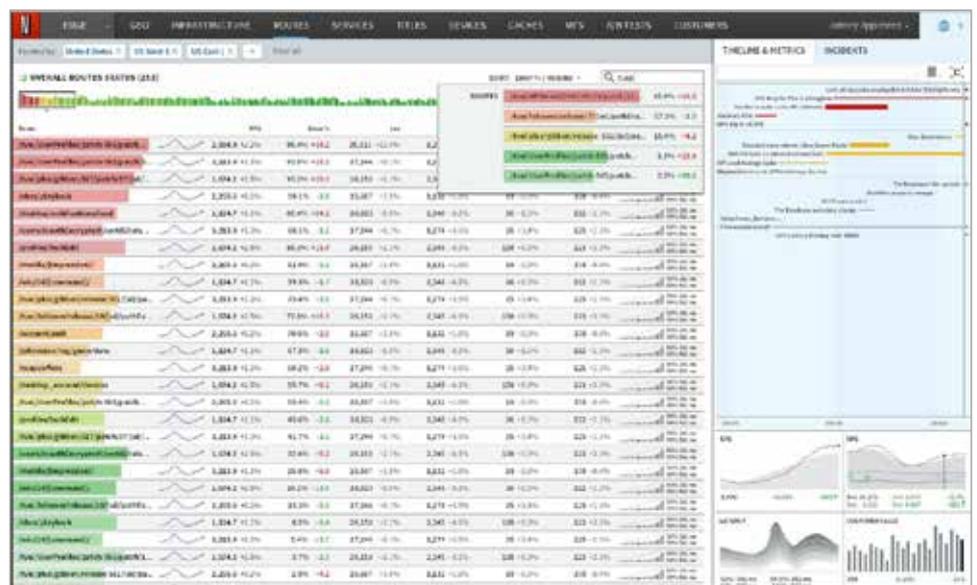


Figure 4. On the front-end can be seen the navigation routes followed by users, request rates, error rates and other key indicators for each route, which are updated almost in real time.

day, or the holiday periods.

As Yellin explains in *How Netflix uses big data* (2015), monitoring is performed on any playback as well as its abandonment. Such choice makes sense when the *Continue watching* kicks in. It creates a category that holds those titles that the user will most likely want to continue watching, whether it is serialized (episodes) or of specific content that the user has opted to watch in separate stages or has had to stop. It is a customized category that takes into account all the data that allows it to know what content to place first (last viewing time, where the reproduction was stopped, viewing frequency, device through which it was viewed, means of connection used, and so forth).

Another of the algorithms highlighted by Gómez-Urbe and Hunt (2015) is *Video-video similarity*, creates the category *Because you watched*. In this category we can find content related to a title that the user has already watched.

Even though there is also a commercial component that gives viewing priority to certain products (like “Recently added” or “Premieres”), *Netflix’s* access to the content page is a dynamic ecosystem that attempts to adapt to each subscriber’s idiosyncrasies. Therefore, faced with an audiovisual culture that is traditionally geared towards offering non-flexible categories, in *Netflix* we observe a highly customized experience due to this feeling of “novel product”, the result of browsing through the platform. In practice, the algorithmic content selection represents 80% of video hours played by its users (Gómez-Urbe; Hunt, 2015).

#### 4.5. Business intelligence

The information generated by data from user browsing experience is linked to the next business layer, known as *Business intelligence*. It has the mission of making better decisions. In this sense, having a content range modelled in accordance with the users’ proven tastes presupposes previous business decision making.

This business logic stems from previous processes, such as the application of algorithms. For users of an online platform it is crucial to be able to establish viewing and consumption patterns. To this end, *Machine learning* (a branch of artificial intelligence aiming to get computers to learn) performs a useful task, thanks to which data models are assessed in order to confirm their validity and, thus, apply business logics. This is *Business intelligence*.

The series *House of cards* is a recurring example when discussing big data applied to content design. As Neira (2015) states,



Figure 5. *Netflix* interface  
<http://www.netflix.com>

“in 2012 *Netflix’s* management board decided to create a new business unit within the company, whose purpose was to make in-house productions for its platform (*Netflix Originals*). With several proposals on the table, they concluded that overlapping three key elements (director David Fincher, The popularity of the original British series, and actor Kevin Spacey) created “proven success circle”, guaranteeing the project’s success (Carr, 2013). That was the conclusion after many months of data mining (...). *House of Cards* was a safe bet due to a confluence of data that already stood out due to individual merit, but formed a much more solid base when combined in one single product. The result was a data based business decision: investment came to 3.8 million dollars per episode”.

The design of this type of “tailor-made” product is a clear example of business intelligence. It is made possible because big data profiles a potential audience with a set of defined dimensions whose readiness to view the content is the path towards achieving their goals; maximizing client fidelity.

#### 5. Data analysis challenges for *Netflix’s* international expansion

For a global-minded company that, as from April 2016, has had an active service in over 190 countries, customizing is the only way to supply audiovisual content to all those territories in which it operates, without neglecting cultural diversity and catering to local audiovisual preferences.

As Raimon and Basilisco (2016) explain, until 2001, recommendations were mainly based on the place of residence. Every time *Netflix* went to a new country the service, as well as local content tagging, was reset manually. This recommendation approach intended to avoid standardization from damaging the viewing experience, and, thus, distancing itself from audiovisual cultures. Global expansion, strengthened in 2016, has imposed an operative change. Rather than sca-

ling the algorithmic approach to each market, the system has evolved by placing countries into regions with similar catalogs and letting the algorithms operate individually in each region, but processing their data globally. To prevent smaller countries from spoiling the algorithm, in larger countries, *engagement* data (identified as playbacks or intensity of use) are corrected based on each respective catalog size. We should not forget that the choice offered by *Netflix* in its various operating territories is different and varies in content and volume. The management of audiovisual rights, despite the lack of territorial boundaries on the internet, is still territorial and tied to the different existing distribution windows: home-video, VOD, and television (Clares-Gavilán, 2014).

“*Netflix* uses data collected from its users, not to negotiate with advertisers or media agencies, in order to improve members’ retention, reduce cancellations, achieve long term fidelity, and obtain positive satisfaction ratings for their product”

On the other hand, its big data management has been growing as more subscribers join the service (Izrailevsky, 2016) and it has been enriched with new layers due to global expansion. Programs can now use new territories as testing ground and it helps them to refine their clients’ audiovisual content preference maps and identify potential audiences for their products.

The confluence of efficiency in program design, the result of profound client awareness and new market openings, has also run parallel to the rise in *Netflix Originals*. This would be the case of *Narcos* (2015), a drama series based on the figure of Pablo Escobar, with a significant proportion of dialogue in Spanish. This production came at a time of strong consolidation and increasing Hispanic audiences, both local, from the United States, as well as international, and parallel to the service’s landing in a number of Latin American countries (Betancourt, 2015).

Global expansion has allowed them to resize their audience and, thanks to big data, identify the same viewer segments in the different territories. Their viewing niches are now global. General viewing content, such as *House of Cards* (2013), *Jessica Jones* (2015), or *Fuller House* (2016), now co-exist with products for the afore-mentioned global niches. *Sitcoms for millennials*, such as *Love* (2016) or *Masters of none* (2015) are a good example of this trend.

## 6. Conclusions

Having developed the paper, we can draw the following conclusions:

In the analyzed case we see how extraction and data management, and subsequent analysis and decision-making, has become a competitive business advantage. A business gearing its supply in relation to the demands and preferences of the user.

Using data mining techniques enables *Netflix* to segment its clients and identify user clusters in order to observe and offer personalized content. In this sense, these techniques have helped to set up its own production brand (*Netflix Originals*) adapted, largely, to subscriber interests and demands.

Through the use of data mining and content recommendation techniques we have been able to see how they manage to also know what their users want or need, and, as a result, lower the number of service cancellations.

We can also conclude that the service’s internationalization has been, curiously, the decisive step towards overcoming cultural fragmentation by countries. Owing to their business intelligence flow, dynamic, circular, and based on *machine learning* models (Govind, 2015), they have been able to evolve in the identification of more global niches and in tailor-made production content for an objective audience that has grown with the addition of new markets. *Netflix* now recommends, acquires, and produces worldwide.

Looking ahead to future research paths, we believe that of all the mentioned aspects it would be pertinent to widen analysis to other VOD areas in order to check if this reality is replicable to other case studies. In this way, we could analyze, evaluate, and identify trends that will help to continue identifying why big data are considered to be a strategic axis in the on demand audiovisual content distribution industry.

“*Netflix* knowledge of their clients is the cornerstone of their *modus operandi*”

Likewise, we consider that research on big data applied to the audiovisual content on demand distribution and consumption business calls for new research paths that can understand the relationship of new technologies with users and content generators.

## 7. References

- Betancourt, Manuel (2015). “*Netflix* unequal marketing of ‘Narcos’ and ‘Club de cuervos’ ignores bilingual latinos”. *Re-mezcla*, 27 de octubre. <http://goo.gl/qcXmTf>
- Brogan, Jacob (2015). “The case of the ornamental anthropologist. How *Netflix* puts a human face on big data”. *Slate*, 31 May. <http://goo.gl/9Tu7Ja>
- Carr, David (2013). “Giving viewers what they want”. *The New York times*, 24 Febr. <http://goo.gl/lwuks8>
- Clares-Gavilán, Judith (2014). *Estructura y políticas públicas ante los nuevos retos de la distribución y consumo digital de contenido audiovisual. Los proyectos de video bajo demanda de cine Filmin y Universciné como estudio de caso*. Universitat Ramon Llull. Tesis doctoral. <http://www.tesisenred.net/handle/10803/247706>
- Clares-Gavilán, Judith; Ripoll-Vaquer, Jaume; Tognazzi-Drake, Alberto (2013). *Distribución audiovisual en internet*.

*VoD y nuevos modelos de negocio*. Barcelona: Editorial UOC. ISBN: 978 84 9029 911 1

**Fernández-Manzano, Eva-Patricia** (2016). *Big data. Eje estratégico en la industria audiovisual*. Barcelona: Editorial UOC. ISBN: 978 84 9116 380 0

**Gómez-Uribe, Carlos A.; Hunt, Neil** (2016). "The Netflix recommender system: Algorithms, business value, and innovation". *AMC transactions on management information systems (TMIS)*, v. 6, n. 4, article n. 13. <http://dx.doi.org/10.1145/2843948>

**Govind, Nirmal** (2014). "Optimizing the Netflix streaming experience with data science". *Netflix tech blog*, 11 June. <http://techblog.netflix.com/2014/06/optimizing-netflix-streaming-experience.html>

**Gubbins, Michael** (2012). "Digital revolution. Active audiences and fragmented consumption". En: lordanova, Dina; Cunningham, Stuart (eds.). *Digital disruption: Cinema moves on-line*. St. Andrews: St. Andrews Film Studies. ISBN: 978 09563730 7 6

**Izquierdo-Castillo, Jessica** (2012). "Distribución online de contenidos audiovisuales: análisis de 3 modelos de negocio". *El profesional de la información*, v. 21, n. 4, pp. 385-390. <http://dx.doi.org/10.3145/epi.2012.jul.09>

**Izquierdo-Castillo, Jessica** (2015). "El nuevo negocio mediático liderado por Netflix: estudio del modelo y proyección en el mercado español". *El profesional de la información*, v. 24, n. 6, pp. 819-826. <http://dx.doi.org/10.3145/epi.2015.nov.14>

**Izrailevsky, Yury** (2016). "Completing the Netflix cloud migration". *Netflix media center*, 11 February. <https://media.netflix.com/en/company-blog/completing-the-netflix-cloud-migration>

**Jenkins, Henry** (2008). *Convergence culture: la cultura de la convergencia de los medios de comunicación*. Barcelona: Ediciones Paidós. ISBN: 978 84 493 2153 5 <http://dx.doi.org/10.1177/1461444814541523>

**Jenner, Mareike** (2016). "Is this TVIV? On Netflix, TVIII and binge-watching". *New media & society*, v. 18, n. 2, pp. 257-273. <http://dx.doi.org/10.1177/1461444814541523>

**Martínez-Martínez, Silvia; Lara-Navarra, Pablo** (2015). "El big data transforma la interpretación de los medios

sociales". *El profesional de la información*, v. 23, n. 6, pp. 575-581.

<http://dx.doi.org/10.3145/epi.2014.nov.03>

**Madrigal, Alexis** (2014). "How Netflix reverse engineered Hollywood". *The Atlantic*, 2 Jan.

<http://www.theatlantic.com/technology/archive/2014/01/how-netflix-reverse-engineered-hollywood/282679>

**McCabe, Jenny** (2013). *Netflix quick guide: How does Netflix decide what's on Netflix*. Jenny McCabe, Director of global media relations, Netflix.

<https://www.youtube.com/watch?v=VvpoUh9gx58>

**Napoli, Philip M.** (2008). "Toward the model of audience evolution: New technologies and the transformation of the media audiences". *McGannon Center working paper series*. Paper 15. [http://fordham.bepress.com/mcgannon\\_working\\_papers/15](http://fordham.bepress.com/mcgannon_working_papers/15)

**Neira, Elena** (2015). *La otra pantalla. Redes sociales, móviles y la nueva televisión*. Barcelona: Editorial UOC. ISBN: 978 8 4911 6116 5

**Raimond, Yves; Basilico, Justin** (2016) "Recommending for the world". *Netflix tech blog*, 17 Febr.

<http://techblog.netflix.com/2016/02/recommending-for-world.html>

**Sabah, Mohammad** (2012). *Hadoop Summit, 2012*. Mohammad Sabah, Principal data scientist, Netflix. Silicon Angle.

[https://www.youtube.com/watch?v=9h6wQXuke\\_E](https://www.youtube.com/watch?v=9h6wQXuke_E)

**Scolari, Carlos A.** (2013). *Narrativas transmedia: cuando todos los medios cuentan*. Barcelona: Deusto Ediciones. ISBN: 978 84 234133 6 2

**Serrano-Cobos, Jorge** (2014). "Big data y analítica web. Estudiar las corrientes y pescar en un océano de datos". *El profesional de la información*, v. 23, n. 6, pp. 561-565.

<http://dx.doi.org/10.3145/epi.2014.nov.01>

**Tse, Eva** (2015). "Netflix case study". *Amazon web service*. <https://aws.amazon.com/es/solutions/case-studies/netflix>

**Yellin, Todd** (2015). *How Netflix uses big data*. Todd Yellin, VP of innovation, Netflix.

<https://www.youtube.com/watch?v=Pu4myXu0ji0>

**Yin, Robert K.** (2009). *Case study research. Design and methods*. Thousand Oaks: Sage. ISBN: 978 1412960991

El profesional de la  
**información**

Bienvenido a **EPI** Indexada por ISI y Scopus  
ISSN 1386-6710 / ISSN-e 1699-2407  
Revista internacional, científica y profesional

<http://www.elprofesionaldelainformacion.com>

Revista internacional de  
**Información y Comunicación**  
indexada por ISI Social Sciences Citation Index (Q3),  
Scopus (Q2) y otras bases de datos

Factor de impacto JCR:  
**IF 2015 = 0,710**

Scopus/SCImago Journal Rank:  
**SJR 2015 = 0,422**



Presentación del Director

<http://www.elprofesionaldelainformacion.com>

# Colección de libros de bolsillo

## *El profesional de la información (Editorial UOC)*



**NOVEDAD**



El profesional de la  
**información**

**EDITORIAL UOC**

Más información:

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

# CIVIC PARTICIPATION AND INTERACTIVE DOCUMENTARIES: A CONTRIBUTION TO THE OPEN GOVERNMENT MODEL

Participación ciudadana y documentales interactivos: una contribución al modelo del gobierno abierto

Laura Cortés-Selva and Marta Pérez-Escolar



**Laura Cortés-Selva** is professor of media studies at the *Catholic University of Murcia*, Spain. She holds bachelors' degrees in media studies from the *University of Navarra* and cinematography from the *University of Barcelona*, and has a *cum laude* international PhD (2012). She has published several papers and book chapters about her main academic interests: visual style, cinematography, documentaries, and methodology. She has also been part of several research projects as well as an invited lecturer in various international universities. She has developed several professional projects as director, director of photography, and producer in short movies, documentaries, music videos, and commercials.

<http://orcid.org/0000-0003-0306-9401>

[lcortes@ucam.edu](mailto:lcortes@ucam.edu)



**Marta Pérez-Escolar** is a PhD student of communication science at the *Catholic University of Murcia* (Spain). She is a member of the *Comunicación, política e imagen* (communication, politics, and image) research group. Her main lines of research are cyber-activism, political participation, civic participation, cyber-democracy, collective intelligence, social movements, social protests, and political communication.

<http://orcid.org/0000-0003-2575-7993>

[mperez726@ucam.edu](mailto:mperez726@ucam.edu)

*Universidad Católica de Murcia, Comunicación Audiovisual*  
Campus de los Jerónimos, s/n. 30107 Guadalupe (Murcia), España

## Abstract

This research introduces the marriage between interactive documentaries (i-docs) and politics as an opportunity to motivate new civic participation and expression in the political sphere. The main objective of this study is to identify the different degrees of civic participation that i-docs permit and to setup a classification according to these levels of engagement. In order to achieve the main objective, this research has analyzed a group of documentaries using a mixture of Gaudenzi, Gifreu-Castells, and Nash methodology. The results show that it is possible to find a parallel between the closed mode and the Orwellian model; the semi-closed mode and the *Tokenism* model; the semi-opened mode and the playful model; and between the opened mode and the Athenian model. However, only the opened mode, used in the *Global lives* documentary, offers a real experience that symbolizes the purest representation of an open government system.

## Keywords

Participation; Documentaries; I-docs; Webdocs; Interactive documentaries; Open government; Civic engagement; Politics.

## Resumen

Esta investigación presenta el maridaje entre i-docs y la política como una oportunidad para motivar la participación ciudadana en la esfera política. El principal objetivo de este estudio es identificar los grados de participación ciudadana que permiten los i-docs y establecer una clasificación según esos niveles de implicación cívica. Para ello, esta investigación ha analizado un conjunto de documentales utilizando la metodología propuesta por Gaudenzi, Gifreu-Castells y Nash. Los resultados muestran que existe un paralelismo entre el modelo cerrado y el paradigma orwelliano; el modelo semi-cerrado y el paradigma de la cortina de humo o *Tokenism*; el modelo semi-abierto y el paradigma solaz y entre el modelo abierto y el paradigma ateniense. Sin embargo, sólo el modelo abierto, como ocurre en el documental *Global lives*, ofrece una experiencia real que simboliza la más pura representación de la filosofía del gobierno abierto u *open government*.

Manuscript received on 15-03-2016

Approved on: 10-04-2016

## Palabras clave

Participación; Documentales; *I-docs*; *Webdocs*; Documentales interactivos; Gobierno abierto, Implicación ciudadana; Política.

**Cortés-Selva, Laura**; Pérez-Escolar, Marta (2016). "Civic participation and interactive documentaries: a contribution to the open government model". *El profesional de la información*, v. 25, n. 4, pp. 578-587.

<http://dx.doi.org/10.3145/epi.2016.jul.07>

## Introduction

Old communication processes (Lasswell, 1948) are obsolete in the current public landscape. The internet and emergence of web 2.0 has changed the traditional distribution and consumption flows (Area-Moreira; Ribero-Pessoa, 2012). The digital revolution has motivated the rise of "participative communication" (Servaes, 1996; Servaes; Malikhao, 2005), encouraging citizens to monitor political elites and to play an active role in the decision-making process.

Following this idea, this research defends the position that civic participation becomes the cross axis of a new politics and the main ground of any cooperative model of open government (Sampedro; Sánchez, 2001; Dader, 2002). The web 2.0 has transformed the nature of the dialogue between the representatives and those who are represented and has triggered the redefinition of traditional democratic tenets. The main cause of the current crisis in representative democracy is the citizens' lack of trust in the political actors and public institutions. For this reason, contemporary social movements stand up for new political reforms that allow citizens to engage and participate in the public sphere. In this context, the open government model could be understood as the best organizational archetype for citizens because it is a hybrid political form that provides participation tools for civic engagement and uses new technologies to influence and control decisions and actions taken by their representatives (Campos-Domínguez, 2011).

“The marriage between interactive documentaries (i-docs) and politics provides a new opportunity for creating and reformulating expressions of civic participation”

This study considers that political i-docs could be understood as a perfect medium to interpret the open government paradigm. Therefore, the objective of this research is to analyze the levels of civic participation that i-docs allow in order to know the boundaries, challenges, and opportunities of the open government archetype. Although this is still an emerging political organization model, the results of this study could offer a prospective landscape about the future of civic participation and new forms of social engagement.

I-docs, as a new documentary genre (Gifreu-Castells, 2013), can provide renewed and exciting possibilities for civic participation in the political sphere: speakers –political elites and institutions– are not the only ones with the power to narrate stories –political storytelling–; but now receivers –

citizens– also find opportunities to divulge their stories and to share their narration with the rest of the receivers and speakers of the dialogical community. In this way, the story converts into a product that can be modified; this means that narration is not rigid, fixed or linear, but it is an organic, flexible, and malleable account. In this way the *i-doc*, or interactive documentary phenomenon (Gaudenzi, 2013), encourages the convergence of new social action forms like activism, in which "ordinary people" (Hibbing; Theis-Morse, 2002) achieve a new role as participative actors involved in configuring the political narration.

According to Dahlgren (2013), contemporary citizens have the abilities and qualities to produce their own content, to share information, to interact with the rest of the members of the community, and to participate in political debates and social movements organized in the digital environment. Thus Dahlgren (2013) proposed the term "civic intellectuals" in reference to a heterogeneous community of citizens that share the same social circumstances:

“the dilemmas of democracy, the character of the media landscape, and not least, the contemporary crises of capitalism” (Dahlgren, 2013, p. 403).

Therefore, the concept of "civic intellectual" implies a new form of collective identity that will

“lead democracy forward to a new golden age, but they do signal a potentially positive step in the chronicles of citizen participation and the evolution of the public sphere” (Dahlgren, 2013, p. 403).

In conclusion, this study supports the position that the marriage between interactive documentaries and politics means a new opportunity for creating and reformulating new civic participation expressions. These new social forms will facilitate the consumption, interaction, and transformation of the political story from speakers to receivers, as well as offer new ways to rethink about more open, horizontal, and more participative democratic models.

Concerning the principal objective aforementioned, this research not only aims to present the different levels of civic participation that i-docs offer, but also to determine if i-docs –as a kind of social activism– guarantee and enable the possibility of an open government model. Our point of view is focused on the idea that civic participation in the content and the structure of i-docs is what makes the difference and delimits the levels. For those reasons, applying the methodology suggested by Gaudenzi, Gifreu-Castells, and Nash, this study has analyzed a set of documentaries –most of them interactive– to match our objectives and expectations.

## 2. Theoretical approach to i-docs. From lineal documentary to interactive documentary

The documentary field has traditionally been characterized by a lack of consensus, especially in terms of its definition, and although there are several scholars<sup>1</sup> dealing with this subject, an agreement has not been reached.

It is necessary to understand and consider that, within film history, documentary as a term has included products such as so-called *actualités*, travelogs, news, educational films, and TV programs of various styles and content (León, 1999, p. 59).

Being aware of the polysemy of the aforementioned term and its openness to different meanings, this research considers the convenience of underlining the definitions of pioneers such as Robert Flaherty. Mainly quoted as the father of the documentary for works such as *Nanook of the North*, Flaherty understands that

“the purpose of the documentary is to represent life in a way in which it is lived. This by no means implies what some people might think; namely, that the task of the documentary director is to film, without making any selection [...]. The task of selection is performed on the documentary material, with the aim of telling the truth in the most appropriate way” (Romaguera-Ramió; Alsina-Thevenet, 1989, p. 152).

Nevertheless, John Grierson, leading historical figure of the British documentary movement, is considered the first to use the term documentary in 1926, which seems to be an adaptation of the French word *documentaire*, used during the 20's as a reference to travelogs (León, 1999, p. 59). Grierson interprets the term as a creative treatment of reality (Grierson, 1966, pp. 36-37), as an intention to include all the different works included in the documentary genre, that offer diverse ways to observe and organize reality.

Bill Nichols (1991) offers multiple perspectives to the term documentary, and his definition is the most accepted in the academic field. Nichols, being sensitive to historical changes, includes in his definition, not only the audience, but a body of texts and a group of filmmakers. From this, Nichols establishes three perspectives to understand documentaries:

- from the filmmaker point-of-view;
- from the text or film itself, and
- from the audience.

In this paper we will consider documentary films as a compendium of Grierson's creative approach to reality, together with Flaherty's perspective, and Nichols' multiple approach.

From the aforementioned, and with the consideration of the evolution of technology related to the web 2.0, i-docs (also known as webdocs, interactive docs, webdocumentaries, and interactive web documentaries) emerge.

Beginning approximately one decade ago, the creation and reflection upon this interactive documentary field has exponentially grown due to projects such as *Gaza/Sderot* (2008); *Highrise* (2009); *Prison Valley* (2010), and *Fort McMoney* (2013).

Among the scholars<sup>2</sup> that dare to try to attempt the complex task of defining and researching this emerging genre, Gifreu-Castells (2011, p. 358) stands out with the definition of interactive documentary as an online/offline interactive application which tries to represent different navigation and interaction modes, depending upon the participation level.

Nash (2011, p. 2) uses the term webdocumentaries in allusion to interactive documentaries and defines them as

“a body of documentary work distributed by the Internet that is both multimedia and interactive”.

Finally, Gaudenzi (2013, pp. 31-32) refers to this new documentary category as i-docs and defines them as

“any project that starts with an intention to document the ‘real’ (here, ‘reality’ is understood as any mediated material where mediation might happen through our senses, mind, or media) that we make sense of or make sense through (to establish a meaningful relationship with what surrounds us), and that does so by using digital interactive technology, will be considered an interactive documentary”.

Gaudenzi pretends to include all the factual narratives concerning present and future devices such as tablets, computers, mobile phones, etc. Thus, in this paper we opt to use Gaudenzi's terminology because it is the most suitable definition concerning the subject of study of this research.

“ The potential of i-docs resides in changing the role of citizens, from passive individuals to active political actors, via the interaction between participants of documentary communities ”

## 3. The next step: creating new spaces for civic participation

Participation is an important factor for the development of democratic societies; however, it also represents an ambiguous concept that can be subject to vagueness within the field of communication and media studies. For that reason, this research aims to consider Carpentier's approach to clarify and to delimit the meaning of participation

“on the basis of a comparison with two other concepts, access and interaction, elucidating the differences between these three concepts” (Carpentier, 2015, p. 9).

According to the *AIP model* (Carpentier, 2012), *Access*, *Interaction*, and *Participation* can be articulated in four areas: technology, content, people, and organizations. However, this paper has placed the focus on media technologies. Thus the concept of *access* could be understood as an opportunity for users to make their voices heard and to provide feedback. Access “implies achieving presence (to technologies or to media content)” (Carpentier, 2015, p. 10) and expresses the metaphor of “coming closer” (Berrigan, 1979; Carpentier, 2015), playing an important role in motivating people to produce and distribute content.

On the other hand, interaction

“refers to the establishment of socio-communicative relationships within the media sphere” (Carpentier, 2012, p. 174).

So the idea of “social interaction” denotes a reciprocity process in which citizens produce, select and interpret content individually or collectively (Jenkins; Carpentier, 2013). Interactivity is a characteristic of media technologies that

“incorporates the possibility of user–content and user–user interaction through the interaction between user and technology” (Carpentier, 2015, p. 17).

Finally, the term “participation” represents those practices that imply a decision-making process. Therefore, participation is a practice that includes different forms of interaction and implies a high level of civic engagement and social commitment. For that reason, Jenkins argues that

“participative culture” is a “culture in which fans and other consumers are invited to actively participate in the creation and circulation of new content” (Jenkins, 2006, p. 331).

Nevertheless, the concept of “participation” is still an ambiguous term that expresses diverse interpretations. Academic debates suggest that there are various definitions about the idea of participation and different levels of participation (Arnstein, 1969; Dahlgren, 2012). As a consequence, this research aims to set out a theoretical framework that delimits the meaning of this polysemic term. In this way, this paper considers participation as a full and conscious action taken by an individual or by a community with the objective of getting specific results (Velásquez-Carrillo; González-Rodríguez, 2003). This means that civic participation differs from any other kind of social or individual action in that it is a civic expression characterized by a specific logic and principles (Espinosa-García, 2009). Moreover, citizen participation is also

“a categorical term for citizen power. It is the redistribution of power that enables the have-not citizens, presently excluded from the political and economic processes, to be deliberately included in the future” (Arnstein, 1969, p. 1).

The ladder of citizen participation designed by Arnstein (1969) represents a symbolic structure of the levels of civic participation that could converge in any kind of social, economic, or political system (table 1).

Table 1. Ladder of citizen participation

8. Citizen control	Citizen power
7. Delegated power	
6. Partnership	
5. Placation	Tokenism
4. Consultation	
3. Informing	
2. Therapy	Non participation
1. Manipulation	

Source: Arnstein (1969)

The five first steps of the ladder include “non participation” forms or an allegorical participation model in the case of *Tokenism*. According to Arnstein (1969), the first two depict a manipulated society where citizens perceive a distorted reality and live in a community controlled by political and economic elites. On the other hand, *Tokenism* is a distraction strategy because, although citizens find opportunities for expressing ideas and opinions, there is no interaction or feedback between citizenship and political power. The peak of the ladder represents three kinds of “citizen power”. In these last levels, participation becomes a tool that gives power to the individual and that guarantees public action legitimacy, justice, and effectiveness (Fung, 2006). It is at this point where the civic participation archetype works and where the open government model emerges: actors that take part in the communication process interact among themselves (Sundar; Kalyanaraman; Brown, 2003) and establish a mutually influential relationship.

“I-docs become the ideal agora for promoting various levels of participation and civic engagement”

#### 4. When i-docs met politics: towards an emerging open government model

There are two ways for citizens to participate in i-docs: “participation on media” and “participation through media” (Jenkins; Carpentier, 2013). The first, “participation on media”, refers to the inclusion of non-professionals in the decision making process around a platform –structural participation– or around content production –content participation–. On the other hand, “participation through media” focuses on the social dimension of the documentary: individuals finding opportunities to get involved in the social debate through the debate itself. This last expression introduces the dilemma of considering if participants show the abilities and the interest:

- in communicating in a dialogical space;
- in spreading a message through a documental or social network –spreadability- (Jenkins; Ford; Green, 2013);
- in connecting with other users from the same community, following the “networked society” paradigm (Castells, 2008); or
- popularizing their own social voice.

Regarding participation through media form, Nash (2014) claims there are two important concepts about the voice that become relevant for the study of the interactive documentary: “voice as authorship” –the ability of participants to be taken into account for the development of the documentary– and “voice as social participation” –the ability to connect and to commit to other users in the development of the documentary–. According to these possibilities for participation, Aston & Gaudenzi (2012) elaborated a taxonomy in which they distinguished four types of interactive documentaries:

- conversational (a conversation between users and the computer);

- hypertext (a limited user experience through preset choices);
- participative (a conversation between authors and users in which users are actively involved in the production of content); and
- experiential (users utilize space to structure and to personalize the user experience until the boundaries between real life and virtual life become diffused).

**Gaudenzi** (2013, p. 252) considers the participative mode the main model that motivates transformations in the documentary since the audience is able to change content and to promote the growth of the documentary as if it was a “living documentary”. In this way, Gaudenzi aims to emphasize the beginning of a new era for documentary production where the real *raison d’être* is not just the representation of the reality or the narration, but also the co-creation of reality or co-creation of the story. Along the same lines, **Darley** (2002, p. 165) summarized that

“the ability to choose to do something to change or alter the scene, to intervene in the action does constitute an increase in participation relative to the quiescent state of the usually sedentary viewer”.

For that reason, **Gaudenzi** (2013, p. 210) proposed three issues to be taken into account before analyzing the levels of participation in the interactive documentary: (1) Who is being called to collaborate? (2) What are they allowed to do? (3) When could they make the i-doc?

I-documentaries enable four levels of civic engagement: closed, semi-closed, semi-open, and open

The marriage between documentaries and politics is not necessarily an innovation. The history of documentaries is full of stories about how social movements have used the genre to participate in the public debate. In fact, **Whiteman** (2004) carried out some case studies in which individuals and activist groups popularized a kind of documentary that directs the spectator’s attention towards a specific social issue. In the same way, interactive documentaries are also starting to be considered a strategy for disseminating and sharing social petitions such as *Fort McMurray’s*. One of the most remarkable characteristics of i-docs is their ability to narrate a unified story; this is mainly what distinguishes i-docs from other multimedia web sites.

In the same way collective collaboration motivates the development of democracy, participation gives the audience the opportunity to change the narration of the story (**Nash**, 2014: 387). However, it is important to consider that the



<http://www.arte.tv/sites/en/webdocs/?lang=en>

levels of collaboration and participation in these virtual spaces depend on the ability and predisposition of the individuals who get involved in the interactive story and the architecture or interface of the documentary.

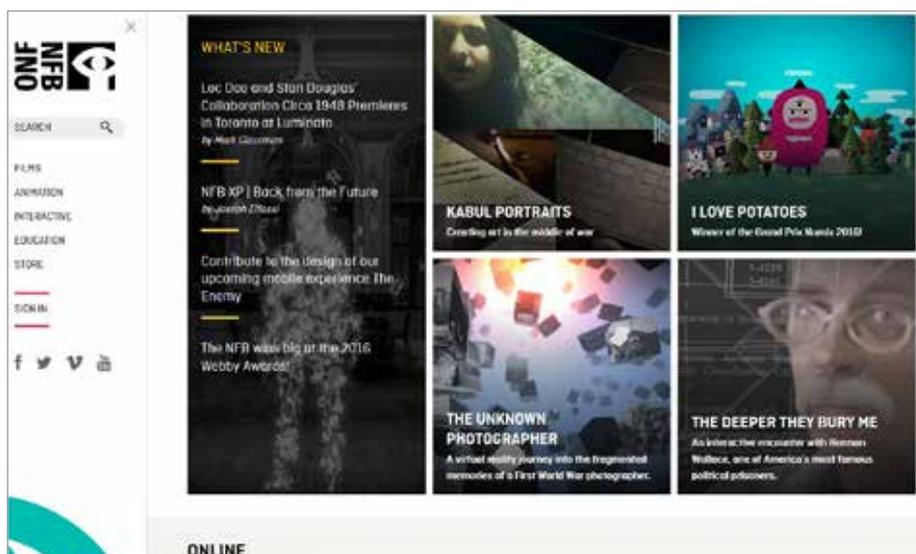
## 5. Methodology

This research has developed a hybrid methodology that combines the methodological approaches proposed by **Gifreu-Castells** (2011), **Gaudenzi** (2013), and **Nash** (2012b). The objective is to obtain a taxonomy of the civic participation forms that could arise from the marriage between i-docs and politics. Firstly, the model suggested by **Gifreu-Castells** (2011) is based on Nichols’ argument and considers that there are three elements that must be taken into account for the analysis of the interactive documentary: the role of the speaker, the message, and the receiver.

Secondly, as we said before, **Gaudenzi** (2013, p. 39-69) formulates the following questions to understand the interactivity level in documentaries: who is being called to participate, what they are allowed to do, and when the moment to accomplish the action is. The answers to these questions determine three modes of interactivity (**Gaudenzi**, 2013, p. 69):

- 1) semi-closed mode, where users can navigate but cannot modify the content;
- 2) semi-open mode, where users can participate in the content but they cannot change the structure of the documentary; and
- 3) open mode, where users and the documentary change constantly with each adapting to the other.

However, we suggest a fourth mode to this classification: the closed mode, regarding those documentaries where there is no interaction at all.



<https://www.nfb.ca/interactive>

Finally, **Nash** (2012b, pp. 195-210) deals with the social dimension of interactive documentaries through participation and highlights the importance of the ability of participants to connect and commit with other users.

In summary, this research has collected some of the most representative documentaries placed on the ARTE database, the National Film Board of Canada, the Open Documentary Lab at MIT, and other documentaries created by minor producers. Thus, the final collection includes the following documentaries: *Triumph des Willens* (*The Triumph of the Will*), *Journey to the end of coal*, *0 responsables* (*Zero responsible*), and *Global lives*. This selection is expected to be useful to illustrate the civic participation models that interactive documentaries allow.

“*Global lives* may be a suitable space to build an open government system because it allows the convergence of the traditional representative model and deliberative democracy”

## 6. Results and discussion

This section introduces an analysis of the documentaries that have been collected to devise a taxonomy about the different civic participation expressions. The results of this analysis reveal that documentaries enable four levels of civic engagement: closed, semi-closed, semi-open, and open. It is important to note that this taxonomy is based on a pure model; although we are aware of hybrid paradigms as well, we have not included them in this analysis because they are not representative or meaningful. This way, we can find a parallel between, for example, the *Triumph des Willens* documentary and the closed mode; *Journey to the end of coal*

and the semi-closed mode; *0 responsables* and the semi-opened mode, and *Global lives* and the open mode.

### 6.1. Closed models: the lack of participation in *Triumph des Willens*

*Triumph des Willens* (Leni Riefenstahl, 1935) is a linear documentary that relates to the 1934 Nazi Party Congress in Nuremberg and it represents a classic in documentary history. *Triumph des Willens* permits a cognitive interaction: spectators are only allowed to reflect upon the story narrated or to act according to what the documentary tells them, but they cannot interact, navigate, or add contents

that modify the structure of the documentary. For these reasons, *Triumph des Willens* constitutes a good example of a closed model. This documentary illustrates a kind of tale in which the speaker –the *Nazi Party* through Leni Riefenstahl– uses unidirectional communication channels to spread propaganda messages with the aim of reaching followers –in fact, *Nazi Party* supporters grew after this film–. In this case, there is no feedback possibility because the documentary presents a hierarchical communication structure. In conclusion, *Triumph des Willens* is a closed model that only motivates individuals in a cognitive mode.

### 6.2. Semi-closed models: *Journey to the end of coal* and *clicktivism*

*Journey to the end of coal* (Samuel Bollendorff and Abel Ségrétin, 2008) was produced by *Honkytonk films* and is an interactive documentary that shows the miserable working conditions of Chinese coal miners. The aim of this documentary is to expose the daily deaths that occur in mines and that have never been reported by the media. In order to achieve this objective, *Journey to the end of coal* provides information that is expected to motivate the audience to reflect upon the working conditions of Chinese coal miners, but the documentary does not ask spectators to respond. The structure of this model of i-doc is closed and hermetic, but allows users to inquire into the problem and delve into the idiosyncrasy of each character. Spectators find the opportunity to plan their own route when they interact with the various options the platform offers, but they cannot modify the story or the documentary structure. Gaudenzi named this kind of interactive documentary hypertext because it does not offer possibilities for real participation. The premise behind *Journey to the end of coal* is that users are just allowed to choose options and click on them. That is what we call “clicktivism”, which means to reduce activism to a mere mouse-click, and it does not represent a real engagement or commitment to the cause.

### 6.3. Semi-open models: *0 responsables* and the virtual agora

*0 responsables* (*Zero responsible*) was produced by *Barrret Films* and the *Asociación de Víctimas del Metro del 3 de Julio*<sup>3</sup>. It is a transmedia project composed through an interactive documentary, a lineal documentary for TV and an interactive product for SmartTV. *0 responsables* narrates the tragic Valencia railway accident that occurred on 3 July 2006. The project –available online– has five episodes that include pictures of the accident and testimonies of victims, journalists, and politicians.

The interactive documentary enables users to navigate through the structure of the platform, look up information and find testimonies. Although viewers cannot modify the documentary structure, they are allowed to participate in the story in two ways: First, individuals can provide documents, videos, and comments and post messages on social media like *Twitter* (#0responsables). In addition, users can also support the cause by participating in the virtual agora of the interactive documentary and joining the demonstrations organized every month in Plaza de la Virgen (Valencia).

In conclusion, the main objective of *0 responsables* is to report an unjust situation and encourage social mobilizations in order to get justice for the victims of the railway station and against the political elite who mismanaged the situation and did not provide solutions. Thanks to the civic engagement in this project and to the broadcast of a report on *Salvados* TV program, the case was re-opened and some people responsible for the accident were forced to declare this in the *Valencia Assembly*.

### 6.4. Open models: *Global lives*, how to create a documentary community

*Global lives* is a transmedia project –website, DVD, and exhibitions– founded in 2004 by David Evan Harris. The objective of this interactive documentary is to become a great database of video resources capable of reflecting the lifestyles of people from different cultures. This project aims to explore human diversity using video tools and to motivate people to debate, investigate, and think about cultures, ethnicities, languages, and religions.

This interactive documentary was made by volunteers from various countries and professionals: documentary makers, academics, journalists, photographers, etc. *Global lives* is

also a collaborative documentary because users are allowed to participate in the content and modify the structure of the platform. Spectators can add new content or just share their content via *Twitter*, *Facebook*, and other social media. The growth of these kinds of documentaries depends on the participation generated by users: user-generated-content strategy. Volunteers become the speakers of the story and the audience turns into a documentary community (Nash, 2014, p. 389). Publics are not just passive observers anymore, but active participants that produce and consume content. In this way, citizens become activists capable of participating in social petitions and can join any kind of cause.

“ In *Global lives*, citizens can collaborate in the structure of the documentary, but there is a committee that makes the final decision ”

In any case, what really makes this project different from the rest of the documentaries presented is that it permits users to participate in the structure; this means that volunteers can get involved in how the project is going to be articulated and organized, as well as participate in debates and forums. This is an example of how society is changing the way citizens participate in the public sphere, but also represents a pretext to motivate a politics’ renovation according to the claims of these emerging social trends.

In short, *Global lives* could be understood as a “living documentary” (Gaudenzi, 2013). It is a platform that accepts individual contributions that affect the content and the structure of the documentary. Moreover, it is a suitable space to build an open government system because it allows the convergence of a traditional representative model and the deliberative democracy (Barber, 1984; Habermas, 1999; Daly; Prugh; Costanza, 2000): citizens can collaborate in the structure of the documentary, but there is a committee that makes the final decision.

## 7. Conclusion

The objective of this paper is to emphasize the interactive potential of interactive documentaries to change the role of citizens from passive individuals to active political actors with their own voice. This way of social activism through i-

Table 2. Marriage between i-docs and politics implications

Level of civic participation	Participation forms allowed in documentaries taxonomy	Arnstein’s ladder of participation	Results from the marriage between i-docs and politics
Passive	Closed mode	1. Manipulation 2. Therapy	Orwellian model
Active	Semi-closed mode	3. Informing 4. Consultation 5. Placation	Tokenism model
	Semi-open mode	6. Partnership	Playful model
	Open mode	7. Delegated power 8. Citizen control	Athenian model

docs creates new spaces for an open government paradigm. In this context, this paper has developed a taxonomy based on the level of interaction that i-docs permit.

Regarding the role that citizens play in the interactive documentary we have identified two kinds of user profiles: passive and active. Passive citizens represent a low level of participation because they only get involved in a cognitive way. Spectators witness the progression of the story and are able to reflect on the information received, but they cannot interact. This passive role is typical in lineal documentaries and defines citizens who live in a closed government model, like in the *Triumph des Willens* documentary where communication is unidirectional and feedback is not permitted as users cannot change the content or the structure of the documentary.

The active citizen also gets involved at a cognitive level, but they interact as well. In this sense, citizens can modify the content, the structure of the documentary, or both elements at the same time. These different possibilities configure different government models: the semi-closed mode, like in *A journey to the end of coal* documentary. It constitutes a *Tokenism* strategy because citizens believe that they belong to a real participative democracy when, in fact, it is just a mirage. The semi-open mode, like *O responsables*, describes citizens who can participate in the story, but they cannot change its structure.

In conclusion, although there is still not a real open mode that represents an open government system, this research proposes that the marriage between i-docs and politics is the most suitable way to create an open social model, like for example in the *Global lives* documentary. Interactive documentaries such as *Global lives* allow citizens to participate in the content -the story- and to modify the structure of the documentary. *Global lives* offers an experience in which individuals get involved and engaged at all levels; for that reason, this documentary symbolizes the purest representation of the philosophical grounds for the open government model.

## Notes

1. Barnouw (1996); Barsam (1992); Nichols (1983; 1991; 2001); Renov (1993; 2004); Rotha (1952; 1970); Rabiger (1997); Català-Domènech (2001; 2008); Breschand (2002); Cerdán; Torreiro (2005); Cerdán; Torreiro (2001); Rosenthal; Corner (2002); León (1999); León; Negrodo (2013), among others.

2. Gaudenzi (2011; 2013); Rose (2014); Whitelaw (2002); Choi (2009); Dinmore (2008); Nash (2011; 2012a; 2012b; 2014); Winston (1995; 2000); or Britain (2006) in the Anglo Saxon countries. In Spain, scholars such as Gifreu-Castells (2011; 2012; 2013; 2014; 2015); Gifreu-Castells; Moreno (2014); León; Negrodo (2013); and Rodríguez-Fidalgo; Molpeceres-Arnáiz (2013) stand out.

3. This Spanish association was founded in 2006 and represents the victims of a railway accident in Valencia, in the south-east of Spain. The Valencia derailment occurred on 3 July 2006, when a commuter train travelling from Plaza de España station (Valencia) to Torrent (Valencia) derailed

at high speed on a bend about 50 metres before arriving at the Joaquin Sorolla-Jesús railway station at Valencia. Of the 150 people aboard, 43 were killed.

## 8. References

- Area-Moreira, Manuel; Ribeiro-Pessoa, Maria-Teresa (2012). "De lo sólido a lo líquido: las nuevas alfabetizaciones ante los cambios culturales de la web 2.0". *Comunicar*, v. 19, n. 38, pp. 13-20.  
<http://dx.doi.org/10.3916/c38-2012-02-01>
- Arnstein, Sherry (1969). "A ladder of citizen participation". *Journal of the American Institute of Planners*, v. 35, n. 4, pp. 216-224.  
<http://dx.doi.org/10.1080/01944366908977225>
- Aston, Judith; Gaudenzi, Sandra (2012). "Interactive documentary: Setting the field". *Studies in documentary film*, v. 6, n. 2, pp. 125-139.  
[http://dx.doi.org/10.1386/sdf.6.2.125\\_1](http://dx.doi.org/10.1386/sdf.6.2.125_1)
- Barber, Benjamin R. (1984). *Strong democracy: Participatory politics for a new age*. Berkeley: University of California Press. ISBN: 978 0520242333
- Barnouw, Eric (1996). *El documental: historia y estilo*. Barcelona: Gedisa. ISBN: 978 8474325638
- Barsam, Richard (1992). *Nonfiction film: A critical history*. Bloomington: Indiana University Press. ISBN: 978 0253207067
- Berrigan, Frances J. (1979). *Community communications. The role of community media in development*. Paris: Unesco. ISBN: 9231017713  
<http://unesdoc.unesco.org/images/0004/000440/044035eo.pdf>
- Breschand, Jean (2002). *Le documentaire: l'autre face du cinéma*. Paris: Cahiers du cinéma. ISBN: 978 2866423483
- Britain, Connor (2009). *Raising reality to the mythic on the Web: The future of interactive documentary film*. North Carolina: Elon University.  
<https://issuu.com/cbritain/docs/taaportfolio>
- Campos-Domínguez, Eva-María (2011). *El desarrollo de la ciberdemocracia en el Congreso de los Diputados: la comunicación e interacción entre ciudadanos y parlamentarios a través de internet (2004-2008)*. Tesis doctoral. Universidad Complutense de Madrid.  
<http://eprints.ucm.es/12401>
- Carpentier, Nico (2012). "The concept of participation. If they have access and interact, do they really participate?". *Fronteras*, v. 14, n. 2, pp. 164-177.  
<http://dx.doi.org/10.4013/fem.2012.142.10>
- Carpentier, Nico (2015). "Differentiating between access, interaction and participation". *Conjunctions*, v. 2, n. 2, pp. 7-28.  
<http://dx.doi.org/10.7146/tjcp.v2i2.23117>
- Castells, Manuel (2008). *La sociedad red*. Madrid: Alianza. ISBN: 978 84 206 4246 8
- Català-Domènech, Josep-Maria (2001). "La crisis de la reali-

dad en el documental español contemporáneo". En: Català-Domènech, Josep-Maria; Cerdán, Josetxo; Torreiro, Casimiro (eds.). *Imagen, memoria y fascinación. Notas sobre el documental en España*. Madrid: Ocho y Medio, pp. 27-44. ISBN: 9788493137687

**Català-Domènech, Josep-Maria** (2008). *La forma del real. Introducció als estudis visuals*. Barcelona: Universitat Obrera de Catalunya. ISBN: 978 8490297469

**Cerdán, Josetxo; Torreiro, Casimiro** (2005). *Documental y vanguardia*. Madrid: Cátedra. ISBN: 978 8437622309

**Choi, Insook** (2009). "Interactive documentary: A production model for nonfiction multimedia narratives". En: Nijholt, Anton; Reidsma, Dennis; Hondorp, Hendri (eds.). *Intelligent technologies for interactive entertainment*. Berlin: Springer, pp. 44-55. ISBN: 978 3642023149

**Corner, John** (2002). "Performing the real: documentary diversions". *Television and new media*, v. 3, n. 3, pp. 255-269. <http://dx.doi.org/10.1177/152747640200300302>

**Dader, José-Luís** (2002). *La ciberdemocracia posible. Reflexión prospectiva a partir de la experiencia en España*. <http://www.saladeprensa.org/art361.htm>

**Dahlgren, Peter** (2012). "Mejorar la participación: la democracia y el cambiante entorno de la Web". En: Innerarity, Daniel; Champeau, Serge (eds.). *Internet y el futuro de la democracia*. Barcelona: Paidós, pp. 45-67. ISBN: 978 8449327407

**Dahlgren, Peter** (2013). "From public to civic intellectuals via online cultures". *Participations. Journal of audience & reception studies*, v. 10, n. 1, pp. 400-404. <http://bit.ly/1TsKpGH>

**Daly, Herman; Prugh, Thomas; Costanza, Robert** (2000). *The local politics of global sustainability*. Washington: Island Press. ISBN: 9781559637442

**Darley, Andrew** (2002). *Visual digital culture: Surface play and spectacle in new media genres*. London: Routledge. ISBN: 978 0415165556

**Dinmore, Stuart** (2008). *The real online: Imagining the future of documentary*. School of Communication. Doctoral thesis. University of South Australia.

**Espinosa-García, Mario** (2009). "La participación ciudadana como una relación socio-estatal acotada por la concepción de democracia y ciudadanía". *Andamios*, v. 5, n. 10, pp. 71-109. <http://bit.ly/1TLgWbB>

**Fung, Archon** (2006). "Varieties of participation in complex governance". *Public administration review*, v. 66, n. S1, pp. 66-75. <http://dx.doi.org/10.1111/j.1540-6210.2006.00667.x>

**Gaudenzi, Sandra** (2011). "The i-doc as a relational object". *i-Docs*, 8 September. <http://goo.gl/yZmkpe>

**Gaudenzi, Sandra** (2013). *The living documentary: From representing reality to co-creating reality in digital interactive documentary*. Doctoral thesis. University of London. <http://research.gold.ac.uk/7997>

**Gifreu-Castells, Arnau** (2011). "The interactive documentary. Definition proposal and basic features of the new emerging genre". *McLuhan Galaxy conf. procs.*, pp. 367-378. ISBN: 978 8493880217 [https://www.academia.edu/1491044/The\\_Interactive\\_Documentary\\_Definition\\_Proposal\\_and\\_Basic\\_Features\\_of\\_the\\_New\\_Emerging\\_Genre](https://www.academia.edu/1491044/The_Interactive_Documentary_Definition_Proposal_and_Basic_Features_of_the_New_Emerging_Genre)

**Gifreu-Castells, Arnau** (2012). "Nuevo modelo de no ficción interactiva móvil. Caracterización del reportaje y el documental interactivo". En: García-Medina, Irene; Contreras-Espinosa, Ruth-Sofía (eds.). *M-todos. Tendencias y oportunidades de la movilidad digital*, pp. 59-66. ISBN: 978 8469520710 [http://repositori.uvic.cat/bitstream/handle/10854/1873/artconlli\\_a2011\\_mtodos.pdf?sequence=1](http://repositori.uvic.cat/bitstream/handle/10854/1873/artconlli_a2011_mtodos.pdf?sequence=1)

**Gifreu-Castells, Arnau** (2013). *El documental interactivo como nuevo género audiovisual. Estudio de la aparición del nuevo género, aproximación a su definición y propuesta de taxonomía y de modelo de análisis a efectos de evaluación, diseño y producción*. Tesis doctoral. Universitat Pompeu Fabra. [http://agifreu.com/interactive\\_documentary/TesisArnauGifreu2012.pdf](http://agifreu.com/interactive_documentary/TesisArnauGifreu2012.pdf)

**Gifreu-Castells, Arnau** (ed. y coord.) (2015). "Narrativas de no ficción audiovisual, interactiva y transmedia". *Obra digital*, n. 8. Vic: Universitat de Vic-Universitat Central de Catalunya. <http://revistesdigitals.uvic.cat/index.php/obradigital/issue/view/12/showToc>

**Gifreu-Castells, Arnau; Moreno, Valentina** (2014). "Estrategias y modelos de financiación del documental interactivo y transmedia". *Fonseca. Journal of communication*, n. 9, pp. 41-63. <http://revistas.usal.es/index.php/2172-9077/article/view/12241/12591>

**Grierson, John** (1966). *Grierson on documentary*. London: Forsyth Hardy Faber. ISBN: 978 0520005341

**Habermas, Jürgen** (1999). *Problemas de legitimación en el capitalismo tardío*. Madrid: Cátedra. ISBN: 978 8437617534

**Hibbing, John R.; Theiss-Morse, Elizabeth** (2002). *Stealth democracy: Americans' beliefs about how government should work*. Cambridge: Cambridge University Press. ISBN: 978 0521009867

**Jenkins, Henry** (2006). *Convergence culture: Where old and new media collide*. NY: NYU Press. ISBN: 978 0814742952

**Jenkins, Henry** (2014). "Participations: Dialogues on the participatory promise of contemporary culture and politics". Forum. Part 3: Politics. *International journal of communication*, v. 8, pp. 1129-1151. <http://goo.gl/t951aC> <http://ijoc.org/index.php/ijoc/article/view/2787/1124>

**Jenkins, Henry; Carpentier, Nico** (2013). "Theorizing participatory intensities: A conversation about participation and politics". *Convergence: The international journal of research into new media technologies*, v. 19, n. 3, pp. 265-286. <http://dx.doi.org/10.1177/1354856513482090>

- Jenkins, Henry; Ford, Sam; Green, Joshua** (2013). *Spreadable media: Creating value and meaning in a networked culture*. New York: New York University Press. ISBN: 978 0814743508
- Lasswell, Harold** (1948). "The structure and function of communications in society". En: Bryson, Lyman (Ed.). *The communication of ideas*. New York: Harper & Row, pp. 37-51.
- León, Bienvenido** (1999). *El documental de divulgación científica*. Barcelona: Paidós. ISBN: 978 8449307270
- León, Bienvenido; Negrodo, Samuel** (2013). "Documental web. Una nueva página para el sueño interactivo". *Telos: Cuadernos de comunicación e innovación*, n. 96, octubre 2013 - enero 2014, pp. 1-10.  
<https://goo.gl/GWkpoY>
- Nash, Kate** (2011). *Clicking on the real: telling stories and engaging audiences through interactive documentaries*.  
<http://goo.gl/Vht9oc>
- Nash, Kate** (2012a). "Goa hippy tribe: Theorising documentary content on a social network site". *Media international Australia*, n. 142, pp. 30-41.  
<http://dx.doi.org/10.1177/1329878X1214200105>
- Nash, Kate** (2012b). "Modes of interactivity: Analyzing the web-doc". *Media, culture and society*, v. 34, n. 2, pp. 195-210.  
<http://dx.doi.org/10.1177/0163443711430758>
- Nash, Kate** (2014). "What is interactivity for? The social dimension of web-documentary participation". *Continuum: Journal of media & cultural studies*, v. 28, n. 3, pp. 383-395.  
<http://dx.doi.org/10.1080/10304312.2014.893995>
- Nichols, Bill** (1983). "The voice of documentary". *Film quarterly*, v. 36, n. 3, pp. 17-30.  
<http://dx.doi.org/10.2307/3697347>
- Nichols, Bill** (1991). *Representing reality: Issues and concepts in documentary*. Bloomington: Indiana University Press. ISBN: 978 0253206817
- Nichols, Bill** (2001). *Introduction to documentary*. Bloomington: Indiana University Press. ISBN: 978 0 253 22260 2
- Rabiger, Michael** (1997). *Directing the documentary*. London: Focal Press. ISBN: 978 0240806082
- Renov, Michael** (1993). "Toward a poetics of documentary". En: Renov, Michael (ed.). *Theorizing documentary*. New York: Routledge, pp. 12-36. ISBN: 978 0415903820
- Renov, Michael** (2004). *The subject of documentary*. Minneapolis: University of Minnesota Press. ISBN: 978 0816634415
- Rodríguez-Fidalgo, María-Isabel; Molpeceres-Arnáiz, Sara** (2013). "Los nuevos documentales multimedia interactivos: construcción discursiva de la realidad orientada al receptor activo". *Historia y comunicación social*, v. 18, pp. 249-262.  
[http://dx.doi.org/10.5209/rev\\_hics.2013.v18.44325](http://dx.doi.org/10.5209/rev_hics.2013.v18.44325)
- Romaguera-Ramió, Joaquim; Alsina-Thevenet, Homero** (eds.) (1989). *Textos y manifiestos del cine. Estética, escuelas, movimientos, disciplinas, innovaciones*. Madrid: Cátedra. ISBN: 978 8437608457
- Rose, Mandy** (2014). "Making publics: Documentary as do-it-with-others citizenship". En: Ratto, Matt; Boler, Megan (eds.). *DIY citizenship: Critical making and social media*. Boston: MIT Press, pp. 201-212. ISBN: 978 0262026819
- Rosenthal, Alan; Corner, John** (2005). *New challenges for documentary*. Manchester: Manchester University Press. ISBN: 978 0520057241
- Rotha, Paul** (1952). *Documentary film: The use of the film medium to interpret creatively and in social terms the life of the people as it exists in reality*. London: Faber and Faber. ISBN: 978 0803815292
- Rotha, Paul** (1970). *Documentary film*. New York: Hasting House Publishers. ISBN: 978 1447439998
- Sampedro, Víctor; Sánchez, José-Manuel** (2011). "La Red era la plaza". *Ciberdemocracia.es*.  
<http://www.ciberdemocracia.es/articulos/RedPlaza.pdf>
- Servaes, Jan** (1996). "Participatory communication research with new social movements: A realistic utopia". En: Servaes, Jan; Jacobson, Thomas L.; White, Shirley A. (eds.). *Participatory communication for social change*. New Delhi: Sage, pp. 82-108. ISBN: 0803992963
- Servaes, Jan; Malikhao, Patchanee** (2005). "Participatory communication: The new paradigm?" En: Hemer, Oscar; Tufté, Thomas (eds.). *Media & global change. Rethinking communication for development*. Buenos Aires: Consejo Latinoamericano de Ciencias Sociales, pp. 91-103. ISBN: 9871183267  
<http://biblioteca.clacso.edu.ar/clacso/coediciones/20100824061923/glocal.pdf>
- Sundar, Shyman; Kalyanaraman, Sriram; Brown, Justin** (2003). "Explicating web site interactivity: Impression formation effects in political campaign sites". *Communication research*, v. 30, n. 1, pp. 30-59.  
<http://dx.doi.org/10.1177/0093650202239025>
- Velásquez-Carrillo, Fabio; González-Rodríguez, Esperanza** (2003). *¿Qué ha pasado con la participación ciudadana en Colombia?* Bogotá: Fundación Corona. ISBN: 958 971997X  
[http://www.dhl.hegoa.ehu.es/ficheros/0000/0120/participacion\\_ciudadana\\_en\\_colombia.pdf](http://www.dhl.hegoa.ehu.es/ficheros/0000/0120/participacion_ciudadana_en_colombia.pdf)
- Whitelaw, Mitchell** (2002). *Playing games with reality: Only fish shall visit and interactive documentary*.  
<http://mtchl.net/playing-games-with-reality-only-fish-shall-visit-and-interactive-documentary>
- Whiteman, David** (2004). "Out of the theatres and into the streets: A coalition model of the political impact of documentary film and video". *Political communication*, v. 21, n. 1, pp. 51-69.  
<http://dx.doi.org/10.1080/10584600490273263-1585>
- Winston, Brian** (1995). *Claiming the real: The documentary film revisited*. London: BFI. ISBN: 978 0851704647
- Winston, Brian** (2000). *Lies, damn lies and documentaries*. London: BFI. ISBN: 978 0851707976

# ANÁLISIS COMPUTACIONAL DE LAS CARACTERÍSTICAS ORTOTIPOGRÁFICAS Y PARALINGÜÍSTICAS DE LOS TWEETS PERIODÍSTICOS

## Computational analysis of the orthotypographic and paralinguistic characteristics of journalistic tweets

Esteban Vázquez-Cano, Javier Fombona y César Bernal



**Esteban Vázquez-Cano** es profesor del *Departamento de Didáctica, Organización Escolar y Didácticas Especiales* de la *UNED*. Doctor en ciencias de la educación con premio extraordinario. Ha publicado más de un centenar de trabajos científicos en las áreas del aprendizaje ubicuo y en el uso del lenguaje digital. Sus publicaciones más recientes abordan el tema del lenguaje digital en *Whatsapp* con el artículo: "Análisis lexicométrico de la especificidad de la escritura digital del adolescente en *Whatsapp*" y el libro *Dispositivos digitales móviles en educación. El aprendizaje ubicuo*, publicado en *Narcea* y considerado uno de los mejores libros en educación.  
<http://orcid.org/0000-0002-6694-7948>

*Universidad Nacional de Educación a Distancia, Facultad de Educación*  
Juan del Rosal, 14. 28040 Madrid, España  
[evazquez@edu.uned.es](mailto:evazquez@edu.uned.es)



**Javier Fombona** es profesor titular en la *Universidad de Oviedo*. Licenciado y doctor en ciencias de la información, y licenciado en filosofía y ciencias de la educación, sus trabajos abordan la incorporación de las nuevas tecnologías audiovisuales aplicadas a la educación. Ex-realizador de TV y profesor de didáctica y comunicación desde 1986 en varios niveles educativos, ha impartido múltiples comunicaciones, conferencias y publicaciones sobre estos temas. Investiga la narrativa en los nuevos soportes tecnológicos y coordina varios proyectos de investigación sobre dispositivos móviles avanzados.

<https://orcid.org/0000-0001-5625-5588>

*Universidad de Oviedo, Facultad de Formación del Profesorado y Educación*  
Aniceto Sela, s/n. 33005 Oviedo, España  
[fombona@uniovi.es](mailto:fombona@uniovi.es)



**César Bernal** es profesor de la *Universidad de Almería*, y colabora con la *UNED* en masters de competencia audiovisual y digital. Participa en varios proyectos de investigación, tales como "La competencia audiovisual de la ciudadanía andaluza. Estrategias de alfabetización mediática en la sociedad del ocio digital (*Alfa@Med*)". Ha publicado numerosos artículos y capítulos de libro sobre jóvenes y redes sociales, impacto social de las TIC, escenarios socio-educativos y de ocio de la sociedad de la información, tecnología educativa, integración curricular de los medios digitales en la formación docente, etc. Es evaluador en varias revistas.

<https://orcid.org/0000-0002-2802-1618>

*Universidad de Almería*  
*Facultad de Educación, Departamento de Educación*  
Ctra. Sacramento, s/n. 04120 La Cañada de San Urbano (Almería), España  
[cbernal@ual.es](mailto:cbernal@ual.es)

### Resumen

Se presenta un análisis las principales características ortotipográficas y paralingüísticas del tweet periodístico en una muestra de diez de los periodistas españoles más mediáticos y con mayor actividad en la red de microblogging *Twitter*. A través de una metodología de investigación que se enmarca dentro del análisis del discurso mediado por ordenador, se procesan 178 tweets periodísticos mediante técnicas de lingüística computacional (tokenización y lematización) y tratamiento estadístico del mensaje digital (estadística inferencial). Los resultados muestran que los elementos ortotipográficos con mayor representación son las comillas y puntos suspensivos en tweets de extensión media que no suelen sobrepasar los 100 caracteres. Asimismo, los elementos paralingüísticos más significativos son los *hashtags*, arrobas e imágenes en tweets con una densidad baja de palabras (9,07%).

Artículo recibido el 25-12-2015  
Aceptación definitiva: 29-03-2016

**Palabras clave**

Periodismo digital; Tweets; *Twitter*; Ortotipografía; Paralingüística; Lingüística computacional; Comunicación audiovisual; Lenguaje digital.

**Abstract**

The main orthotypographic and paralinguistic characteristics of journalistic tweets using a sample of ten of the most media active Spanish journalists in the network of microblogging *Twitter* are analyzed. Through a research methodology based on the computer-mediated discourse of computational linguistics and statistical processing of digital messages, 178 journalistic tweets are analyzed using computational linguistics techniques (tokenization and lemmatization) and statistical processing of digital messages (inferential statistics). The results show that the orthotypographic elements with greater representation are the quotes and suspension points in tweets of average length not exceeding 100 characters. Also, the most significant paralinguistic elements are the hashtags, ats, and images in tweets with a low density of words (9.07%).

**Keywords**

Digital journalism; *Twitter*; Orthotypography; Paralinguistics; Computational linguistics; Audiovisual communication; Digital language.

**Vázquez-Cano, Esteban; Fombona, Javier; Bernal, César (2016).** "Análisis computacional de las características ortotipográficas y paralingüísticas de los tweets periodísticos". *El profesional de la información*, v. 25, n. 4, pp. 588-598.

<http://dx.doi.org/10.3145/epi.2016.jul.08>

**1. Introducción**

El uso incesante de redes sociales y de aplicaciones de mensajería instantánea y microblogging está generando una nueva forma de escritura digital en todos los ámbitos de la vida: personal, social, académico y profesional. La red de microblogging *Twitter* ha irrumpido con fuerza entre el público en general y en el mundo periodístico en especial (**Ahmad**, 2010; **Lasorsa**; **Lewis**; **Holton**, 2012; **Broersma**; **Graham**, 2013). El flujo bidireccional de la información que se establece y la reacción que su contenido tiene en las redes sociales fomentan una interacción constante y una escritura digital que tiene que adaptarse a los nuevos escenarios audiovisuales y digitales y a la condensación formal de los 140 caracteres en redes como *Twitter* (**Honeycutt**; **Herring**, 2009; **Hong**; **Convertino**; **Chi**, 2011; **Pano-Alamán**; **Mancera-Rueda**, 2014). Estos flujos se complican al irrumpir en ellos la narrativa audiovisual, aplicando potentes estrategias comunicacionales extraídas de los medios de comunicación y de las redes sociales (**Fombona**; **Agudo**; **Fernández**, 2013).

El lenguaje periodístico, por lo tanto, se ve afectado por esta actividad en red que genera una nueva forma de construir el mensaje periodístico a través de un nuevo discurso digital que altera significativamente la construcción sintáctica, la selección léxica, la ortotipografía y se produce una nueva incorporación de elementos paralingüísticos (icónico-visuales) que enriquecen y modalizan la opinión, la noticia y la argumentación periodística (**Anderson**, 2012; **Androutopoulos**, 2014; **Domingo**; **Le-Cam**, 2014).

No se han realizado hasta la fecha estudios sobre la construcción de los tweets periodísticos desde el análisis de sus elementos ortotipográficos y paralingüísticos. En este estudio se presenta un análisis de la escritura digital periodística en la red *Twitter* en una muestra limitada a periodistas con alto impacto en *Twitter* (determinado por el número de seguidores y tweets retuiteados, entre otras variables) a tra-

vés de una metodología basada en la minería de textos y el análisis computacional y estadístico conforme a parámetros de tokenización, lematización y significatividad de variables formales y su incidencia en la construcción formal de los tweets periodísticos.

***Twitter*-lenguaje y *Twitter*-periodismo**

Desde su creación en 2006 por Jack Dorsey, *Twitter* se convirtió en un referente importante para el mundo. Sus usuarios podían comunicarse en tiempo real y seguir a personas, conociendo lo que les sucedía e interactuando con ellas con mensajes cortos de no más de 140 caracteres (**Java et al.**, 2007). *Twitter* como fenómeno social y tecnológico se encuentra a medio camino entre una red social al uso y un chat de actualidad mundial, lo que ha generado un código comunicativo y unas pautas de interacción específicas (**Mancera-Rueda**; **Pano-Alamán**, 2013). La encuesta *Join the conversation: How Spanish journalists are using Twitter* (**Carrera-Álvarez et al.**, 2012) realizada por un grupo de estudiantes de la *Universidad Carlos III de Madrid* concluye que *Twitter* es una de las redes sociales mejor vistas por los periodistas.

Hoy en día el concepto de escritura digital es poliédrico con multitud de situaciones y dispositivos en los que materializarla y en múltiples contextos digitales de carácter personal, académico y profesional en los que el ciudadano debe desenvolverse. La interacción tecnológica ha generado una escritura de la ubicuidad (**Vázquez-Cano**, 2012; **Vázquez-Cano**; **Mengual-Andrés**; **Roig-Vila**, 2015) y en el periodismo una especie de red de alerta temprana con un renacimiento de los antiguos flashes periodísticos (**De-Ramón-Carrión**, 2013). Este nuevo contexto ha fomentado que los profesionales del periodismo tengan que adaptar su estilo y sus técnicas de redacción a los requerimientos de estos nuevos canales y en el que los elementos ortotipográficos y paralingüísticos adquieren una nueva dimensión y función informativa (**Knaš**, 2009; **Thurlow**; **Poff**, 2011).

Elementos internos del tweet como la etiqueta (*hashtag*) y los enlaces a otras noticias en la Red posibilitan que el relato se enriquezca y adopte nuevas formas interpretativas. El empleo de alteraciones ortotipográficas, emoticonos o la inserción de elementos audiovisuales (fotos, fragmentos de audio, imágenes, ilustraciones, montajes, memes, etc.) puede ayudar a transmitir la intención comunicativa del periodista desde parámetros diferentes a los medios periodísticos más tradicionales.

El empleo de alteraciones ortotipográficas, emoticonos o la inserción de elementos audiovisuales puede ayudar a transmitir la intención comunicativa del periodista desde parámetros diferentes a los medios periodísticos tradicionales

En *Twitter* predomina la inmediatez comunicativa y se entrelaza la comunicación privada con la pública. Las implicaciones lingüísticas afectan a diferentes planos de la lengua: ortotipográfico, morfológico, sintáctico, semántico y pragmático. En el intercambio digital, síncrono y ubicuo no se controla la narración sino la interacción, favoreciendo una “ecología conversacional” (Boyd; Golder; Lotan, 2010; Markman, 2013). La autoría es compartida, pues las sucesivas intervenciones de los interlocutores son las que construyen la continuidad del relato, que acostumbra a ser improvisado, poco sopesado, con una sintaxis no muy elaborada y relajamiento ortográfico en el contexto informal (Gómez-Camacho, 2014). La investigación del uso del lenguaje en *Twitter* y, por extensión, en los subgéneros como el *Twitter*-periodismo es más factible que en otras redes sociales porque en ésta los mensajes presentan formas y contenidos más estables respecto a otros canales (Lomborg, 2011; Cortés-Rodríguez, 2012).

## 2. Elementos ortotipográficos y paralingüísticos en la escritura digital

A partir de una revisión de las principales investigaciones llevadas a cabo para el análisis del discurso mediado por ordenador, Androutsopoulos (2011) afirma que el aspecto que ha recibido mayor atención por parte de los investigadores es el carácter oral u oralizado de la lengua escrita en prácticamente todos los géneros electrónicos. Así, en el ámbito de la lengua española, Yus (2001) analiza el “texto escrito oralizado” y Sanmartín-Sáez (2007), la “conversación escrita”. A pesar de que el lenguaje oralizado es uno de los aspectos más significativos de los mensajes digitales que se intercambian en la Red, hay que analizar este tipo de mensajes y la presencia paralingüística en cada uno de los géneros y medios de comunicación conforme a la finalidad comunicativa de los hablantes, su situación y dispositivo desde el que se realiza el intercambio comunicativo (Almela, 2003; Vázquez-Cano; Mengual-Andrés; Roig-Vila, 2015).

Los signos paralingüísticos y ortotipográficos son un conjunto de signos no verbales que acompañan la comunicación lingüística y que complementan la comunicación, especial-

mente la interpersonal, a través de signos de puntuación y de expresión (: , ; ... ¿?), y en la comunicación digital se incrementan con el empleo de emoticonos y complementos audiovisuales (Vázquez-Cano; Mengual-Andrés; Roig-Vila, 2015). El propósito fundamental del empleo de los elementos paralingüísticos en la escritura consiste en restituir el elemento prosódico y socioemocional del discurso oral (Thurlow; Poff, 2011, p. 11).

En el discurso digital escrito la riqueza en símbolos, iconos, emoticonos y material audiovisual enriquece los recursos paralingüísticos de la escritura tradicional. Así podemos encontrar variaciones ortotipográficas y paralingüísticas con base en la alteración, ausencia y/o presencia de: letras, puntuación, acentuación, comillas, paréntesis, signos interrogación y exclamación, etiquetas temáticas (*hashtags*), arroba —que favorece el enlace al discurso escrito de otro interlocutor—, inclusión de imágenes, enlaces a la Red, acortamientos, símbolos y emoticonos, entre otros. A este tipo de alteraciones se les ha dado en otras lenguas como el inglés la denominación de *phoneticons* (Knaś, 2009, p. 2). La inclusión de este tipo de elementos paralingüísticos es bastante común en los tweets periodísticos, aunque no hay un acuerdo unánime entre los investigadores sobre el uso e importancia de cada uno de estos elementos en el discurso digital. Por ejemplo, para algunos autores el uso de las mayúsculas es poco relevante (Randall, 2002, p. 12), aunque otros consideran que su uso es significativo cuando el emisor quiere dejar constancia de su sorpresa o mostrar algún elemento afectivo (Thurlow; Mroczek, 2011). Otro elemento como es el uso de la puntuación adquiere diferentes interpretaciones que ayudan a modalizar el texto, principalmente desde el uso no canónico de signos de interrogación y exclamación, los puntos suspensivos y su combinación con los emoticonos (Huffaker; Calvert, 2005; Dresner; Herring, 2010; Menna, 2012).

Por lo tanto, la construcción del tweet que por su limitación de caracteres tiene si cabe una mayor dificultad de condensación del mensaje, requiere de elementos ortotipográficos y paralingüísticos que contextualicen el mensaje, que orienten su interpretación y que ayuden a mostrar en un número tan reducido de caracteres cuáles son las intenciones comunicativas de sus emisores.

## 3. Método

El objetivo de este artículo es establecer las características ortotipográficas y paralingüísticas que caracterizan la construcción lingüística de los tweets periodísticos y así poder determinar cuáles son los patrones de uso de la lengua en *Twitter* por parte de una muestra de periodistas españoles conforme a parámetros de tokenización y lematización y su tratamiento computacional y estadístico. Para ello se ha empleado una metodología de investigación que se enmarca dentro del análisis del discurso mediado por ordenador mediante técnicas de lingüística computacional para el análisis textual (Fletcher, 2004; Parodi, 2010; Vázquez-Cano; Mengual-Andrés; Roig-Vila, 2015) y procesamiento estadístico inferencial en aplicación al análisis de la construcción lingüística del mensaje digital. El análisis lingüístico realizado en este estudio ha constado de cuatro fases:

I) extracción de tweets y métricas de los periodistas en *Twitter*;

II) detección de parámetros de tokenización y lematización;

III) análisis estadístico descriptivo e inferencial del empleo de elementos ortotipográficos y paralingüísticos en los tweets periodísticos;

IV) determinación de la influencia del género en la muestra.

Fase I): se ha empleado la herramienta *TweetChup*, que permite analizar métricas de perfiles en *Twitter* en un período de tiempo limitado.

<http://tweetchup.com>

Posteriormente, hemos exportado los tweets al programa *Excel* para generar un archivo con extensión .xls que pueda ser sometido al análisis mediante técnicas de minería textual y estadística inferencial en la segunda fase del procedimiento metodológico.

Fase II): se ha empleado la API para minería de texto *Lemmatization, PoS and Parsing of Meaning Cloud*.

<https://www.meaningcloud.com/developer/lemmatization-pos-parsing>

Se han aplicado algoritmos automatizados para la tokenización (segmentación del tweet), y lematización (desambiguación de semas) para identificar los elementos paralingüísticos y ortotipográficos relevantes. El procedimiento para la obtención del texto tokenizado responde a una estructura matemática como la siguiente:

$$t_i = ("w_{1\oplus\boxtimes} w_{2\oplus\boxtimes} \dots \oplus w_{m\oplus\boxtimes} ")$$

donde  $t_i$  es una cadena "i<sup>th</sup> tweet" con un número 'm' de palabras determinado, el operador " $\oplus$ " marca las palabras separadas por espacios y el operador "x" marca el elemento paralingüístico separado por pausas.

Fase III): los resultados obtenidos se sometieron a análisis estadístico con el programa *SPSS 19*. Se analizó la "asimetría" y la "curtosis" y la posible significatividad de la ausencia y/o presencia de: letras, puntuación, acentuación, comillas, paréntesis, signos de interrogación y exclamación, *hashtags*, arrobas, inclusión de

Tabla 1. Periodistas y actividad en *Twitter*

Periodistas	Nº de seguidores	Tweets en 14 días	Promedio de tweets por día	Veces que los tweets fueron retuiteados
Pedro J. Ramírez	342.465	464	33,1	8.451
Juan Ramón Lucas	215.958	107	7,6	2.488
Ignacio Escolar	500.134	297	21,2	51.810
Jesús Maraña	138.059	191	13,6	12.376
Melchor Miralles	103.578	135	9,6	371
Pepa Bueno	140.195	268	19,1	24.301
Ana Pastor	1.334.927	1.068	76,3	104.015
Susana Griso	403.669	147	10,5	6.200
Esther Palomera	57.120	116	8,3	3.350
Julia Otero	529.178	108	7,7	8.227
Media	37.652,83	290,1	20,7	22.158,9
Total	3.765.283	2.901	207	221.589

imágenes, enlaces a la Red, acortamientos, símbolos y emoticonos; mediante múltiples pruebas de regresión lineal.

Fase IV): se analizó si existían diferencias significativas atendiendo a la variable género.

#### 4. Muestra y selección

El criterio de selección de los 10 periodistas se basó en los siguientes criterios:

- número de seguidores;
- promedio de tweets por día;
- número de veces que los tweets fueron retuiteados.

Se eligieron 5 periodistas varones y 5 periodistas mujeres:

- Pedro J. Ramírez
- Juan Ramón Lucas
- Ignacio Escolar
- Jesús Maraña

<b>Pedro J. Ramírez</b>	22:48 - 27 de may./22:52 - 28 de may./14:04 - 7 de jun. de 2015/22:36 - 10 de jun./13:05 - 29 de may./22:38 - 1 de jun./11:26 - 1 de jun./22:40 - 2 de jun./22:47 - 9 de jun./0:06 - 2 de jun./22:52 - 3 de jun./22:48 - 4 de jun. /0:21 - 7 de jun./6:25 - 6 de jun./13:47 - 29 de may./1:52 - 29 de may./1:01 - 31 de may./13:06 - 28 de may./23:01 - 28 de may.
<b>Juan R. Lucas</b>	3:12 - 30 de may./10:43 - 30 de may./14:24 - 6 de jun./9:24 - 30 de may./22:11 - 9 de jun./14:16 - 6 de jun./0:09 - 4 de jun./4:16 - 31 de may./3:46 - 6 de jun./0:38 - 31 de may./10:27 - 29 de may./2:04 - 7 de jun./7:35 - 7 de jun./10:19 - 30 de may./22:29 - 1 de jun.
<b>Ignacio Escolar</b>	9:49 - 28 de may./4:12 - 29 de may./8:36 - 5 de jun./8:25 - 28 de may./9:57 - 30 de may./4:26 - 28 de may./4:52 - 9 de jun./2:57 - 2 de jun./0:48 - 6 de jun./13:52 - 9 de jun./0:02 - 9 de jun./2:49 - 29 de may./2:23 - 28 de may./3:25 - 29 de may./3:50 - 3 de jun./11:52 - 1 de jun./10:02 - 30 de may./5:57 - 5 de jun.
<b>Jesús Maraña</b>	14:15 - 6 de jun./7:58 - 11 de jun./6:27 - 29 de may./10:56 - 2 de jun./2:56 - 1 de jun./14:54 - 5 de jun./1:13 - 3 de jun./9:20 - 8 de jun./15:03 - 3 de jun./1:52 - 4 de jun./2:51 - 2 de jun./1:19 - 3 de jun./1:38 - 2 de jun./14:27 - 1 de jun.
<b>Melchor Miralles</b>	15:28 - 3 de jun./3:48 - 28 de may./23:34 - 7 de jun./9:41 - 31 de may./6:13 - 9 de jun./14:42 - 30 de may./2:09 - 31 de may./15:49 - 28 de may./18:35 - 4 de jun./13:17 - 29 de may./2:08 - 31 de may./23:31 - 8 de jun./9:14 - 4 de jun./12:14 - 7 de jun./15:27 - 3 de jun./16:38 - 9 de jun.
<b>Pepa Bueno</b>	14:21 - 7 de jun./13:48 - 31 de may./14:04 - 31 de may./13:46 - 31 de may./14:08 - 31 de may./12:42 - 31 de may./14:33 - 7 de jun./13:29 - 31 de may./13:12 - 31 de may./13:28 - 7 de jun./11:30 - 7 de jun./13:01 - 31 de may./13:38 - 31 de may./13:33 - 30 de may./14:16 - 7 de jun./13:01 - 7 de jun./14:03 - 31 de may./14:33 - 7 de jun./13:36 - 31 de may.
<b>Ana Pastor</b>	5:01 - 30 de may./6:53 - 1 de jun./11:41 - 3 de jun./13:13 - 9 de jun./7:42 - 2 de jun./12:29 - 9 de jun./3:32 - 9 de jun./6:24 - 29 de may./13:37 - 1 de jun./13:32 - 7 de jun./13:29 - 8 de jun./0:48 - 10 de jun./4:15 - 4 de jun. 0:57 - 9 de jun./4:19 - 1 de jun./1:56 - 28 de may.
<b>Susana Griso</b>	5:11 - 9 de jun./4:20 - 4 de jun./5:28 - 29 de may./3:53 - 29 de may./5:04 - 9 de jun./4:47 - 11 de jun./1:48 - 4 de jun./1:51 - 11 de jun./1:56 - 4 de jun./1:41 - 29 de may./3:47 - 9 de jun./3:32 - 4 de jun./1:47 - 8 de jun./3:15 - 4 de jun./3:47 - 9 de jun./1:43 - 10 de jun./4:26 - 1 de jun.
<b>Esther Palomera</b>	6:07 - 11 de jun./23:38 - 1 de jun./22:13 - 7 de jun./13:29 - 9 de jun./22:23 - 2 de jun./8:28 - 9 de jun./3:45 - 2 de jun./2:07 - 30 de may./7:22 - 3 de jun./2:09 - 30 de may./0:10 - 31 de may./12:38 - 28 de may./22:48 - 7 de jun./23:44 - 3 de jun./22:34 - 31 de may./1:12 - 29 de may./11:16 - 2 de jun.
<b>Julia Otero</b>	15:50 - 8 de jun./12:41 - 7 de jun./15:55 - 8 de jun./12:37 - 7 de jun./5:16 - 5 de jun./16:00 - 8 de jun./3:16 - 3 de jun./14:24 - 7 de jun./14:40 - 31 de may./0:02 - 9 de jun./11:33 - 7 de jun./14:09 - 28 de may./14:06 - 31 de may./11:29 - 7 de jun./14:16 - 7 de jun./5:09 - 5 de jun./13:12 - 7 de jun.
<b>Total</b>	<b>10 periodistas / 178 tweets</b>

Figura 1. Periodistas y tweets analizados

- Melchor Miralles
- Pepa Bueno
- Ana Pastor
- Susana Griso
- Esther Palomera
- Julia Otero.

La tabla 1 muestra los datos de la actividad y repercusión de los tweets de cada periodista en un período aleatorio de 14 días entre el 28 de mayo y el 11 de junio de 2015.

Los tweets que han sido analizados en este estudio se presentan en la figura 1 con indicación del periodista que lo emite y la fecha y hora en la que lo hace durante el año 2015.

Tabla 2. Caracterización formal y lingüística de la muestra

Periodistas	Total de palabras	Total de caracteres	Palabras diferentes	Promedio de palabras en oración prototípica	Oraciones empleadas
Pedro J. Ramírez	419	2.018	238	11,82	40
Juan R. Lucas	181	790	120	7,60	30
Ignacio Escolar	291	1.874	190	10,13	33
Jesús Maraña	252	1.284	126	10,26	21
Melchor Miralles	180	825	105	8,87	8
Pepa Bueno	280	1.337	180	8,36	18
Ana Pastor	265	1.272	159	7,66	38
Susana Griso	210	1.045	137	7,83	30
Esther Palomera	192	993	131	10,14	21
Julia Otero	253	1.272	156	7,89	35
Media	252,3	1.271	154,2	8,042	27,4
Total	2.523	12.710	1.542	80,42	274

### 5. Resultados y discusión

En primer lugar, se muestran los datos descriptivos y formales de los tweets emitidos por los 10 periodistas analizados. Para ello se procedió mediante tokenización y lematización a establecer una caracterización formal lingüística de todos los tweets (tabla 2).

Como podemos comprobar en la tabla 2, los periodistas varones y mujeres emplean un número total de palabras muy parecido (varones: 1.323) y (mujeres: 1.200). Asimismo, el número de oraciones empleadas en el conjunto de tweets es muy similar (varones: 132) y (mujeres: 142); así como el empleo de palabras diferentes (varones: 779) y (mujeres: 763). Esto muestra que la diferencia entre sexos en el empleo formal del tweet en las categorías analizadas es muy similar y no se producen diferencias significativas. Por el contrario, sí es significativa la diferencia entre periodistas de forma individual. El periodista con un empleo más acusado de palabras y oraciones es Pedro J. Ramírez (419 palabras en 40 oraciones con una densidad media de palabras por oración de 11,82). Los dos periodistas con un menor empleo de palabras y oraciones son respectivamente:

- Melchor Miralles (180 palabras en 8 oraciones con una densidad media de palabras por oración del 8,87);

- Juan Ramón Lucas (181 palabras en 30 oraciones con una densidad media de palabras por oración del 7,6).

En primer lugar, el proceso de tokenización permite suprimir las palabras bien escritas o que no presentan ninguna alteración, eliminando los espacios para obtener la cadena de elementos ortotipográficos y paralingüísticos constituyente de cada tweet conforme a la fórmula (tabla 3):

$$t_i = ("w_{1\oplus\boxtimes} w_{2\oplus\boxtimes} \dots \oplus w_{m+\boxtimes} ")$$

Estos resultados se sometieron a análisis estadístico para determinar si el empleo y/o alteración de estos elementos en los tweets comportaba significación estadística con respecto al número total de caracteres posibles en su construcción. Para ello, presentamos los estadísticos descriptivos de la muestra tomando en consideración si la asimetría y curtosis son significativas (tabla 4).

El análisis de la asimetría y curtosis de una muestra lingüística nos permite identificar si los datos se distribuyen de forma uniforme alrededor de la media. En la muestra analizada, los resultados muestran que la media es mayor que la mediana y esto genera una asimetría y curtosis po-

Tabla 3. Resultados del proceso de tokenización de elementos ortotipográficos

Periodistas	Letra	Puntuación	Acentuación	Puntos suspensivos	Comillas	Paréntesis	Acortamiento	Signo de exclamación
Pedro J. Ramírez	0	2	0	1	2	2	1	8
Juan R. Lucas	0	4	0	1	2	0	0	1
Ignacio Escolar	0	6	0	0	2	1	0	0
Jesús Maraña	0	0	0	0	2	1	0	0
Melchor Miralles	2	10	0	0	6	0	0	0
Pepa Bueno	0	0	0	2	14	0	1	0
Ana Pastor	0	0	0	6	3	0	0	4
Susana Griso	0	0	0	11	4	1	0	0
Esther Palomera	1	1	1	0	2	1	0	5
Julia Otero	1	2	0	0	2	0	1	0

Tabla 4. Estadísticos descriptivos de la densidad de los elementos ortotipográficos

		Letra	Puntuación	Acentuación	Puntos suspensivos	Comillas	Paréntesis	Acortamiento	Signo de exclamación
Nº	Válidos	10	10	10	10	10	10	10	10
	Perdidos	0	0	0	0	0	0	0	0
Media		1,0000	1,9000	0,1000	2,1000	3,8000	0,7000	1,8000	0,3000
Mediana		0,0000	0,5000	0,0000	0,5000	2,0000	0,5000	0,0000	0,0000
Desviación típica.		1,88562	3,14289	0,31623	3,63471	3,85285	0,82327	2,85968	0,48305
Varianza		3,556	9,878	0,100	13,211	14,844	0,678	8,178	0,233
Asimetría		2,486	2,262	3,162	2,069	2,478	0,687	1,462	1,035
Curtosis		6,572	5,480	10,000	3,921	6,518	-1,043	1,109	-1,224
Mínimo		0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00
Máximo		6,00	10,00	1,00	11,00	14,00	2,00	8,00	1,00
Suma		10,00	19,00	1,00	21,00	38,00	7,00	18,00	3,00

sitiva (leptocúrtica) en las categorías: “letra”, “puntuación”, “acentuación”, “puntos suspensivos”, “comillas” y “acortamientos”; y, por lo tanto, presenta un elevado grado de concentración alrededor de los valores centrales de la variable, por lo que estos elementos ortotipográficos se distribuyen de forma más homogénea entre los diferentes tweets analizados y podemos considerarlos de distribución y empleo más general entre los periodistas. Por el contrario, la curtosis es negativa en las categorías: “paréntesis” y “signos de exclamación”; lo que genera una distribución platicúrtica, es decir, con una reducida concentración alrededor de los valores centrales de la distribución. Esto nos indica que los “paréntesis” y los “signos de exclamación” se emplean de forma más reducida en los tweets periodísticos.

Posteriormente, analizamos mediante sucesivas pruebas de regresión lineal simple la significatividad de cada uno de los elementos ortotipográficos conforme al número total de caracteres por tweet. De esta manera, podemos delimitar la significatividad o no de la presencia o alteración de elementos ortotipográficos con respecto a la estructura formal del tweet (tabla 5).

Los resultados de las múltiples regresiones lineales muestran que ninguna de las categorías ortotipográficas anali-

zadas es altamente significativa en la construcción de los tweets periodísticos. Entendiendo esta significatividad por su presencia es necesaria u obligatoria para su construcción. Esto muestra que son elementos modalizadores del discurso pero no esenciales para su construcción. El empleo más generalizado entre los periodistas es el de “comillas” y “puntos suspensivos” (figura 2). Estos signos aparecen con mayor frecuencia en tweets de extensión media que no suelen sobrepasar los 100 caracteres.

Las “comillas” se suelen emplear con su uso canónico para indicar citas en estilo directo, títulos de artículos, etc. y los “puntos suspensivos” para aspectos de suspensión del discurso con diversa interpretación pragmática. La figura 3 muestra algunos ejemplos de tweets que emplean comillas y puntos suspensivos en su construcción.

El tweet remitido por periodistas tiene un índice bajo de error en la ortografía de la letra y en la acentuación ( $n = 10 / n = 1$ ). Tienen una incidencia ligeramente más acusada las alteraciones o ausencias en los signos de puntuación ( $n = 19$ ), principalmente en la ausencia del punto al final del tweet o en la ausencia del punto en los acortamientos. Estos resultados coinciden con otros estudios realizados en español como los de **De-Ramón-Carrión** (2013). Las alteraciones

de la ortografía se pueden considerar heterografías (**Martínez-de-Sousa, 2004**) y constituyen lo que podríamos llamar desviaciones intencionadas. Asimismo, el uso de acortamientos no es muy acusado ( $n = 18$ ) en comparación con otros estudios de escritura digital (**Vázquez-Cano;**

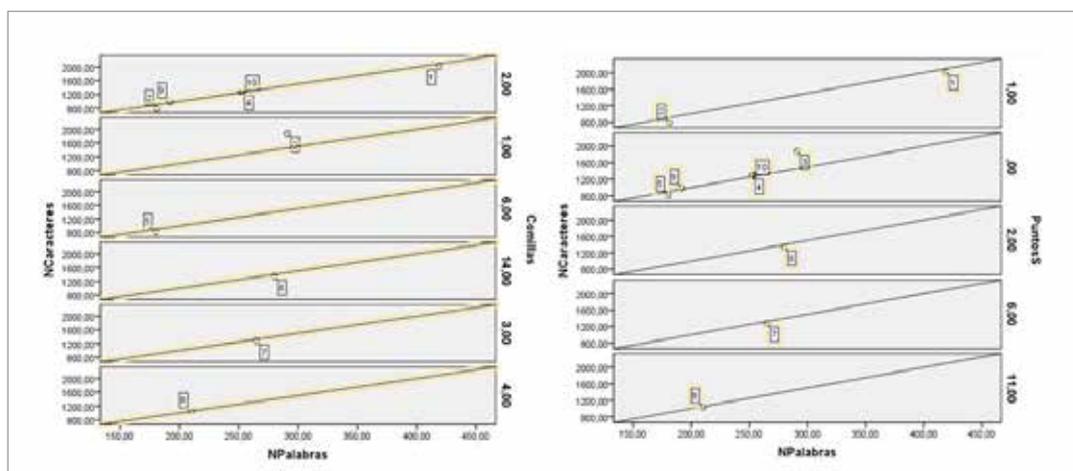


Figura 2. Densidad de comillas y puntos suspensivos en los tweets periodísticos

Tabla 5. Regresión lineal de los elementos ortotipográficos de los tweets periodísticos

Ortotipo-grafía	R	R cuadrado	R cuadrado corregida	Error típico de la estimación	Estadísticos de cambio					Durbin-Watson
					Cambio en R cuadrado	Cambio en F	gl1	gl2	Significatividad cambio en F	
Letra	0,355 <sup>a</sup>	0,126	0,017	402,70289	0,126	1,156	1	8	0,314	2,222
Punt.	0,450 <sup>a</sup>	0,202	0,102	384,80263	0,202	2,027	1	8	0,192	2,109
Acentuación	0,240 <sup>a</sup>	0,058	-0,060	418,16726	0,058	0,491	1	8	0,503	2,592
Puntos suspensivos	0,157 <sup>a</sup>	0,025	-0,097	425,48871	0,025	0,201	1	8	0,666	2,527
Comillas	0,145 <sup>a</sup>	0,021	-0,101	426,24066	0,021	0,172	1	8	0,689	2,563
Paréntesis	0,734 <sup>a</sup>	0,539	0,481	292,58210	0,539	9,345	1	8	0,016	2,149
Signo exclamación	0,461 <sup>a</sup>	0,213	0,114	382,30552	0,213	2,159	1	8	0,180	2,246
Acortamiento	0,393 <sup>a</sup>	0,155	0,049	396,13467	0,155	1,462	1	8	0,261	2,649

Mengual-Andrés; Roig-Vila, 2015). Los paréntesis y signos de exclamación son utilizados en su forma canónica y su uso es relativamente bajo (paréntesis n = 7 / signos de exclamación n = 3).

Los *hashtags*, arrobas e imágenes son los elementos que más contribuyen a otorgar información adicional y complementaria al tweet periodístico

Otro aspecto importante en la construcción del tweet es delimitar la presencia y significatividad de los elementos paralingüísticos característicos de la escritura digital (iconos, emoticonos, imágenes, enlaces) y aquellos elementos más significativos del tweet como son la arroba y el *hashtag* (etiqueta). El proceso de tokenización nos ha permitido exportar al programa *Excel* los elementos paralingüísticos por periodista que se muestran en la tabla 6.

Estos resultados, al igual que los elementos ortotipográficos, se sometieron a análisis estadístico para determinar si su empleo en los tweets comportaba significación estadística con respecto al número total de caracteres posibles en su construcción. Para ello, se presentan en la tabla 7 los

Tabla 6. Resultados del proceso de tokenización de elementos paralingüísticos

Periodistas	Hashtag y @	Imagen	Enlace	Emoticonos
Pedro J. Ramírez	7	0	3	0
Juan R. Lucas	16	7	4	3
Ignacio Escolar	0	0	14	0
Jesús Maraña	23	5	2	0
Melchor Miralles	19	0	2	0
Pepa Bueno	30	0	0	0
Ana Pastor	24	5	6	4
Susana Griso	7	18	0	0
Esther Palomera	21	11	4	0
Julia Otero	16	9	2	1

estadísticos descriptivos de la muestra tomando en consideración si la asimetría y curtosis son significativas.

Los resultados muestran que la media es mayor que la mediana y esto genera una asimetría y curtosis positiva (leptocúrtica) en las categorías: "imagen", "enlace" y "emoticonos"; y por lo tanto, estos elementos paralingüísticos se distribuyen de forma más homogénea y general entre los diferentes

tweets analizados. El uso del *hashtag* y la arroba obtiene una curtosis negativa pero es la categoría más recurrente en el recuento (n = 163) y sigue una distribución normal (figura 4). Esto indica que son los elementos paralingüísticos con mayor representatividad en los tweets periodísticos.

Posteriormente, se analizó mediante sucesivas pruebas de regresión lineal simple la significatividad de cada uno de los ele-

Periodistas	Tuits	Ncar.	C/PS <sup>1</sup>
Juan R. Lucas	En <a href="#">@elconfidencial</a> : Hablan las esclavas sexuales del ISIS: "Nos olían, nos violaban, nos torturaban..."	103	C
Ignacio Escolar	Una concejala valenciana del PP: "Ahora empezarán la quema de iglesias y la violación de monjas"	96	C
Jesús Maraña	"Los antisistema del sistema". El artículo de El Gran Wyoming no tiene desperdicio.	83	C
Pepa Bueno	"Aznar fue quien vendió armas al gobierno de Chávez" <a href="#">#ChesterBono</a>	65	C
Ana Pastor	Preparando la entrevista del domingo con <a href="#">@ZaidaCantera</a> .. Madre mía.. Y este tipo encima la denuncia...	103	PS
Susana Griso	Y al tercer día....salió.	25	PS
Esther Palomera	C's afirma que la imputación de Victoria y Figar complica "cualquier acercamiento" con el PP (EP, EFE)	102	C

<sup>1</sup>C/PS (Comillas/Puntos Suspensivos) Ncar. (Número de caracteres)

Figura 3. Tweets con elementos ortotipográficos (comillas y puntos suspensivos)

Tabla 7. Estadísticos descriptivos de la densidad de los elementos paralingüísticos

		Hashtag y @	Imagen	Enlace	Emoticonos
N	Válidos	10	10	10	10
	Perdidos	0	0	0	0
Media		16,3000	5,5000	3,7000	0,8000
Mediana		17,5000	5,0000	2,5000	0,0000
Desviación típica		9,18997	5,98609	4,05654	1,47573
Varianza		84,456	35,833	16,456	2,178
Asimetría		-0,445	0,971	2,057	1,722
Curtosis		-0,458	0,615	5,077	1,696
Mínimo		0,00	0,00	0,00	0,00
Máximo		30,00	18,00	14,00	4,00
Suma		163,00	55,00	37,00	8,00

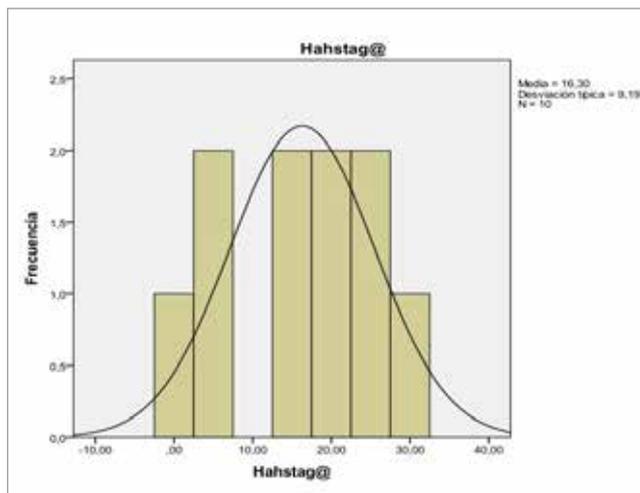


Figura 4. Densidad de *hashtags* y arrobas en los tweets

mentos paralingüísticos conforme al número total de caracteres por tweet (tabla 8). De esta manera, podemos delimitar la significatividad o no de la presencia o alteración de estos elementos con respecto a la estructura formal del tweet.

Los resultados de las múltiples regresiones muestran, al igual que para los elementos ortotipográficos, que no existe significatividad de los elementos paralingüísticos con respecto a la estructura formal del tweet (por su referencia al número de caracteres totales, palabras o número de tweets).

Los *hashtags*, arrobas e imágenes son los elementos que más contribuyen a otorgar información adicional y complementaria al tweet periodístico. Estos tres elementos paralingüísticos ya habían sido reseñados como significativos en la construcción formal del tweet en todos los subgéneros (Honeycutt; Herring,

2009). Por lo tanto, aunque su presencia no es obligatoria para la construcción formal de los tweets periodísticos, sí tienen un empleo acusado cuando el periodista quiere orientar el contenido. Es significativo que estos tres elementos paralingüísticos se utilicen con mayor frecuencia en tweets con baja densidad de palabras (9,07%) (figura 5).

La figura 6 muestra algunos ejemplos de tweets que emplean *hashtags*, arrobas e inserción de imágenes en su construcción.

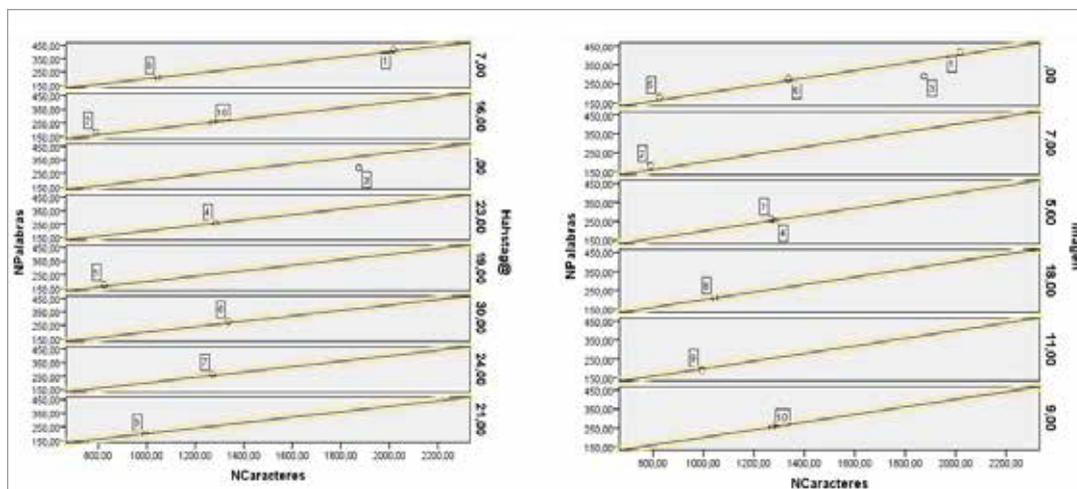


Figura 5. Densidad de *hashtags/arrobas* e imágenes en los tweets

Tabla 8. Regresión lineal de los elementos paralingüísticos

Elemento paralingüístico	R	R cuadrado	R cuadrado corregida	Error típico de la estimación	Estadísticos de cambio					Durbin-Watson
					Cambio en R cuadrado	Cambio en F	gl1	gl2	Signif. cambio en F	
Hashtag/@	0,478 <sup>a</sup>	0,228	0,132	378,51910	0,228	2,363	1	8	0,163	2,586
Imagen	0,476 <sup>a</sup>	0,226	0,130	378,89619	0,226	2,342	1	8	0,164	2,386
Enlace	0,458 <sup>a</sup>	0,210	0,111	383,00944	0,210	2,121	1	8	0,183	2,472
Emoticonos	0,267 <sup>a</sup>	0,071	-0,045	415,22279	0,071	0,612	1	8	0,457	2,199

Tabla 9. Prueba de Mann-Whitney (diferencias entre la variable género)

	Letra	Puntuación	Acen-tuación	Puntos suspensivos	Comillas	Paréntesis	Signo de exclamación	Acortamiento	Hashtag /@	Imagen	Enlace
U de Mann-Whitney	10,500	7,500	10,000	7,000	7,000	8,000	10,000	12,500	7,000	5,000	8,500
W de Wilcoxon	25,500	22,500	25,000	22,000	22,000	23,000	25,000	27,500	22,000	20,000	23,500
Z	-0,472	-1,118	-1,000	-1,230	-1,226	-1,021	-0,655	0,000	-1,156	-1,622	-0,851
Sig. asintót. (bilateral)	0,637	0,264	0,317	0,219	0,220	0,307	0,513	1,000	0,248	0,105	0,395
Sig. exacta [2*(Sig. unilateral)]	0,690 <sup>a</sup>	0,310 <sup>a</sup>	0,690 <sup>a</sup>	0,310 <sup>a</sup>	0,310 <sup>a</sup>	0,421 <sup>a</sup>	0,690 <sup>a</sup>	1,000 <sup>a</sup>	0,310 <sup>a</sup>	0,151 <sup>a</sup>	0,421 <sup>a</sup>

Por último, otro aspecto que nos parecía pertinente analizar era delimitar si existían diferencias significativas en la construcción de los tweets periodísticos según la variable género. Para ello, realizamos la prueba estadística no paramétrica de *Mann-Whitney* que permite incluso en muestras de 10 sujetos establecer resultados satisfactorios (Gómez-Gómez; Danglot-Banck; Vega-Franco, 2003) (tabla 9).

Los resultados de la prueba muestran que no hay diferencias significativas conforme a la variable género en el empleo de elementos ortotipográficos y paralingüísticos ya que los niveles de significación asintótica no son significativos al nivel alfa de 0,05.

### 6. Conclusión

La investigación realizada muestra que 10 de los periodistas con mayor repercusión en la red social *Twitter* realizan una media de 20,7 tweets por día con una estructura formal de 252,3 palabras y 1.271 caracteres de media de forma homogénea entre varones y mujeres. Conforme a estas variables de construcción formal, se ha analizado la influencia y significatividad que tienen los elementos ortotipográficos y paralingüísticos en la modalización de los tweets periodísticos. Los resultados de las múltiples regresiones lineales de los elementos ortotipográficos y paralingüísticos muestran que ninguna de las categorías analizadas es altamente significativa en la construcción de los tweets periodísticos. Entendiendo esta significatividad por si su presencia es necesaria u obligatoria para su construcción. Esto muestra que son elementos modalizadores del discurso, pero no esenciales para su construcción. Con respecto a los elementos ortotipográficos, el empleo más generalizado entre los periodistas es el de “comillas” y “puntos suspensivos” en tweets de extensión media que no suelen sobrepasar los 100 caracteres.

Asimismo, el tweet remitido por periodistas tiene un índice bajo de error en la ortografía de la letra, acentuación y puntuación. Los paréntesis y signos de exclamación son utilizados en su forma canónica y su uso es relativamente bajo.

Periodistas	Tuits	N.ºPal.
Juan R. Lucas	Nos hemos venido arriba ;) y el dom.7 volvemos a #EoriaDelLibrodeMadrid #HablemosSobreFelicidad @lideditorial/c318 (Imagen)	9
Jesús Maraña	Portada de #infoLibre para soci@s: la Iglesia católica recibió más de 11,5 millones de la Fundación #CajaMadrid (Imagen)	13
Pepa Bueno	Le brillan los ojos a @alejandrosanz cuando habla de sus hijos. #ChesterAlejandro	10
Ana Pastor	Premio doble porque @ObjetivoLaSexta lo comparte con @jordievole y @salvados tv #PremiosACPE (Imagen)	7

Figura 6. Tweets con elementos paralingüísticos (hashtag, arroba e imagen)

En relación a los elementos paralingüísticos, a pesar de que tampoco son elementos imprescindibles para la construcción de los tweets periodísticos, los *hashtags*, arrobas e imágenes son los elementos que más contribuyen a otorgar información adicional y complementaria con un empleo acusado cuando el periodista quiere orientar el contenido en tweets con poca densidad de palabras (9,07%). Por último, los resultados muestran que no hay diferencias significativas conforme a la variable género en el empleo de elementos ortotipográficos y paralingüísticos en el tweet periodístico.

Esta investigación se ha llevado a cabo desde la limitación de la muestra empleada, por lo tanto, sería interesante seguir investigando la construcción del tweet desde estudios longitudinales con muestras mayores y desde la comparación de géneros diferentes para establecer las diferencias y similitudes del lenguaje en *Twitter*.

### 7. Bibliografía

Ahmad, Ali (2010). “Is Twitter a useful tool for journalists?”. *Journal of media practice*, v. 11, n. 2, pp. 145-55. [http://www.tandfonline.com/doi/abs/10.1386/jmpr.11.2.145\\_1?journalCode=rjmp20](http://www.tandfonline.com/doi/abs/10.1386/jmpr.11.2.145_1?journalCode=rjmp20)

Almela, Ramón (2003). “Inconsistencias de la oposición oral/escrito”. *Oralia. Análisis del discurso oral*, v. 6, pp. 63-88.

Anderson, Christopher W. (2012). *Rebuilding the news: Metropolitan journalism in the digital age*. Philadelphia: Temple University Press. ISBN: 978 1439909348

Androustopoulos, Jannis (2011). “From variation to heteroglossia in the study of computer-mediated discourse”. En: Thurlow, Crispin; Mroczek, Kristine (eds.). *Digital discourse:*

- language in the new media*, pp. 277-298. New York: Oxford University Press. ISBN: 978 0199795444  
[https://jannisandroutsopoulos.files.wordpress.com/2012/01/2011\\_from-variation-to-heteroglossia-in-cmda-published.pdf](https://jannisandroutsopoulos.files.wordpress.com/2012/01/2011_from-variation-to-heteroglossia-in-cmda-published.pdf)
- Androutsopoulos, Jannis** (2014). *Mediatization and socio-linguistic change*. Berlin: De Gruyter Mouton. ISBN: 978 3110346831
- Boyd, Danah; Golder, Scott; Lotan, Gilad** (2010). "Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter". En: *Procs of the 43<sup>rd</sup> Hawaii intl conf on system sciences*. Hawaii, HI, IEEE Press, pp. 1-10.  
<http://dx.doi.org/10.1109/HICSS.2010.412>
- Broersma, Marcel; Graham, Todd** (2013). "Twitter as a news source". *Journalism practice*, v. 7, n. 4, pp. 446-464.  
[http://www.rug.nl/research/portal/files/12824342/Broersma\\_and\\_Graham\\_2013.pdf](http://www.rug.nl/research/portal/files/12824342/Broersma_and_Graham_2013.pdf)  
<http://dx.doi.org/10.1080/17512786.2013.802481>
- Carrera-Álvarez, Pilar; Sainz-de-Baranda, Clara; Herrero-Curiel, Eva; Limón-Serrano, Nieves** (2012). "Journalism and social media: How Spanish journalists are using Twitter". *Estudios sobre el mensaje periodístico*, v. 18, n. 1, pp. 31-53.  
[http://dx.doi.org/10.5209/rev\\_ESMP.2012.v18.n1.39353](http://dx.doi.org/10.5209/rev_ESMP.2012.v18.n1.39353)
- Cortés-Rodríguez, Luis** (2012). "Los límites del discurso: condicionantes y realizaciones". *Círculo de lingüística aplicada a la comunicación*, v. 51, pp. 3-49.  
<https://revistas.ucm.es/index.php/CLAC/article/viewFile/40627/38953>
- De-Ramón-Carrión, Manuel** (2013). "Un nuevo lenguaje para los medios periodísticos digitales. La necesidad de adaptarse al continuum informativo. Primeras experiencias en España". *Revista de comunicación de la Seeci*, año XV, v. 32, pp. 127-140.  
<http://dx.doi.org/10.15198/seeci.2013.32.127-140>
- Domingo, David; Le-Cam, Florence** (2014). "Journalism in dispersion". *Digital journalism*, v. 2, n. 3, pp. 310-321.  
<http://dx.doi.org/10.1080/21670811.2014.897832>
- Dresner, Eli; Herring, Susan C.** (2010). "Functions of the non-verbal in CMC: Emoticons and illocutionary force". *Communication theory*, v. 20, n. 3, pp. 249-268.  
<http://dx.doi.org/10.1111/j.1468-2885.2010.01362.x>
- Fletcher, William H.** (2004). "Facilitating the compilation and dissemination of *ad hoc* web corpora". En: Aston, Guy; Bernardini, Silvia; Stewart, Dominic (eds.). *Corpora and language learners*, pp. 273-300. ISBN: 978 9027222886
- Fombona, Javier; Agudo, Susana; Fernández, Alberto** (2013). "Management and teaching experiences in a context of socio-economic crisis in Spain: a new vocational training centre in communication". *Croatian journal of education*. v. 15, n. 2, pp. 11-39.  
<http://hrcak.srce.hr/file/157304>
- Gómez-Camacho, Alejandro** (2014). "La norma disortotográfica en la escritura digital". *Didac*, v. 63, pp. 19-25.  
[http://revistas.iberomx.com/didac/articulo\\_detalle.php?id\\_volumen=17&id\\_articulo=207](http://revistas.iberomx.com/didac/articulo_detalle.php?id_volumen=17&id_articulo=207)
- Gómez-Gómez, Manuel; Danglot-Banck, Cecilia; Vega-Franco, Leopoldo** (2003). "Nonparametric statistical tests synopsis. When are they used?". *Revista mexicana de pediatría*, v. 70, n. 2, pp. 91-99.  
<http://goo.gl/tb3WK7>
- Honeycutt, Courtenay; Herring, Susan C.** (2009). "Beyond microblogging: Conversation and collaboration via Twitter". En: *Procs of the 42<sup>nd</sup> Hawaii intl conf on system sciences*, Hawaii, HI: IEEE, pp. 1-10.  
<http://dx.doi.org/10.1109/HICSS.2009.89>
- Hong, Lichang; Convertino, Gregorio; Chi, Ed** (2011). "Language matters in Twitter: A large scale study". En: *Intl AAAI conf on weblogs and social media*, pp. 518-521.  
<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2856/3250>
- Huffaker, David A.; Calvert, Sandra L.** (2005). "Gender, identity, and language use in teenage blogs". *Journal of computer-mediated communication*, v. 10, n. 2.  
<http://dx.doi.org/10.1111/j.1083-6101.2005.tb00238.x>
- Java, Akshay; Song, Xiodan; Finin, Tim; Tseng, Belle** (2007). "Why we Twitter: Understanding micro-blogging usage and communities". En: *Procs 9<sup>th</sup> WebKDD and 1<sup>st</sup> SNA-KDD 2007 Workshop on web mining and social network analysis*, San Jose, California, USA.  
[http://ebiquity.umbc.edu/\\_file\\_directory\\_/papers/369.pdf](http://ebiquity.umbc.edu/_file_directory_/papers/369.pdf)
- Knaś, Agnieszka** (2009). "Read my voice: Expressing silence and sound in text-messages". *Networking knowledge*, v. 2, n. 1, pp 1-18.  
<http://www.ojs.meccsa.org.uk/index.php/netknow/article/download/45/45>
- Lasorsa, Dominic; Lewis, Seth; Holton, Avery** (2012). "Normalizing Twitter: Journalism practice in an emerging communication space". *Journalism studies*, v. 13, n. 1, pp. 19-36.  
<http://dx.doi.org/10.1080/1461670X.2011.571825>
- Lomborg, Stine** (2011). "Negotiating the Twitter self. On networks of affiliation and relational pressures". En: *International communication association*, Boston, MA, USA, 26-30 May.
- Mancera-Rueda, Ana; Pano-Alamán, Ana** (2013). *El español coloquial en las redes sociales*. Madrid: Arco Libros. ISBN: 978 8476358795
- Markman, Kris M.** (2013). "Conversational coherence in small group chat". En: Herring, Susan; Stein, Dieter; Virtanen, Tuija (eds.). *Pragmatics of computer-mediated communication*, pp. 539-564. ISBN: 978 3110214468
- Martínez-de-Sousa, José** (2004). *Ortografía y ortotipografía del español actual*. Gijón: Trea. ISBN: 978 8497047241
- Menna, Laura** (2012). "Nuevas formas de significación en red: el uso de las #etiquetas en el movimiento 15M". *Estudios de lingüística del español*, v. 34, pp. 1-61.  
<http://www.raco.cat/index.php/Elies/article/view/271561/359189>
- Pano-Alamán, Ana; Mancera-Rueda, Ana** (2014). "La conversación en Twitter: las unidades discursivas y el uso de marcadores interactivos en los intercambios con parlamentarios españoles en esta red social". *Estudios de lingüística*

del español, v. 35, n. 1, pp. 243-268.

<http://infoling.org/elies/35/elies35.1-10.pdf>

**Parodi, Giovanni** (2010). *Lingüística de corpus: de la teoría a la empiria*. Frankfurt: Iberoamericana/Vervuert. ISBN: 978 8484895015

**Randall, Neil** (2002). *Lingo online: A report on the language of the keyboard generation*. Department of English, University of Waterloo; MSN.ca.

<http://www.arts.uwaterloo.ca/~nrandall/LingoOnline-finalreport.pdf>

**Sanmartín-Sáez, Julia** (2007). *El chat. La conversación tecnológica*. Madrid: Arco Libros. ISBN: 978 8476357101

**Thurlow, Crispin; Mroczek, Kristine** (2011). *Digital discourse: Language in the new media*. New York: Oxford University Press. ISBN: 978 0199795437

**Thurlow, Crispin; Poff, Michele** (2011). "Text-messaging". En: Herring, Susan; Stein, Dieter; Virtanen, Tuija (eds.). *Handbook of the pragmatics of CMC*, pp. 1-24. <http://nl.ijs.si/janes/wp-content/uploads/2014/09/thurlowpoff11.pdf>

**Vázquez-Cano, Esteban** (2012). "Mobile learning with Twitter to improve linguistic competence at secondary schools". *The new educational review*, v. 29, n. 3, pp. 134-147.

**Vázquez-Cano, Esteban; Mengual-Andrés, Santiago; Roig-Vila, Rosabel** (2015). "Análisis lexicométrico de la especificidad de la escritura digital del adolescente en *Whatsapp*". *Revista de lingüística teórica y aplicada*, v. 53, n. 1, pp. 83-105. <http://rua.ua.es/dspace/handle/10045/48306>

**Yus, Francisco** (2001). *Ciberpragmática. El uso del lenguaje en internet*. Barcelona: Ariel. ISBN: 978 8434482425

## No descuides mantener actualizados tus datos en el *Directorio EXIT* (*EX*pertos en el *T*ratamiento de la *I*nformación)



Países	nº
España	2.287
Colombia	187
México	126
Argentina	113
Brasil	107
Reino Unido	81
Cuba	72
Perú	69
Chile	68
EUA	61
Italia	46
Portugal	40
Francia	32
Holanda	30
Ecuador	28
Venezuela	21
Bélgica	19
Alemania	17
Uruguay	13
Suiza	12



Puedes incluir enlaces a tus páginas de las redes sociales

### Especialidades más representadas en EXIT

Especialidad	nº
Gestión de la información y del conocimiento	942
Web 2.0 y redes sociales	747
Biblioteca universitaria	706
Recuperación de información y búsquedas	636
Información científico-técnica	577
Biblioteca digital	550
Planificación y gestión	483
Bibl. especializada / Centro de documentación	454
Análisis y diseño de sistemas de información	453
Gestión de contenidos	437
Open access	427
Comunicación	422
Catalogación, clasificación e indexación	401
Bases de datos (contenidos)	388
Educación, formación, alfin	337
Revistas electrónicas	335
Bibliometría y cibermetría	324
Marketing y promoción	317
Documentación en general	316
Arquitectura de la información	311
Servicios de internet en general	297
Información biomédica	290
Diseño de webs	284
Usabilidad, interfaces e interacción	274
Biblioteca pública	270
Software documental	245
Ontologías, metadatos, taxonomías y tesauros	230
Periodismo	229
Multimedia y documentación audiovisual	229
Información humanística	227

<http://directorioexit.info>



# JOUPRENEUR: AN ORIGINAL METHODOLOGY FOR RAISING ENTREPRENEURIAL INTENTIONS AMONG JOURNALISM STUDENTS

*Joupreneur: una metodología docente para incrementar la intención emprendedora en estudiantes de periodismo*

**Manuel Goyanes y Marcello Serra**



**Manuel Goyanes**, PhD, teaches at the *Carlos III University* in Madrid and his main interests are in media management and sociology of organizations. He has written about leadership, news overload and business models. He has attended several international congresses and published in international journals such as *Journalism practice*, *International journal on media management*, and *International journal of communication*.  
<http://orcid.org/0000-0001-6537-9777>

[manuel.goyanes@uc3m.es](mailto:manuel.goyanes@uc3m.es)



**Marcello Serra**, PhD, teaches at the *Carlos III University* in Madrid and his main interests are in semiotics and media theory. Among other subjects, he has written about comics, transparency, social movements, and Diego Maradona as a contemporary idol. He has coordinated several books and, in particular, he is the editor, with Mario Pireddu, of *Mediología. Una disciplina attraverso i suoi classici* (Spanish version: *Mediología. Cultura, tecnología y comunicación*).  
<http://orcid.org/0000-0002-6412-5078>

[marcello.serra@uc3m.es](mailto:marcello.serra@uc3m.es)

*Carlos III University of Madrid*  
C/Madrid, 133. 28903 Getafe (Madrid), Spain

## Abstract

This article describes in detail, and empirically validates a methodology for raising entrepreneurial intentions among journalism students. The research is based on pre-post surveys of four groups of students (88 in total), ages 18-25, from the *Carlos III University of Madrid*, and measures the change in the students' intentions after a period of three months in a six-credit, third-year, journalism class. The findings contribute to the theories of education and have wider implications for the practice of entrepreneurship teaching among journalism students.

## Keywords

Entrepreneurship; Entrepreneurial intention; Journalism; Education; University students.

## Resumen

El presente artículo describe en detalle y valida empíricamente una metodología docente para incrementar la intención emprendedora de estudiantes de periodismo. La investigación se basa en un diseño experimental mediante encuestas pre-post en cuatro grupos de 18-25 estudiantes (88 en total) de la *Universidad Carlos III de Madrid*, y mide el cambio en la intención después de un período de tres meses de clase. Los resultados contribuyen a la bibliografía teórica sobre educación y tienen amplias implicaciones para la práctica de la enseñanza sobre emprendimiento en estudiantes de periodismo.

## Palabras clave

Emprendimiento; Intención emprendedora; Periodismo; Educación; Estudiantes universitarios.

**Goyanes, Manuel; Serra, Marcello** (2016). "Joupreneur: An original methodology for raising the entrepreneurial intention among journalism students". *El profesional de la información*, v. 25, n. 4, pp. 599-605.

<http://dx.doi.org/10.3145/epi.2016.jul.09>

## 1. Introduction

In recent years, fostering entrepreneurship education has become a topic of the highest priority in public policy throughout most industrial countries (Galán-Gamero; Aguado-Guadalupe, 2014). Research typically supports the idea that high levels of entrepreneurship can be achieved through education (Tkachev; Kolvereid, 1999) and, as a result, a wide range of education initiatives related to entrepreneurship have been initiated. In addition, well-educated entrepreneurs have established themselves as important leaders throughout the world.

In the United States, entrepreneurship and entrepreneurship education (EE) are widely recognized to have made tremendous progress (Lüthje; Franke, 2004). Spain, however, has been comparatively slow to recognize these opportunities. Until recently, fostering entrepreneurial values has not been regarded as a primary task in Spanish universities. In this sense, the vast majority of Spanish public universities have lacked specific education on this field, and “entrepreneurship” as an educational discipline has not been recognized by the official educational authorities (ANECA). However, in recent years, fostering entrepreneurial inclinations through education and training has received increasing attention from Spanish public institutions and, even more, from private ones. Since the beginning of the new century and particularly as a result of the bursting of the Spanish housing bubble and the subsequent economic downturn, the public authorities that champion entrepreneurship education as a fundamental contributor for the future economic development are growing louder (Goyanes, 2015).

The important assumption underlying the promotion of entrepreneurship education follows the idea that participating in an entrepreneurship course can increase a student’s intention to become an entrepreneur and equip would-be entrepreneurs with relevant skills (Lorz; Mueller; Volery, 2013). Most studies support this hypothesis, and conclude that entrepreneurship education has a positive impact on entrepreneurial intentions and behaviors (Liao; Gartner, 2007). Nevertheless, this stance does not imply that entrepreneurial education can create the next Spanish generation of Mark Zuckerbergs, Steve Jobs, or Larry Pages. Instead, most of the programs are aimed at teaching students to put theory into practice and to better understand entrepreneurship (Meyer, 2011). In this way, according to Sánchez (2013), students are expected to gain self-confidence and motivation to establish their own business ventures, become proactive, creative, and learn how to work in a team.

On the other hand, the multiplication of entrepreneurship programs has generated a growing interest from fund providers and academics about issues of effectiveness and efficiency in these programs, as well as the identification and diffusion of best practices (Fiet, 2001). In this sense, despite the importance of the political and educational promotion of entrepreneurship programs, to our knowledge, there has been a noticeable lack of research into entrepreneurship programs that have been exclusively designed for students pursuing journalism studies. In addition, very little is known about the specific impact of such programs on students’

intentions and the possibilities of their adoption by other different universities.

In this paper, we empirically examine how a specific entrepreneurship program influenced the entrepreneurial intentions of journalism students from the *Carlos III University of Madrid*. In total, 138 students participated in the study: 88 were in the experimental group and 50 in the control group. We conducted a pre-post-test in order to describe in detail, and empirically validate an entrepreneurship program specifically designed to promote an entrepreneurial career path in journalism.

With these goals in mind, the paper is structured as follows: initially, we present a theoretical framework in which we describe the recent literature on entrepreneurial intention and entrepreneurship education programs; this leads to our research question: Does *Joupreneur* significantly increase the entrepreneurial intention of students pursuing journalism studies? Subsequently, we present the methodology of the research and describe in detail the organization, contents, and evaluation system of *Joupreneur*. We then present the results and, finally, we link the findings with the wider entrepreneurship education literature.

## 2. Theoretical framework

### 2.1. Entrepreneurial intention and entrepreneurship education programs

Despite the wide-ranging debate regarding the theoretical framework for analyzing intentions (Thompson, 2009), the vast majority of empirical models depart from the theory of planned behavior (Ajzen, 1991). In this sense, intention would be a decisive and determinant element in predicting behaviors (Fayolle; DeGeorge, 2006). Any planned behavior is best predicted by observing intentions towards that behavior, not by measuring attitudes, beliefs, personality traits, or demographics (Bagozzi; Baumgartner; Yi, 1989). Therefore, the prediction of future behavior should be obtained through the measurement of this intention which, in our case, is the intention to become an entrepreneur. Based on this logic, entrepreneurial intention could be conceptualized as the self-acknowledged conviction or intent “to set up a new business venture and consciously plan to do so at some point in the future” (Thompson, 2009, p. 676). However, depending on what people learn (in this case on what students learn) entrepreneurial intentions can be enhanced or diminished (Von-Graevenitz; Harhoff; Weber, 2010).

In the context of this study we define entrepreneurship education programs as any pedagogical program or process of education for entrepreneurial attitudes and skills, which involves developing certain personal qualities (Fayolle; Gailly; Lassas-Clerc, 2006). This definition is not exclusively focused on the immediate creation of new business ventures and it covers a wide variety of situations, aims, methods, and teaching approaches. Based on these preliminary observations, we draw, in our research, on Donckels’ (1991) claim that at the university level the principal role of entrepreneurship education programs is to increase student awareness and to highlight the entrepreneurial path as a viable career option. The purpose of this study is to test the impact of a specific

program (*Joupreneur*) on the entrepreneurial intentions of students pursuing journalism studies.

Previous empirical studies have focused their analysis on the general theme of entrepreneurial education and entrepreneurial intention. Results show a widespread consensus regarding the positive effect of entrepreneurial education programs on entrepreneurial intention, but the evidence is still not strong. Some empirical studies do confirm that there is a positive impact of entrepreneurship education courses or programs at universities on perceived attractiveness and perceived feasibility of new venture initiation (Tkachev; Kolvereid, 1999). However, other studies find evidence that the effects are negative (Oosterbeek; Van-Praag; Ijsselstein, 2010). In this sense, there may be methodological reasons why the literature has not yet generated consistent assessments on this topic (Von-Graevenitz; Harhoff; Weber, 2010). One such methodological limitation is the focus on self-selected participants with some existing predisposition towards entrepreneurship, which is likely to bias results in favor of educational intervention (Gorman; Hanlon; King, 1997). But above all, the main limitation is that few studies employ a pre-post design, and even fewer use a control group.

Despite these limitations there is a general consensus in the academic literature about the positive relationship between entrepreneurial education programs and entrepreneurial intention (Souitaris; Zerbinati; Al-Laham, 2007). However, it remains to be seen whether and how those results can be generalized to various settings, such as journalism. Based on this, we proposed the following research question: Does *Joupreneur* significantly increase the entrepreneurial intentions of students pursuing journalism studies?

### 3. Methodology of the study

#### 3.1. About the course

“Organization and Management of Media Enterprises” is a mandatory course that is worth 6 ECTS practical-theoretical credits and is offered in the third year of the BA degree in journalism and bilingual journalism at the *Carlos III University of Madrid*. The course is based on the methodology that have been developed for years by Javier Galán, Guadalupe Aguado, José Fernández-Beaumont and Enrique Bullido. The course theoretically explores basic business concepts that are required in order to design, develop, and manage a media company in a relatively competitive and realistic environment. The course has a practical and didactic approach and each topic includes case studies and exercises that help students to understand the theoretical contents under study. In general, the main objective of the course is to provide students with the basic and fundamental theoretical and conceptual tools to develop new business ventures in a relatively real and competitive context.

#### 3.2. Procedure

Intention towards entrepreneurship was analyzed with four different groups of students that took the mandatory course “Organization and Management of Me-

dia Enterprises” in the first semester of the academic year 2014/2015. Of the four groups, three of them were taught in Spanish, whilst the fourth one was in English. In the English group, the origin of the students was as follows (most of the non-Spaniards were part of the international program): USA ( $n = 1$ ; 4.76%), Portugal ( $n = 2$ ; 9.52%), Australia ( $n = 2$ ; 9.52%), Germany ( $n = 3$ ; 14.28%), England ( $n = 5$ ; 23.80%), and Spain ( $n = 8$ ; 38.09%). The total number of students in all four courses amounted to 120, of which 88 were surveyed (the response rate was 73.33%). The total sample of 88 students included slightly more women ( $n = 61$ ; 69.3%) than men ( $n = 27$ , 30.7%), a gender distribution that aligns with the vast majority of journalism courses in Spanish universities. The average age was 20.79 (max.: 33; min.: 19).

The distribution of the students by group was as follows: a total of 25 students were surveyed in group one, 21 students were surveyed in group two (in the bilingual journalism BA degree), a total of 23 students were surveyed in group three, and 19 in the fourth group. The same students participated in both the pre and post-test. As previously stated, the surveys were conducted at the beginning and end of the first semester of the 2014/2015 academic year in all four groups. Data for the pre-test was collected on the 15<sup>th</sup> of September of 2014 while data for the post-test was collected on the 15<sup>th</sup> of December 2014, that is, three months later and after *Joupreneur*.

With respect to the control group, we chose 50 journalism students who took the same subject in a different public Spanish university. The rationale to this selection was the fact the students had similar demographic characteristics and the subject was imparted in the same degree, but with a different learning methodology and topics covered. The control group was formed by 28 women (56%) and 22 men (44%) with an average age of 22 (max: 27; min: 21).

Different methodological approaches have been used to examine differences between treatment and control groups. However, we use the methodological approach proposed by Souitaris, Zerbinati, and Al-Laham (2007) because it meets a high standard of methodological rigor (Martin; MacNally; Kay, 2012) and it is more suited to our interests. Thus, to test the effect of the program on the changes in intentions, we used one-way anova on the difference scores (total sample) with group membership (experimental versus control) as the independent variable. The “difference scores” method is preferable to split-plot repeated measures anova for pretest–posttest designs with “treatment” and control groups as it gives equivalent results in a simpler and less confusing way (Girden, 1992).

Table 1. Demographic characteristics

	Experimental (N = 88)		Control (N = 50)		Total (N = 138)	
	Nº	Percent	Nº	Percent	Nº	Percent
Sex	88	100	50	100	138	100
Men	27	30.7	22	44	49	35.50
Women	61	69.3	28	56	89	64.49
Age (m)	20.79		22		21.30	

#### 4. *Joupreneur* program

The duration of the course is 14 weeks, with one three-hour class per week. The course design is identical for all the groups and complies with the following pattern. From the beginning, according to their number (usually around 25), the students are randomly divided into three or four groups of 6-8 members and they start working on the idea of a business project in the media industry that will be presented at the end of the term. In the first part of the course (weeks 1-6) classes are divided into two equal parts: during the first half, the lecturer introduces some theoretical concepts and managerial tools, whereas the second part is dedicated to team-work skill-building and students try to apply these notions to their projects, which grow week by week.

The different teachers are free to organize their lectures according to their own personal inclinations, but they all typically cover the following topics: mission, vision and values of the company; industry analysis; product development and innovation process; value chain and competitive advantage; organizational structure and marketing. The contents tackled in the lectures provide an introduction to management for journalists and represent a guide for the realization of the final project, a 5 to 10 pages document, which is organized as follows: First, it must contain a mission statement, a vision statement, and a values statement of the company. The second part is dedicated to an analysis of the industry in which the firm is entering, of its direct competitors, and of their potential response. In the third section of the project, the students must describe the company's value chain and core processes, its position in the value system and then indicate how they think it is possible to achieve a competitive advantage in the chosen industrial sector. After this, the document illustrates the formal organization of the firm and provides a possible salary frame. The successive paragraph deals with the competitive strategy of the company and its middle-term strategy. Then there is a description of the launching process and of the launching timing plan, illustrated by a Gantt chart. The following section of the project envisions a marketing strategy and the last one examines the sources of revenues that are expected to provide profitability. Given that the course is part of a journalism degree, no financial analysis is required. Above all, the final project design attempts to give a taste of what an entrepreneurial process can be and it provides students their first introduction to basic managerial tools.

The second part of the course (weeks 7-11) is characterized by the absence of lectures and every week each group presents a different study on a national media industry. Ideally, the chosen country coincides with the one in which the group is considering as the place to start the final project's company, but this is not mandatory. The first presentation is dedicated to the press industry, then the industries of radio, television, and advertising are covered; lastly, the groups provide a general framework of the chosen national media industry<sup>1</sup>.

Every member must participate in the public presentation, which lasts 25-30 minutes and is accompanied by a multimedia document (no longer than 10 slides) and a two-sheet

pdf document containing subsidiary information and all the sources of the data<sup>2</sup>. At the end of each presentation, the teacher comments on the research and a class discussion follows. Also, the other groups deliver a brief written analysis/consideration and an evaluation of the presentation to the lecturer. These "peer-reviews", that may or may not be taken into account by the teacher, have the purpose of maintaining the attention of the students and motivating a critical view.

Overall, the goal of these five presentations is to develop the ability of studying a media industry from an entrepreneurial point of view, of providing a concise account of its situation, and of seeing the business opportunities within the distinct sectors. This also aims to improve the final projects, which acquire details and background information during these weeks.

The class of week twelve is dedicated to a group tutorship in which each team receives final advice and feedback from the teacher. Then, on the last two days of class, students work in teams to give 30-45 minute public presentations of their final projects.

#### 5. Evaluation system

The final grade is the result of the team's work, which is evaluated during the course of the quarter, and of a final individual exam. Both the exam and the continuous evaluation account for half of the final grade, which is expressed, according to the Spanish system, in a 0-10 scale where 5 is the minimum mark needed to pass. The final exam consists of 10 questions about the topics covered in the lectures and a reference book [*Organización y gestión de la empresa informativa* by **Aguado-Guadalupe et al.** (2008) for the Spanish group and *Strategic management in the media: Theory to practice* by **Lucy Küng** (2008) for the English group].

The final project accounts for 25% of the grade and its mark is up to 250 points. To avoid cases of free-riding, this mark is multiplied by the number of the members in the groups and the students split it and share it according to the perceived individual contribution. So, if the final project grade is 200 and the group has 6 members, the mark will be 1200, and the individual marks may be 200 for everybody if the group agrees to share equally, or may be different in case it decides otherwise. Finally, each of the five presentations of the national media industry account for 5% of the final grade. Their marks are up to 50 and they follow the same sharing rule of the final project's grade.

#### 6. Results

Prior to conducting the experimental analysis, we analyzed both pre and post data. The mean scores of the entrepreneurial intention are very similar in pre data (in ascending order): course 3 has the lowest entrepreneurial intention ( $m = 2.13$ ;  $SD = 1.14$ ), followed by course 1 ( $m = 2.48$ ;  $SD = 1.12$ ), course 2 ( $m = 2.61$ ;  $SD = 1.11$ ), and finally course 4 has the highest entrepreneurial intention ( $m = 2.84$ ;  $SD = 0.95$ ). The mean of the entrepreneurial intention in all courses in pre data is 2.50 ( $SD = 1.10$ ). With regard to the control group the mean scores of the entrepreneurial intention is 2.40 ( $SD = 1.16$ ).

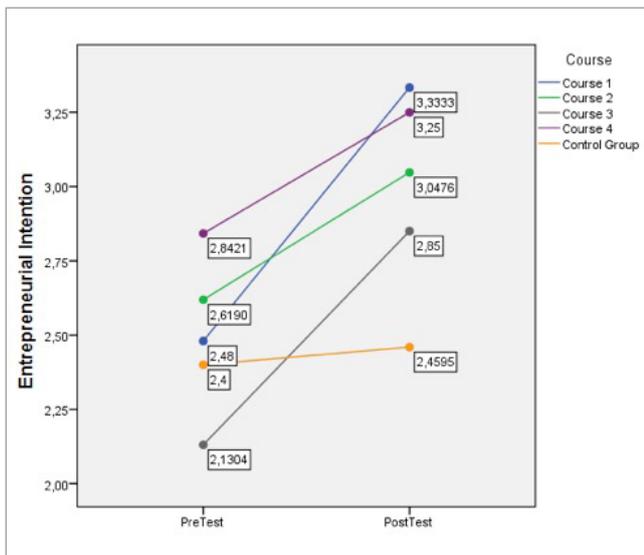


Figure 4. Pre and post-data in the four groups and the control group.

In relation to the post analysis, data is very similar with regard to the parity between means of entrepreneurial intention, although there is a notable increase with respect to pre data (to be in ascending order): course 3 has the lowest mean in entrepreneurial intention in post data ( $m = 2.85$ ;  $SD = 1.13$ ), followed by course 2 ( $m = 3.04$ ;  $SD = 1.02$ ), course 4 ( $m = 3.25$ ;  $SD = 1.01$ ) and finally, course 1 ( $m = 3.33$ ;  $SD = 1.06$ ) has the highest entrepreneurial intention. The mean of the entrepreneurial intention in all four courses in post data is 3.12 ( $SD = 1.05$ ). On the other hand, the mean scores of the control group are very similar to pre data ( $m = 2.45$ ;  $SD = 1.19$ ) suggesting, *a priori*, no significant differences.

As we previously noted, the data of the entrepreneurial intention scores in pre data was 2.50 ( $SD = 1.10$ ) while in post data was 3.12 ( $SD = 1.05$ ). To verify if our methodology (*Joupreneur*) significantly increases the entrepreneurial intention among journalism students we conduct a one-way analysis of variance (anova) for the entire sample. Our results indicate that there are significant differences in relation to the pre/post scores and the level of entrepreneurial intention ( $p < 0.05$ ). Furthermore, there is no significant difference between the level of entrepreneurial intention between pre and post data in the control group. Therefore, we conclude that *Joupreneur* increases the overall entrepreneurial intention among journalism students.

## 7. Discussion

The present article is one of the first studies to describe in detail, and empirically validate a methodology for raising the entrepreneurial intention among journalism students. Based on pre-post surveys of four different groups of students (in total), ages 18 to 25, from the *Carlos III University of Madrid*, this study measured the change in students' intentions after a period of three months of *Joupreneur*, that is, the entrepreneurship program implemented.

First, this study shows a generally low entrepreneurial intention among the students pursuing journalism studies in

Spain. After *Joupreneur*, the analysis of variance (anova) suggests a significant increase in intention in relation to pre-data, even if it still remains low. If intention is considered to be one of the most important predictors of future behavior (Ajzen, 1991), self-employment as a career path in the journalism sector remains as a subordinate labor alternative. Therefore, being a salaried employee continues to be deemed as one of the most important and desirable labor preferences for future professionals.

There are several different reasons that might explain this lack of entrepreneurial intention. Between them, we can highlight the lack of educational support for the students pursuing journalism studies in developing new business ventures, the current Spanish socioeconomic context, and the professional culture of this particular sector (their habits, experiences, and lifestyles). In particular, we consider that the most important factor in determining low entrepreneurial intention is the security provided by the sense of belonging to a workforce, that is, the vision of journalism as an occupation that implies working as an employee in a traditional media company.

In Spain, journalism studies' curricula promote a system of internships in news organizations that start from the early years of the bachelor's degree. Beyond its utility for professional development —something that doesn't concern the scope of this article— it could be suggested that this educational model acts as an inhibitor of future entrepreneurial behaviors. Taking into account the current situation of the sector, this strategy might not be the best for giving the students an opportunity to find a job in the journalism industry. On the contrary, in these days of economic instability, where media companies tend to implement radical austerity measures fundamentally through layoffs and wage cuts, the promotion of entrepreneurship as a preferred career path for future professionals of the journalism sector can be seen as a possible solution (Goyanes, 2014; Manfredi-Sánchez; Rojas-Torrijos; Herranz-de-la-Casa, 2015).

The current Spanish situation of the news industry leaves no room for doubt. Between 2008 and 2013, 284 news organizations were closed and the unemployment rates have increased by 132%. In 2013 alone, 4,434 reporters were dismissed and 73 news organizations disappeared. Nevertheless, in 2012, 2,909 new journalists graduated. Given this data and according to our research, we would suggest the implementation by academic institutions of complementary programs aimed at changing the students' traditional expectations. We believe that a stronger focus on entrepreneurship would provide students with an alternative mental framework while showing a different approach to their prospective careers. To achieve this goal, the universities may, for example, introduce some courses oriented to the development of entrepreneurial attitudes from the beginning of the degree, encourage extracurricular activities focused on entrepreneurship, or create master's degrees or curricula with a strong emphasis on self-employment. As a result, we could witness the creation of new environments and professional fields where the skills acquired during the journalism degree could flourish.

Moreover, our research shows how increases in entrepreneurial intentions may be caused by the implementation of simple managerial and entrepreneurial tools and concepts. This is to say, to increase the entrepreneurial intention of journalism students it is sufficient enough to merely create awareness that being an employee in a traditional media company is not the only career option available.

## Notes

1. For a better understanding of the details, we briefly describe here the structure of the presentation:

1) The press industry: evolution of the industry; industry's volume (readers and advertisement investments); general audience's evolution (last 5-10 years minimum); cost structure; revenues structure; promotion / marketing; industry analysis (Michael Porter's 5 competitive forces); impact of digital media; segmentation and regionalization; business opportunities.

2) The radio industry: regulation of the industry; volume of the industry (listeners and advertising investments); evolution of the audiences; timeline of daily consumption (who watches what at what time); industry analysis; segmentation of the contents and the markets; radio station groups; digital radio; business opportunities.

3) The television industry: regulation of the industry (including advertising regulation); volume of the industry (audiences and advertising investments); evolution of the audiences; timeline of daily consumption; segmentation of the contents and the markets; television groups; industry analysis; new ways of watching television; business opportunities.

4) The advertising industry: regulation; evolution of the investments in all media industries (general overview), in each traditional media industry (radio, television, press), and by advertisers; structure of the industry (number and size of companies); factors in media selection; new advertising trends in traditional media companies and new media companies.

5) General framework: presentation that provides a description of the present situation of the national media industry mostly using data gathered in previous weeks. The primary aim is to highlight the most important features of the industry and, secondly, present the business opportunities within the industry.

2. These restrictions are not applied in the case of the last presentation.

## 8. References

- Aguado-Guadalupe, Guadalupe et al.** (2008). *Organización y gestión de la empresa informativa*. Madrid: Síntesis. ISBN: 9788497566025
- Ajzen, Icek** (1991). "The theory of planned behavior". *Organizational behavior and decision processes*, v. 50, n. 2, pp. 179-211. [http://dx.doi.org/10.1016/0749-5978\(91\)90020-T](http://dx.doi.org/10.1016/0749-5978(91)90020-T)
- Bagozzi, Richard; Baumgartner, Hans; Yi, Youjae** (1992). "State versus action orientation and the theory of reasoned action: an application to coupon usage". *Journal of consumer research*, v. 18, n. 4, pp. 505-518. <http://dx.doi.org/10.1086/209277>
- Donckels, Rik** (1991). "Education and entrepreneurship experiences from secondary and university education in Belgium". *Journal of small business & entrepreneurship*, v. 9, n. 1, pp. 35-42. <http://dx.doi.org/10.1080/08276331.1991.10600389>
- Fayolle, Alain; Degeorge, Jean-Michel** (2006). "Attitudes, intentions, and behaviour: New approaches to evaluating entrepreneurship education". En: Fayolle, Alain; Klandt, Heinz. *International entrepreneurship education. Issues and newness*. Cheltenham, UK: Edward Elgar Publishing, pp. 74-89. ISBN: 978 184542 179 3 <http://goo.gl/GD4xWr>
- Fayolle, Alain; Gailly, Benoit; Lassas-Clerc, Narjisse** (2006). "Assessing the impact of entrepreneurship education programmes: a new methodology". *Journal of European industrial training*, v. 30, n. 9, pp. 701-720. <http://dx.doi.org/10.1108/03090590610715022>
- Fiet, James** (2001). "The theoretical side of teaching entrepreneurship". *Journal of business venturing*, v. 16, n. 1, pp. 1-24. [http://dx.doi.org/10.1016/S0883-9026\(99\)00041-5](http://dx.doi.org/10.1016/S0883-9026(99)00041-5)
- Galán-Gamero, Javier; Aguado-Guadalupe, Guadalupe** (2014). *Periodismo emprendedor: planificación y desarrollo de empresas informativas*. Madrid: McGraw-Hill. ISBN: 9788448190392
- Girden, Ellen R.** (1992). *Anova: Repeated measures*. Thousand Oaks, CA: Sage. ISBN: 978 0803942578
- Gorman, Gary; Hanlon, Dennis; King, Wayne** (1997). "Some research perspectives on entrepreneurship education, enterprise education and education for small business management: a ten-year literature review". *International small business journal*, v. 15, n. 3, pp. 56-77. <http://dx.doi.org/10.1177/0266242697153004>
- Goyanes, Manuel** (2015). "Apoyo estructural en la intención emprendedora de estudiantes de periodismo y comunicación audiovisual en España". *El profesional de la información*, v. 24, n. 1, pp. 55-61. <http://dx.doi.org/10.3145/epi.2015.ene.07>
- Küng, Lucy** (2008). *Strategic management in the media: Theory to practice*. London: Sage. ISBN: 1412903130
- Liao, Jianwen; Gartner, William B.** (2007). "The influence of pre-venture planning on new venture creation". *Journal of small business strategy*, v. 18, n. 2, pp. 1-21. <https://www.questia.com/read/1P3-1462628221/the-influence-of-pre-venture-planning-on-new-venture>
- Lorz, Michael; Mueller, Susan; Volery, Thierry** (2013). "Entrepreneurship education: a systematic review of the methods in impact studies". *Journal of enterprising culture*, v. 21, n. 2, pp. 123-151. <http://dx.doi.org/10.1142/S0218495813500064>
- Lüthje, Christian; Franke, Nikolaus** (2004). "Entrepreneur-

ial intentions of business students: A benchmarking study". *International journal of innovation and technology management*, v. 1, n. 3, pp. 269-288.  
[http://epub.wu.ac.at/3097/1/entrepreneurial\\_intentions\\_benchmark.pdf](http://epub.wu.ac.at/3097/1/entrepreneurial_intentions_benchmark.pdf)

**Manfredi-Sánchez, Juan-Luis; Rojas-Torrijos, José-Luis; Herranz-de-la-Casa, José-Luis** (2015). "Innovación en el periodismo emprendedor deportivo. Modelo de negocio y narrativas". *El profesional de la información*, v. 24, n. 3, pp. 265-273.  
<http://dx.doi.org/10.3145/epi.2015.may.06>

**Martin, Bruce C.; McNally, Jeffrey J.; Kay, Michael J.** (2012). "Examining the formation of human capital in entrepreneurship: A meta-analysis of entrepreneurship education outcomes". *Journal of business venturing*, v. 28, n. 2, pp. 211-224.  
<http://dx.doi.org/10.1016/j.jbusvent.2012.03.002>

**Meyer, G. Dale** (2011). "The reinvention of academic entrepreneurship". *Journal of small business management*, Special issue: Global research and policy implications for today's SMEs, v. 49, n. 1, pp. 1-8.  
<http://dx.doi.org/10.1111/j.1540-627X.2010.00311.x>

**Oosterbeek, Hessel; Van-Praag, Mirjam; Ijsselstein, Auke** (2010). "The impact of entrepreneurship education on entrepreneurship skills and motivation". *European economic review*, v. 54, n. 3, pp. 442-454.  
<http://dx.doi.org/10.1016/j.eurocorev.2009.08.002>

**Sánchez, José C.** (2013). "The impact of an entrepreneur-

ship education program on entrepreneurial competencies and intention". *Journal of small business management*, v. 51, n. 3, pp. 447-465.  
<http://dx.doi.org/10.1111/jsbm.12025>

**Souitaris, Vangelis; Zerbinati, Stefania; Al-Laham, Andreas** (2007). "Do entrepreneurship programmes raise entrepreneurial intention of science and engineering students? The effect of learning, inspiration and resources". *Journal of business venturing*, v. 22, n. 4, pp. 566-591.  
<http://dx.doi.org/10.1016/j.jbusvent.2006.05.002>

**Thompson, Edmund R.** (2009). "Individual entrepreneurial intent: Construct clarification and development of an internationally reliable metric". *Entrepreneurship theory and practice*, v. 33, n. 3, pp. 669-694.  
<http://dx.doi.org/10.1111/j.1540-6520.2009.00321.x>

**Tkachev, Alexei; Kolvereid, Lars** (1999). "Self-employment intentions among Russian students". *Entrepreneurship & regional development*, v. 11, n. 3, pp. 269-280.  
<http://dx.doi.org/10.1080/089856299283209>

**Von-Graevenitz, Georg; Harhoff, Dietmar; Weber, Richard** (2010). "The effects of entrepreneurship education". *Journal of economic behavior & organization*, v. 76, n. 1, pp. 90-112.  
<http://dx.doi.org/10.1016/j.jebo.2010.02.015>

**Zahra, Shaker; Wright, Mike; Abdelgawad, Sondas** (2014). "Contextualization and the advancement of entrepreneurship research". *International small business journal*, v. 32, n. 5, pp. 479-500.  
<http://dx.doi.org/10.1177/0266242613519807>

## 7ª Conferencia internacional sobre revistas de ciencias sociales y humanidades Cuenca, 4-5 mayo 2017



# SITUACIÓN DE LOS CENTROS DE DOCUMENTACIÓN EN LOS MEDIOS ESCRITOS ANDALUCES

## Situation of documentation services in the Andalusian written media

Natalia Meléndez-Malavé y Rocío Hirschfeld-Suárez



**Natalia Meléndez-Malavé**, doctora en periodismo (2005) por la *Universidad de Málaga (UMA)*, es profesora en el *Departamento de Periodismo* de la *Facultad de Ciencias de la Comunicación* de la *UMA*, donde imparte la asignatura *Documentación informativa* en el grado de periodismo. Es integrante de varios grupos de investigación nacionales y de proyectos de innovación educativa en la enseñanza universitaria. Ha publicado diversas aportaciones en obras, actas de congresos y artículos en revistas especializadas con material hemerográfico como fuente para la investigación, así como sobre la adaptación de metodologías didácticas innovadoras para la impartición de la asignatura *Documentación informativa*.

<http://orcid.org/0000-0003-0835-3997>

*Universidad de Málaga, Facultad de Ciencias de la Comunicación, Departamento de Periodismo  
Campus Teatinos, s/n, 29071 Málaga, España  
natalia@uma.es*



**Rocío Hirschfeld-Suárez** es estudiante de periodismo en la *Facultad de Ciencias de la Comunicación* de la *Universidad de Málaga*, en la cual además ha realizado diversos cursos de formación complementaria en la rama del periodismo político o referentes a temas como el lenguaje de los medios en la formación de la imagen femenina. Efectuó sus prácticas laborales en la cadena local *Onda Azul Málaga RTV*, realizando tareas de redacción, montaje y locución.

<http://orcid.org/0000-0003-0602-1761>

*rociohirschfeld@hotmail.com*

### Resumen

Este trabajo analiza la situación actual de los departamentos de documentación en 14 de las principales cabeceras locales andaluzas. El objetivo es actualizar los datos de un estudio similar titulado "Situación de los servicios de documentación de la prensa diaria de Andalucía" (Aquesolo-Vegas, 1996) que realizaba una radiografía de estos departamentos en siete diarios. Este trabajo recoge, adapta y amplía la metodología de aquella investigación con el fin de conocer cómo ha evolucionado desde entonces el panorama de estos centros en los periódicos andaluces más representativos. Con ese fin se ha realizado un estudio de campo basado en cuestionarios, enviados a 16 diarios andaluces, dos por provincia, de los cuales 14 fueron contestados. Se analizan aspectos como las tareas, el personal, las instalaciones y el equipamiento. Los resultados reflejan, a la par que una reconversión tecnológica, la reducción e incluso la desaparición de estos servicios, en una etapa lastrada por la crisis económica y caracterizada por la búsqueda autónoma de documentos por parte de los redactores.

### Palabras clave

Servicios de documentación; Documentación periodística; Periodismo; Prensa; Diarios; Medios; Andalucía; Crisis de la prensa escrita.

### Abstract

This paper analyzes the situation of the archive departments in 14 of the Andalusian local newspapers. The aim is to update a similar study entitled "Situation of archive services in the daily press of Andalusia" (Aquesolo-Vegas, 1996) in which an analysis of these departments was performed for seven newspapers. This work studies, adapts, and extends the methodology of the above cited study to know how the situation has evolved since 1996. To that end, this field study is based on questionnaires sent to 16 Andalusian newspapers, two per province; 14 of which were completed. Aspects such as tasks, staff, facilities, and equipment are analyzed. The results show a technological restructuring, and also a reduction, or even the disappearance, of these services, due to the economic downturn. Therefore, journalists must now search for documents by themselves.

Artículo recibido el 28-01-2016  
Aceptación definitiva: 04-04-2016

## Keywords

Document management; Archive department; News librarianship; Journalism; Press; Newspapers; Media; Andalusia; Crisis of the press.

**Meléndez-Malavé, Natalia; Hirschfeld-Suárez, Rocío** (2016). "Situación de los centros de documentación en los medios escritos andaluces". *El profesional de la información*, v. 25, n. 4, pp. 606-615.

<http://dx.doi.org/10.3145/epi.2016.jul.10>

## 1. Introducción

La documentación forma parte de cualquier proceso de conocimiento y, por ende, de todo proceso informativo (Galdón-López, 2002, p. 67). La profesión periodística, debido a la naturaleza de su labor, es una de las que más necesita "documentarse para completar y verificar las noticias recientes que se van a publicar" (Torres-Rodríguez, 2012, p. 29). Es decir, una información fundamentada que permita un conocimiento crítico del pasado, y que ayude al público a comprender las causas y consecuencias de los hechos que acontecen.

Con ese fin, desde sus orígenes entre finales del siglo XIX y principios del XX, se fueron configurando en los medios de comunicación los centros de documentación como el perfecto intermediario entre la información y el periodista al "ser los recursos más capacitados para buscar y obtener información" (Rubio-Lacoba, 2005, p. 156). Paralelamente, se va desarrollando la Documentación Informativa como disciplina, que hoy conforma una materia indiscutible en los planes de estudio, entre otras, de las titulaciones de Periodismo.

En España, hasta mediados de los 70, la mayoría de los medios de comunicación escritos españoles "carecían incluso de terminales para el procesamiento de textos. La automatización de los periódicos se centró en la automatización de los procesos de composición e impresión". Aunque, cabe reseñar que, en la década de los 80, ya existían proyectos serios para la informatización de archivos (Moreiro-González, 2000, p. 60). Como afirman Marcos-Rocio y Edo (2015, p. 391):

"La transformación de los medios de comunicación se inició con la implantación de los ordenadores en las redacciones. Allá por los años ochenta del siglo pasado se rompió una estructura de hacer y servir la información que se conservaba desde los tiempos de la imprenta".

Así, a partir de la década de los 90, el panorama de la actividad documental informativa española experimenta grandes cambios. No sólo porque en ese periodo comienzan a proliferar los manuales sobre Documentación Informativa (González-Quesada, 2003, p. 77), sino, además, a causa de que la ineludible informatización, en mayor o menor grado, de los departamentos de documentación se convierte en un hecho. La digitalización de los archivos traerá grandes mejoras para los documentalistas, pero pese a esta circunstancia, es un número muy elevado el de periodistas que optan por buscar la información que requieren para la elaboración de sus trabajos en el ciberespacio antes que en los departamentos de documentación. En 2009, un es-

tudio de Micó-Sanz; Masip y García-Avilés (2009, p. 287) en el que fueron entrevistados más de 70 profesionales de la información y se les realizó una encuesta a los responsables de los centros de documentación de las televisiones y los periódicos más relevantes del país, en ese momento, reflejaba que el empleo de buscadores como *Google* y distintas hemerotecas accesibles por la red se ha extendido entre los profesionales de la información por encima del uso de la base de datos del propio medio, colocándose como primeras fuentes de consulta.

De este modo, llegamos al contexto actual de descenso del peso de los servicios de documentación en los medios, debido en parte a que los usuarios han ido adquiriendo mayores competencias en la búsqueda autónoma de documentación, lo que no debe ser obstáculo, sino, al contrario, un impulso a los profesionales para contrarrestar el escenario que describen algunos autores:

"Esta situación de autonomía en la búsqueda y recuperación de información en sistemas integrados puede dar lugar a un menor contacto y diálogo entre usuarios y documentalistas, lo que conlleva un distanciamiento entre el análisis que realizan los documentalistas y las necesidades de información de los usuarios" (Estrada-Nora González; Patallo-Fernández; Pastor-Blanco, 2009, p. 331).

Por su parte, Orero y Cebrián-Enrique concluyen que

"expertos y documentalistas consideran que las fuentes digitales han mejorado la calidad del servicio de las unidades de información por el crecimiento exponencial de los recursos y la rapidez en la respuesta de las consultas. Sin embargo, se confirma que el uso de internet en las redacciones ha provocado también una pérdida de control de los documentalistas sobre las búsquedas de los usuarios" (Orero; Cebrián-Enrique, 2014, p. 615).

Paralelamente, como explica Javier Guallar:

"la crisis en la que están inmersos los medios de comunicación en general y en especial la prensa escrita en la actualidad, está diezmando el personal de las redacciones, afectando de lleno a documentalistas y a centros de documentación" (Guallar, 2011b, p. 54).

Así, la situación económica de los últimos años –sumada al surgimiento de nuevas plataformas– ha afectado gravemente a las empresas informativas y, si bien existen numerosas noticias que hablan del número de medios de comunicación cerrados o de periodistas despedidos, son mucho menos comunes los textos que hablan de las consecuencias sufridas en los centros documentales, quizá porque en los medios

empiezan a percibirlos como obsoletos con la llegada de las nuevas tecnologías. Como bien explica **Rubio-Lacoba** (2010, p. 645):

“La crisis está acorralando a los servicios de documentación, que ven como sus plantillas se esquilman, sus recursos se congelan y su razón de ser es cuestionada (...) Digitalizando el archivo y contando con internet como oráculo documental, ¿para qué seguir gastando en un departamento cuyas tareas han sido absorbidas por informáticos y periodistas?”.

Esta situación repercute directamente en la credibilidad y la calidad de los contenidos informativos ya que si el medio carece de departamento de documentación, o sus informaciones carecerán de datos que enriquezcan el futuro texto, o éstos no serán fiables. De cualquier manera, el producto final se verá afectado, de modo que será de menor calidad que la proporcionada por medios que sí posean centros de documentación, estos últimos, conseguirán a su vez niveles más altos de credibilidad (**Codina**, 2000, p. 49).

Ante la constatación de esta realidad, consideramos de interés revisar la situación actual de los servicios de documentación en prensa. Se juzgó más operativo ceñirnos a un espacio geográfico manejable y no cabe duda que el ámbito autonómico constituía un marco del máximo interés toda vez que, como afirman **Ros-Martín y Rodero-Susiac** (2009, p. 316), en España las empresas editoras de medios de comunicación regionales no han dedicado demasiados recursos a la gestión de la documentación tanto gráfica como textual. Así pues, se optó por actualizar el trabajo de **Aquesolo-Vegas** (1996) sobre los medios escritos en Andalucía, en el que se propuso estudiar la situación de los servicios de documentación de los principales periódicos andaluces en ese momento.

## 2. Objetivos y método

El presente trabajo tiene como objetivo conocer el estado actual de los centros documentales en los principales diarios andaluces, averiguando si cuentan con servicios de documentación o si fueron suprimidos desde el estudio anterior que nos sirve de referencia (**Aquesolo-Vegas**, 1996). Somos conscientes y queremos hacer explícita la limitación de este trabajo al comparar dos momentos separados dos décadas en el tiempo –con la enorme revolución que se ha experimentado en este terreno- pero que suponen la única renovación disponible de datos de este tipo, pues no contamos con estudios de estas características realizados en un periodo más reciente. En todo caso, creemos que los resultados arrojan luz sobre algunas cuestiones descriptivas que pueden ser de interés académico, docente y profesional al contar con datos actualizados que describen la realidad del presente partiendo de atrás, de donde venimos.

En concreto, entre los propósitos establecidos se incluye confirmar si el personal en los departamentos de documentación se ha visto disminuido, además de llevar a cabo una comprobación del progreso en la preparación de dicho personal. Asimismo, se pretende explorar la evolución de estos departamentos en cuanto al espacio físico que se les destina y, especialmente su situación en lo referente al equipamien-

to con el que cuentan. Por último, pretendemos observar los cambios experimentados por estos centros con respecto a los servicios que ofrecen y la relación que mantienen con el equipo de redacción.

Dichos objetivos están guiados por algunas premisas de partida que intentamos corroborar, como la idea de que el departamento de documentación va a ser una de las parcelas perjudicadas en las empresas informativas, suprimiendo el servicio o mermando su personal y sus recursos en gran parte, al verse afectadas por recortes en sus presupuestos debido a la situación de crisis experimentada tanto en el sector de la prensa, como a nivel general desde aproximadamente 2008.

A la par –o incluso a pesar de- esta última proposición, nos planteamos a su vez constatar si, debido al arraigo de las TIC, los servicios han visto cómo sus infraestructuras experimentaban una mejora generalizada respecto a su equipamiento tecnológico.

El método de investigación seleccionado para el estudio del estado actual de los centros de documentación de los principales diarios andaluces ha sido la encuesta mediante cuestionario. Otros estudios precedentes –aunque no dedicados al estudio de los servicios documentales en prensa regional, cuya carencia queremos constatar a pesar del indudable interés que tendrían- han demostrado la eficacia del uso de encuestas a profesionales para obtener datos que nos ayuden en esta tarea descriptiva. Nos referimos por ejemplo a los ya citados de **Micó-Sanz; Masip y García-Avilés** (2009) y el estudio de **Marcos-Recio y Edo** (2015) sobre el uso de la documentación por parte de los periodistas españoles, así como la eficacia del funcionamiento de sus centros de documentación, realizado mediante encuesta aplicada a través de la *Asociación de la Prensa de Madrid* a un total de 246 profesionales; en aspectos más concretos, pero relacionados con el anterior estudio de carácter más integral, encontramos los trabajos de **Guallar y Cornet** (2004 y 2005) que cursaron encuestas, en este caso más similares a la de nuestro trabajo, a los departamentos de documentación de los diez diarios españoles de información general de más audiencia, tomando como referencia el *Estudio general de medios (EGM)*, para indagar sobre la utilización de bases de datos comerciales y sobre la utilización de recursos web gratuitos, respectivamente; más recientemente, **Orero y Cebrián-Enrique** (2014) realizaron una encuesta a 81 documentalistas de medios de comunicación generalistas españoles sobre el uso de fuentes de información en centros de documentación, completada además mediante un estudio Delphi con ocho expertos en Documentación informativa; asimismo, en un campo más específico como el de la documentación fotográfica en prensa se han efectuado también investigaciones basadas en información suministrada por los responsables de dichos servicios (**Sánchez-Vigil; Marcos-Recio; Olivera-Zaldua**, 2009 o, más recientemente, **Guallar**, 2011a).

Dado que nuestro propósito se basaba en actualizar los datos de la década de los 90, la base principal fue la encuesta de **Aquesolo-Vegas** (1996), combinada con la configuración de un corpus más acorde con los trabajos arriba referidos de

Tabla 1. Datos de difusión diaria según OJD de los principales diarios andaluces

Periódico	Difusión (enero de 2015)
<i>Ideal</i>	20.520
<i>Sur</i>	17.547
<i>Diario de Cádiz</i>	16.414
<i>Diario de Sevilla</i>	15.241
<i>Córdoba</i>	8.105
<i>La voz de Almería</i>	5.211
<i>Huelva información</i>	4.803
<i>Jaén</i>	4.022
<i>La opinión de Málaga</i>	2.690
<i>Diario de Jerez</i>	4.465
<i>Europa Sur</i>	2.366
<i>Granada hoy</i>	2.320
<i>Diario de Almería</i>	1.914
<i>El día de Córdoba</i>	1.101
Total	106.719

Guallar y Cornet y enviada además a través de los medios online que actualmente están a nuestro alcance. Así pues, se diseñó y envió una encuesta con preguntas acerca del personal adscrito al servicio y su titulación, el horario y tipo de usuario, la ubicación del local y equipamiento, fondos disponibles o nivel de informatización y equipos y tipo de información que facilitan.

Frente al corpus tomado por Aquesolo que estaba compuesto en un primer momento por 11 diarios, -*La voz de Almería*, *El diario de Cádiz*, *Córdoba*, *Ideal* (Granada), *Huelva información*, *Jaén*, *Sur* (Málaga), *Diario 16 Málaga*, *ABC Sevilla*, *Diario 16 Andalucía* (Sevilla) y *El correo de Andalucía* (Sevilla)-, de los que finalmente sólo pudo contar con la colaboración de siete, la presente investigación ha aumentado el número de rotativos a los que someter a examen. Debido a evidentes cambios en el panorama mediático andaluz no se pudo contar con las mismas cabeceras, pero se seleccionaron dos periódicos por provincia en función de sus datos de difusión (tabla 1).

Posteriormente, se tomaron algunas determinaciones para hacer el material más operativo: se incluyeron los títulos de *Viva Jaén*, *Viva Huelva* y *La voz* (Cádiz) al no haber segundos periódicos en la tabla para estas provincias. Se excluyó a *Europa Sur* y al *Diario de Jerez*, a pesar de ser importantes en difusión, porque elegimos solo capitales de provincia. Además, añadimos a *ABC Sevilla* por la importancia de este periódico, pero finalmente no contestaron la encuesta, al igual que *Diario Jaén*, lo que redujo la muestra definitiva de 16 a 14 rotativos:

- *Diario de Sevilla*  
<http://www.diariodesevilla.es>
- *Sur*  
<http://www.diariosur.es>
- *La opinión de Málaga*  
<http://www.laopiniondemalaga.es>

- *Diario de Cádiz*  
<http://www.diariodecadiz.es>
- *La voz*  
<http://http://www.lavozdigital.es>
- *Ideal*  
<http://www.ideal.es>
- *Granada hoy*  
<http://www.gradahoy.com>
- *Diario Córdoba*  
<http://www.diariocordoba.com>
- *El día de Córdoba*  
<http://www.eldiadecordoba.es>
- *La voz de Almería*  
<http://www.lavozdealmeria.es>
- *Diario de Almería*  
<http://www.elalmeria.es>
- *Viva Jaén*  
<http://andaluciainformacion.es/jaen>
- *Huelva información*  
<http://www.huelvainformacion.es>
- *Viva Huelva*  
<http://andaluciainformacion.es/huelva>

A continuación, se procedió a localizar a los responsables de documentación de los diarios seleccionados. En este proceso aparecieron ciertos obstáculos: parte de los encuestados, debido a su limitado tiempo, no se ofrecía proclive a colaborar, pese a tratarse de preguntas muy breves y concretas expuestas en un sencillo cuestionario de *Google Drive*. En concreto, la herramienta adoptada a la hora de conocer la situación actual de estos departamentos consistió en un cuestionario idéntico diseñado expresamente para este trabajo, que utilizaba de base el empleado en el estudio de Aquesolo, pero adaptado al contexto actual que constaba de las siguientes preguntas:

- Personal adscrito al servicio y titulación: ¿Cuántas personas se encargan de este departamento? ¿Qué tipo de formación tiene el personal adscrito?
- Horario: ¿Qué horario tienen los trabajadores de este departamento?
- Ubicación e instalaciones: ¿Se encuentra el servicio de documentación próximo a la redacción? ¿Dónde está ubicado concretamente? ¿Qué tamaño posee?
- Equipamiento: ¿El equipamiento está informatizado? ¿Qué software emplea? ¿Sobre qué sistema operativo trabaja? ¿Qué hardware emplea? (describa sus características y número de aparatos) ¿Qué tipo de clasificación utiliza el departamento de documentación?
- Información que suministran: ¿Qué tipo de información es la más requerida?

Por último, recibidas las respuestas a lo largo del tercer trimestre de 2014, se extrajeron los resultados y se compararon los datos obtenidos en el trabajo publicado previamente por José Antonio Aquesolo con los derivados de nuestro análisis de las cabeceras andaluzas seleccionadas.

### 3. Resultados

Uno de los grandes obstáculos en ambos estudios ha sido la comunicación con los diarios seleccionados en la muestra. Podemos decir que se repitieron los problemas expuestos en el estudio de Aquesolo, que no pudo contar con las respuestas de algunos de los periódicos escogidos. Tras intentar contactar en reiteradas ocasiones con los responsables de los departamentos seleccionados (en 1996 por teléfono y fax; en esta última muestra se intenta la comunicación mediante teléfono y correo electrónico), quedó constancia desde un inicio de que no todos los medios se mostrarían igual de accesibles. Así pues, de los 16 diarios a los que se intenta acceder en un principio, pese a los múltiples esfuerzos por lograr la comunicación, fue imposible contactar con el departamento de Documentación del *Diario Jaén (Diario Jaén SA)* y con el de *ABC Sevilla (Vocento)* con el fin de que rellenasen el formulario. En el caso de Aquesolo, los periódicos que, a pesar de prometer una respuesta inmediata, no contestaron fueron *Sur* de Málaga, *ABC* de Sevilla y los dos pertenecientes al *Grupo 16*.

Otra cuestión destacable es el hecho de que, aunque fueron 16 los diarios escogidos en un principio para el análisis, hay que tener en cuenta que son seis los grupos editoriales a los que pertenecen: *El diario de Sevilla*, el *Diario de Cádiz*, *Granada hoy*, *El día de Córdoba*, el *Diario de Almería* y *Huelva información* pertenecen al *Grupo Joly*. *ABC de Sevilla*, *La voz*, *Sur* e *Ideal* al *Grupo Vocento*. *La Opinión de Málaga*, a *Prensa Ibérica*. El *Grupo Zeta* posee el *Diario Córdoba*. *Prisa* controla *La voz de Almería*. Por último, *Viva Jaén* y *Viva Huelva* forman parte del grupo editorial *Publicaciones del Sur SA*. Es importante hacer alusión a esta situación, toda vez que uno de los primeros resultados es el hecho de que en algunos casos ha habido una centralización del departamento de documentación, como veremos más adelante.

El proceso de recogida de datos y la interpretación de las encuestas recibidas aporta detalles, quizá anecdóticos, pero muy significativos sobre la realidad de los servicios documentales en las empresas informativas. En una de las tomas de contacto se nos llegó a preguntar “qué entendíamos por un centro de documentación”, por lo que no es raro que no tuvieran claro si disponían o no de tal servicio, cuando ni tan siquiera conocían las características de un departamento de este tipo. En otro caso, nos encontramos con que el diario *Córdoba* al rellenar el cuestionario afirma que posee un servicio de documentación pero a la par señala que no hay ninguna persona que se encargue de gestionarlo.

La información que los diarios suministraron en la encuesta en ocasiones fue incompleta o difusa. Como ejemplo, *Ideal* respondió a la pregunta de “qué horarios tienen los trabajadores de este departamento”, indicando que la jornada laboral del responsable de este servicio suele ser de ocho horas, sin especificar una franja horaria aproximada. En todo caso, al igual que en el estudio de hace dos décadas se han procurado salvar estas lagunas y podemos decir que en la mayor parte de los casos, salvo las excepciones mencionadas, esta investigación no hubiera sido posible sin la colaboración de las fuentes implicadas, por lo que consideramos que debemos dejar constancia de nuestro agradecimiento.

Tabla 2. La tabla muestra, de entre los diarios analizados, cuáles tienen departamento de documentación (el símbolo “X” significa que el diario pertenece al *Grupo Joly* y que tiene este servicio centralizado en Sevilla)

Cabecera	Tiene depto. de documentación	No tiene	No contesta
<i>Diario de Sevilla</i>	X		
<i>Abc de Sevilla</i>			X
<i>Sur</i>	X		
<i>La opinión de Málaga</i>		X	
<i>Diario de Cádiz</i>	X		
<i>La voz</i>		X	
<i>Ideal</i>	X		
<i>Granada hoy</i>	X		
<i>Córdoba</i>	X		
<i>El día de Córdoba</i>	X		
<i>La voz de Almería</i>	X		
<i>Diario de Almería</i>	X		
<i>Diario Jaén</i>			X
<i>Viva Jaén</i>		X	
<i>Huelva información</i>	X		
<i>Viva Huelva</i>		X	

#### 3.1 Existencia de departamento de documentación

En 1996, la investigación de Aquesolo reveló que todos los diarios, afirmaban tener “un archivo o centro de documentación que se pudiera identificar como tal”, a excepción del *Diario de Cádiz* que tan sólo contaba con un archivo fotográfico. Casi todos los departamentos disponían de una persona que se encargase de la labor de documentalista, *Ideal* con dos (**Aquesolo-Vegas, 1996**).

En la actualidad no todos los diarios cuentan con un departamento de documentación. El diario *La voz* del grupo *Vocento* afirma contar tan sólo con una base de datos; *Diario Córdoba*, afirma que tiene un servicio de este tipo pero que no hay ninguna persona encargada de su gestión, es decir, cuenta más bien con un archivo digital que cualquier redactor puede gestionar desde su ordenador; *Viva Jaén* y *Viva Huelva* del grupo *Publicaciones del Sur SA* dejan esta labor en manos de los periodistas; en el caso de *La opinión de Málaga*, indican que fue suprimido. Entre los que sí tienen un departamento de documentación se encuentran los diarios del *Grupo Joly*: *Diario de Sevilla*, *Diario de Cádiz*, *Granada hoy*, *Diario de Almería*, *El día de Córdoba* y *Huelva información*, si bien comparten el mismo departamento centralizado en Sevilla del cual se encargan dos personas; *La voz de Almería (Prisa)* con una persona al cargo; y *Sur* e *Ideal* de *Vocento* tienen dos personas en el primero y una en el segundo que se encargan de la administración de dichos servicios. Los datos en lo referente a la cantidad de personal del departamento no salen bien parados de la comparación con los que arroja el estudio de **Marcos-Recio y Edo (2015)**, que pidieron a los periodistas encuestados en su investigación que indicaran el número de documentalistas que trabajan en su medio:

“un 14% respondió entre 5 y 10 documentalistas, un 17% entre 10 y 20, un 18% más de 20 y, por último, un 48% entre 1 y 5” (Marcos-Recio; Edo, 2015, p. 403).

A lo que dichos autores añaden:

“según estas respuestas ha bajado el número con respecto a hace una década, donde la mayoría de los centros contaban con más de 20 documentalistas. Se considera que la crisis que atraviesan los medios también se ha reflejado en los centros de documentación, ya que han perdido documentalistas cuando el medio ha despedido a parte de su plantilla” (Marcos-Recio; Edo, 2015, p. 403).

### 3.2. Personal adscrito al servicio y titulación

Dentro de los trabajadores de este departamento, en lo que se refiere a la titulación, en 1996, dos eran procedentes de formación profesional, otros dos de estudios de bachillerato superior, uno poseía un título de grado medio (sin especificar), uno era redactor gráfico y dos no contestaron. Es decir, no se encontraba ningún documentalista titulado, bien en ciencias de la información o de la comunicación, bien en biblioteconomía y documentación, que son las titulaciones más afines.

Respecto a la titulación actual de dicho personal, en el caso del *Grupo Joly* poseen una licenciatura y cursos en formación de biblioteconomía y documentación; en *La voz de Almería* la formación se queda en la de estudios de “administración”, según se expresó en la encuesta; licenciados en periodismo en el diario *Sur* y, en *Ideal*, una licenciada en ciencias de la información con un máster en documentación digital, con lo que se advierte una notable mejora en la formación de los profesionales. La figura 1 muestra cómo en 1996 el 75% del personal que se encontraba a cargo de este servicio provenía de estudios de grado medio, mientras que en la figura 2 se advierte como ese porcentaje baja hasta el 25% y nos encontramos con que el 75% del personal posee estudios de grado superior.

### 3.3. Horario

En el estudio de 1996 la mayoría de los diarios afirmaban que la tarde era el período más importante –las horas iban entre las 16:00 y las 24:00-. En el caso de *El correo de An-*

*dalucía y Huelva información* abrían el servicio por las mañanas. *Córdoba* y *La voz de Almería* no respondieron a esta pregunta en concreto (Aquesolo-Vegas, 1996).

En nuestro estudio el horario varía bastante de un diario a otro: desde la existencia de dos turnos divididos entre las 9:30h y las 21:00h en el *Grupo Joly*, o el caso de *Sur* que también se divide entre la mañana y la tarde, al de *La voz de Almería* que tan sólo cuenta con turno de mañana (de 9:00h a 14:00h); *Ideal* no especifica el horario pero afirma que suele ser de 8 horas (no indica la franja horaria) y añade a su vez que “depende del día”.

### 3.4. Ubicación e instalaciones

En la década de los 90 los locales habilitados con el fin de ser utilizados como servicios de documentación se hallaban próximos a la redacción (incluso formaban parte de ésta). Con la excepción de *Huelva información* que lo situaba en otra planta, junto a la sala de administración. El espacio asignado variaba bastante de unos diarios a otros, yendo desde los 4 m<sup>2</sup> de *Huelva información*, a los 14 m<sup>2</sup> de *Córdoba*, pasando por los 35 m<sup>2</sup> de *El correo de Andalucía* y los 50 m<sup>2</sup> de *Diario de Cádiz*, hasta alcanzar los 95 m<sup>2</sup> de *Ideal* (Aquesolo-Vegas, 1996).

En nuestros días la situación también es bastante diversa dependiendo de la cabecera a examinar. Los espacios habilitados con este fin están situados cerca de la redacción o forman parte de la misma. Su tamaño varía de las “dos mesas” en el diario *Ideal*, a los 2 m<sup>2</sup> de *La voz de Almería*, pasando por los 12 m<sup>2</sup> del diario *Sur*, hasta llegar a los aproximadamente 25 m<sup>2</sup> del *Grupo Joly*.

### 3.5. Equipamiento

En los resultados obtenidos por Aquesolo, *El correo de Andalucía* era el único que tenía parte de su archivo informatizado (la recepción, búsqueda y archivo de fotografías). En los demás diarios ninguno contaba con archivos informatizados; su sistema de archivos, en general, se basaba en carpetas en las cuales los recortes se clasificaban y almacenaban por temas y estructuras temáticas. *Córdoba*, *El correo de Andalucía* y *Jaén* disponían de disco óptico para almacenar las imágenes. El resto del equipamiento se limitaba a alguna fotocopiadora o fax (Aquesolo-Vegas, 1996).

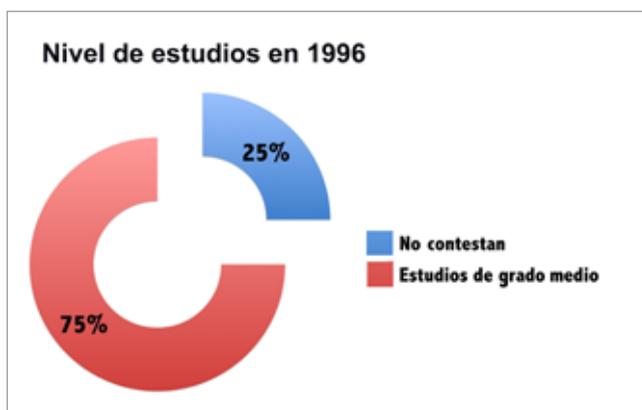


Figura 1. Personal con estudios de grado medio en las cabeceras andaluzas en 1996. Fuente: Aquesolo (1996).

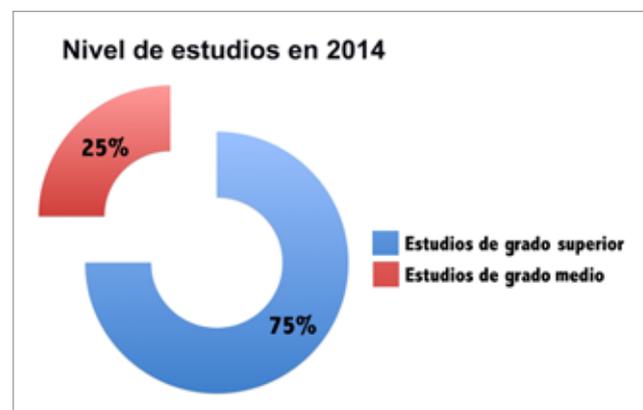


Figura 2. Personal con estudios de grado medio y de grado superior en las cabeceras andaluzas

Mientras, nuestro análisis muestra cómo en estos momentos todos poseen un equipamiento informático suficiente: en lo que se refiere al hardware, el *Grupo Joly* habla de dos ordenadores, *Córdoba* entre sus equipos cuenta con un servidor *HP Proliant DL360* y *La Voz de Almería* confirma que la información puede ser consultada desde “cualquier ordenador de redacción”. Sin embargo, *Sur* e *Ideal* directamente no responden a esta pregunta del cuestionario. Respecto al software de dichos departamentos, el *Grupo Joly* afirmaba que es “una producción propia del departamento de informática”; *La voz de Almería* indicó que dispone de “un buscador creado por la propia empresa”; en el diario *Córdoba* el software fue “realizado por el departamento de I+D de *Grupo Zeta*”, al cual pertenece dicha cabecera; *Sur* se limitaba a confirmar que era “propio del periódico”; y, por último, *Ideal* nos refiere a “*Aplidoc*, elaborado para los periódicos pertenecientes al *Grupo Vocento*”.

### 3.6. Información que suministran

En el estudio de la década de los 90 los periódicos afirmaban que su fondo de archivo estaba compuesto fundamentalmente por revistas a las que estaban abonados, boletines oficiales e información proveniente de publicaciones anteriores del propio diario (**Aquesolo-Vegas**, 1996). Estos contenidos, como vemos, fundamentalmente en papel, eran los que entonces los departamentos podían proporcionar para completar las informaciones.

Cuando preguntamos a los responsables del departamento de documentación de cada diario cuál era la información más requerida, el *Grupo Joly* manifestó que en su caso se trataba de “noticias anteriores del caso que se investiga, publicadas por su propio medio”, además de “fotografías de archivo”; *Sur* afirmó que tomaban como base “la información publicada en el periódico” y la complementaban “con información directa proporcionada por los redactores”. *La voz de Almería* simplemente se limitaba a responder que “de todo tipo”; el diario *Córdoba* contestaba vagamente que las más solicitadas eran las “informaciones publicadas de temas de actualidad”; e *Ideal* declara que su mayor demanda es la de información “local de Granada, Jaén y Almería (datos estadísticos, dossiers históricos sobre temas locales, biografías sobre personajes locales)”. Estos tres últimos no especificaban si la información era propia o proveniente de otras fuentes.

De lo anterior se deduce que, de todos los datos que suministran los departamentos en la actualidad, los más requeridos son noticias y fotos de archivo del propio periódico, en su mayoría de informaciones locales que puedan ser demandadas por los redactores con posterioridad. Puede decirse que, aunque con soportes diferentes, las demandas son parecidas y derivan también del tipo de periódicos que estamos estudiando, apegados como es lógico a las informaciones territoriales de más corto alcance.

## 4. Discusión y conclusiones

Observando los resultados obtenidos, coincidimos con la apreciación de Javier Guallar cuando afirma que, pese a la importancia del trabajo documental, éste está escasamente valorado, sobre todo cuanto más pequeña es la empresa

informativa. Según **Guallar** (2011b, pp. 65-66), son numerosas las redacciones que cuentan con un departamento de documentación y con profesionales de esta especialidad. En general, las empresas de tamaño medio o grande cuentan con un centro de este tipo, siendo las más pequeñas las que tienen dificultades para mantener estos servicios en funcionamiento y actualizados.

Aunque después de comparar los datos obtenidos en ambos estudios se pueda llegar a la rápida e incompleta conclusión –fruto quizá de las limitaciones más arriba comentadas– de que, en proporción, hay el mismo número de periódicos que cuentan con un servicio de documentación que en 1996, la realidad es más complicada: si bien podría afirmarse lo anterior, al realizar un examen más detallado se percibe que el número de departamentos se ha visto reducido al centralizar sus servicios entre los mismos periódicos de un grupo editorial.

Para soportar los costes se tiende a una fuerte concentración mediática a fin de compartir recursos y abaratar gastos. Se generan

“fuertes procesos de reconversión y la afloración de intensivas dinámicas de innovación” (**Campos-Freire**, 2010, p. 14).

Todo esto se traduce en que, por ejemplo, un grupo mediático como *Joly* haya centralizado en Sevilla el departamento de documentación para todos sus diarios dentro de la Comunidad Andaluza, tanto en papel como en digital.

Hecha esta primera constatación, con todo, tras analizar los resultados obtenidos y compararlos con los que cosechó José-Antonio Aquesolo-Vegas en 1996, podemos afirmar que se observan aspectos tanto positivos como negativos.

Entre los positivos destaca sin duda la mejora en la preparación del personal adscrito a estos departamentos: el estudio de Aquesolo apuntaba cómo el perfil profesional del personal no era acorde al trabajo a desempeñar –graduados en enseñanzas de grado medio– sin embargo, los datos obtenidos en las encuestas suministradas a los periódicos en esta última muestra perfilan una notable mejora. Lo general es que el personal cuente con un título de enseñanza superior relacionado directamente con la documentación (a excepción de *La voz de Almería* cuyo responsable está formado en enseñanzas administrativas).

Además, hoy podemos resaltar el avance en la informatización del equipamiento, que en 1996 era prácticamente inexistente y ahora es un hecho: toda la información se encuentra automatizada y plenamente interconectada de forma que los redactores pueden acceder a ésta cuando es precisa su recuperación para la contextualización de una noticia. Según evidencian los datos obtenidos, el avance en los equipos ha experimentado un impulso abrumador. El estudio de Aquesolo reflejaba que eran muy pocos los periódicos que tenían automatizado su centro de documentación, salvo excepciones como el *Correo de Andalucía* que tenía informatizada la recepción, búsqueda y archivo de material fotográfico y cuyos servicios ya permitían la conexión en red y la accesibilidad a la información almacenada. No obstante, en dicho estudio ya se presentía su pronta implantación:

los que afirmaban carecer de informatización en el servicio señalaban su intención de incorporarla.

Los datos obtenidos casi dos décadas más tarde son una clara manifestación del arraigo de las TIC en nuestra sociedad, pues no hay un sólo diario, entre aquellos que cuentan con un servicio de este tipo, cuyo equipamiento no se halle informatizado y en el cual la conexión en red y la accesibilidad no sean una realidad; es más, lo habitual es que cada grupo editorial cuente con un software de creación propia que rija su sistema informático.

Ambos resultados, sobre personal cualificado y equipamiento actualizado, resultan contradictorios con respecto a la casi extinción de los departamentos de documentación, aunque se explican sobre todo por la gran tendencia a la centralización que se deriva de la concentración de las empresas editoriales. No obstante, no basta con dotar un espacio del equipamiento y personal adecuado para conseguir un servicio de documentación satisfactorio: hace falta un esfuerzo motivador para alcanzar una cultura organizacional orientada a compartir y generar conocimientos, mediante el aprovechamiento de las capacidades creativas e innovadoras del factor humano. Los trabajadores originan el capital intelectual, por lo que sería saludable que la empresa potenciase las diferencias intelectuales, facilitando la generación de ideas y su puesta en práctica (Reyes-Ramírez, 2007).

Debemos efectuar un último apunte antes de cerrar esta cuestión llamando la atención ante el hecho de que, pese a que el volumen de información que tiene que gestionar los departamentos de documentación aumenta cada día de manera ingente, éstos cuentan en la actualidad con aproximadamente el mismo número de personas para su administración que hace dos décadas cuando Aquesolo realizó su estudio.

Entramos así en los aspectos negativos que se derivan de los resultados de este trabajo: se observa como en 1996 los diarios implementaban la funcionalidad de sus departamentos documentales, en una clara etapa de despunte tecnológico lo cual suponía introducir mejoras en un servicio insustituible. Sin embargo, en el momento actual, algunos de los diarios analizados perciben una decadencia en su razón de ser y los consideran prescindibles —como ocurre con *La opinión de Málaga* donde, en determinados momentos como el actual, se prescinde de este departamento— e, incluso, obsoletos —en el caso de *Viva Huelva* y *Viva Jaén* la persona responsable de atender la llamada de contacto confiesa que esa labor recae sobre los propios periodistas— ya que se considera innecesario e, incluso, para algunos resulta impensable realizar una inversión dirigida a impulsar estos servicios.

Con todo, los periodistas aseguran no tener dudas de la importancia que posee la documentación como parte de un adecuado proceso informativo. Resulta irónico, si se tiene en cuenta que son contadas las cabeceras españolas que cuentan con un equipo de investigación, hecho que se achaca a la idea de que todo periodista realiza una labor de investigación innata a su labor profesional y al elevado coste que supone el mantenimiento de un equipo de este tipo (Díaz-Güell, 2003, pp. 58-59). Pero quizá porque piensen que el departamento de documentación es tan sólo un es-

pacio físico, y desconocen el perfil y aptitudes del documentalista, viéndolos como meros organizadores del archivo, como si se encargaran en exclusiva del trabajo que ejecuta un bibliotecario. En este sentido, el 50% de los periodistas encuestados en el estudio de 2009 referido con anterioridad apostaban por la futura desaparición de los departamentos de documentación al ser suplidos por internet (Micó-Sanz; Masip; García-Avilés, 2009, p. 288).

Y sin embargo, curiosamente, el centro de documentación no tendría por qué ser un desconocido para los redactores, teniendo en cuenta su integración y su proximidad física. La cercanía de la sala de redacción respecto a la del departamento de documentación es la opción habitual, aunque tampoco se ve como imprescindible actualmente debido a la tecnología empleada. Cualquier redactor puede acceder al fondo documental sin tener que desplazarse de su espacio de trabajo, basta con un clic en el teclado. Así pues, resulta interesante destacar que, aunque tanto en el estudio de 1996 como en el actual, los departamentos de documentación cuentan con un tamaño que varía drásticamente de unos a otros, rozando en algunos casos lo incomprensible (por ejemplo, las “dos mesas” disponibles para esa tarea en el diario *Ideal* o los dos metros cuadrados que le reserva *La voz de Almería*), en general, las dimensiones que se dedicaban a estos espacios se han visto reducidas: el caso más llamativo lo encontramos al analizar el servicio de documentación del diario *Ideal* que antes contaba con 95 m<sup>2</sup>. Lo expuesto no es de extrañar si se tiene en cuenta que se ha pasado de un sistema de almacenamiento físico a otro digital, pero también nos aporta un dato simbólico respecto a la consideración de estos espacios.

Hay que reseñar además que la crisis económica sumada a la, digamos, identitaria de la prensa también ha hecho mella, como se pudo apreciar en la conversación telefónica mantenida con una trabajadora de *La opinión de Málaga* que confirmaba que esa era una de las principales razones por las que se había suprimido el departamento de documentación de dicho diario.

Así pues, ante este escenario, queda en entredicho la figura del documentalista informativo tal y como hasta ahora estaba planteada y surgen nuevas propuestas:

“En el centro de documentación hemos pasado de un simple diálogo —y en ocasiones monólogo— a una conversación permanente en la que se han de facilitar respuestas rápidas y precisas sobre las necesidades informativas, no sólo de los propios periodistas, sino de aquellos usuarios que estén dispuestos a pagar por ella” (Marcos-Recio, 1998, p. 124).

Siguiendo a Guallar, sin embargo, no todos los profesionales parecen entender por igual la importancia de estos nuevos productos y servicios que, pese a no ser plenamente explotados, son su mejor arma. Los documentalistas y los centros de documentación deben encaminarse a la creación de contenidos, superando en la medida de lo posible anteriores roles subordinados y de intermediación (Guallar, 2011b, p. 66).

Así, entre las nuevas funciones que se vienen proponiendo, destaca también la de la gestión y difusión de recursos disponibles en el ciberespacio dentro de la propia empresa,

que ya se empezaba a subrayar en estudios de hace dos décadas como citábamos más arriba:

“Con expertos documentalistas en el manejo de estas redes, para hacer una acertada selección, de tal forma que sirvan como filtro de documentos que han de entregar al periodista, quien ha de asumir por su parte la confianza plena de que esa selección se ha llevado a cabo con criterio y rigor periodístico y documental, lo que no incidirá de forma negativa en la elaboración final de la información” (Marcos-Recio, 1998, p. 118).

Posteriormente se ha continuado defendiendo la idea de que el documentalista puede ser hoy un orientador que ayude a seleccionar las mejores fuentes (García-González, 2006, p. 6).

Pese a los limitados conocimientos y a la falta de formación que los periodistas admiten tener de la búsqueda de información en la Red, dichos profesionales parecen preferirla frente a sus propias bases de datos. Cabe recordar que son minoría los diarios que ofrecieron cursos de formación a sus trabajadores sobre un uso adecuado de internet, hecho que repercute negativamente en la información ofrecida por los periodistas y que se refleja en un descenso en la calidad de la información (Micó-Sanz; Masip; García-Avilés, 2009, p. 287).

Sin embargo, hoy comprobamos cómo estas propuestas, aun siguiendo vigentes, no se han secundado en la práctica. Con todo, esperamos que estudios como éste que llamen la atención sobre la situación actual de los servicios de documentación en la prensa, contribuyan a una reflexión pues coincidimos en que los documentalistas informativos pueden convertirse en el remo del que se vale el periodista para navegar en el mar de datos que ha provocado el acceso masivo a la Web y la consiguiente saturación informativa, pues los documentalistas

“aportan una perspectiva totalmente diferente a las redacciones” al centrarse “en el pasado y el futuro” (Paul, 2009).

## 5. Bibliografía

**Aquesolo-Vegas, José-Antonio** (1996). “Situación de los servicios de documentación de la prensa diaria de Andalucía”. *Cuadernos de documentación multimedia*, n. 5, pp. 69-73. <http://pendientedemigracion.ucm.es/info/multidoc/multidoc/revista/cuadern5/aquesolo.htm>

**Campos-Freire, Francisco** (2010). “Los nuevos modelos de gestión de las empresas mediáticas”. *Estudios sobre el mensaje periodístico*, n. 16, p. 13-30. <http://revistas.ucm.es/index.php/ESMP/article/view/ESMP1010110013A>

**Codina, Lluís** (2000). “La documentación de los medios de comunicación: Situación actual y perspectivas de futuro”, *Cuadernos de documentación multimedia*, n. 10, pp. 49-66. <https://dialnet.unirioja.es/servlet/articulo?codigo=1408512>

**Díaz-Güell, Luis** (2003). *Periodismo y periodistas de investigación en España, 1975-2000: contribución al cambio político, jurídico, económico y social*. Tesis doctoral. Universidad

Complutense de Madrid, Departamento de Periodismo III. <http://eprints.ucm.es/4843>

**Estrada-Nora-González, Belén; Patallo-Fernández, Cristina; Pastor-Blanco, Marcela** (2009). “Puesta en marcha de un servicio de documentación digital en la *Televisión del Principado de Asturias (TPA)*”. *El profesional de la información*, v. 18, n. 3, pp. 326-332. <http://dx.doi.org/10.3145/epi.2009.may.11>

**Galdón-López, Gabriel** (2002). *Teoría y práctica de la documentación informativa*. Barcelona: Ariel. ISBN: 978 8434412934

**González-Quesada, Alfonso** (2003). “La investigación en documentación informativa en España”. *Documentación de las ciencias de la información*, v. 26, pp. 71-97. <http://revistas.ucm.es/index.php/DCIN/article/view/DCIN0303110071A>

**Guallar, Javier** (2011a). “Documentación fotográfica en la prensa. Casos de *El país*, *El periódico* y *La vanguardia*”. *El profesional de la información*, v. 20, n. 4, pp. 392-398. <http://dx.doi.org/10.3145/epi.2011.jul.05>

**Guallar, Javier** (2011b). “La documentación en la prensa digital. Nuevas tendencias y perspectivas”. En: *III Congreso internacional de ciberperiodismo y web 2.0. La transformación del espacio mediático*, pp.52-68. <http://eprints.rclis.org/16326>

**Guallar, Javier; Cornet, Anna** (2004). “Fuentes de información digitales en los centros de documentación de prensa: las bases de datos comerciales”. *El profesional de la información*, v. 13, n. 2, pp. 107-117. <http://www.elprofesionaldelainformacion.com/contenidos/2004/marzo/4.pdf>

**Guallar, Javier; Cornet, Anna** (2005). “Fuentes de información digitales en los centros de documentación de prensa: recursos web gratuitos”. *El profesional de la información*, v. 14, n. 4, p. 296-308. <http://www.elprofesionaldelainformacion.com/contenidos/2005/julio/9.pdf>

**Marcos-Recio, Juan-Carlos** (1998). “Una nueva concepción de la documentación en los medios electrónicos: Retos y nuevas tareas profesionales”. *Documentación de las ciencias de la información*, n. 21, pp. 113-130. <http://eprints.ucm.es/7023>

**Marcos-Recio, Juan-Carlos; Edo, Concha** (2015). “Análisis de la nueva perspectiva de la documentación periodística en los medios de comunicación españoles”. *Revista general de información y documentación*, v. 25, n. 2, pp. 389-423. [http://dx.doi.org/10.5209/rev\\_RGID.2015.v25.n2.51235](http://dx.doi.org/10.5209/rev_RGID.2015.v25.n2.51235)

**Marcos-Recio, Juan-Carlos; Sánchez-Vigil, Juan-Miguel; Olivera-Zaldua, María** (2010). “Modelos de gestión documental en las agencias de publicidad”. *El profesional de la información*, v. 19, n. 2, pp. 175-183. <http://dx.doi.org/10.3145/epi.2010.mar.08>

**Micó-Sanz, Josep-Lluís; Masip, Pere; García-Avilés, José-Alberto** (2009). “Periodistas que ejercen de documentalistas (¿y viceversa?). Nuevas relaciones entre la redacción y el

archivo tras la digitalización de los medios". *El profesional de la información*, v. 18, n. 3, pp. 284-290.

<http://dx.doi.org/10.3145/epi.2009.may.05>

**Moreiro-González, José-Antonio** (coord.) (2000). *Manual de documentación informativa*. Madrid: Cátedra. ISBN: 978 84 376 1798 5

**Orero, Paz; Cebrián-Enrique, Bernardino J.** (2014). "Criterios de evaluación y aplicación de fuentes de información web en centros de documentación periodística". *El profesional de la información*, v. 23, n. 6, pp. 612-617.

<http://dx.doi.org/10.3145/epi.2014.nov.08>

**Paul, Nora** (2009). "Elegía del centro de documentación de prensa", *El profesional de la información*, v. 18, n.3, pp. 249-253.

<http://dx.doi.org/10.3145/epi.2009.may.01>

**Reyes-Ramírez, Livia-Mercedes** (2007). "Sistemas de información para la prensa: la gestión de la información y el conocimiento en el contexto de los sistemas integrados de información". *Acimed*, v. 15, n. 2.

[http://eprints.rclis.org/9104/1/Sistemas\\_de\\_informaci%C3%B3n\\_para\\_la\\_prensa.pdf](http://eprints.rclis.org/9104/1/Sistemas_de_informaci%C3%B3n_para_la_prensa.pdf)

**Ros-Martín, Marcos; Rodero-Susiac, Alfonso** (2009). "El

servicio de documentación de un medio impreso regional: evolución en el diario *Levante-EMV*". *El profesional de la información*, v. 18, n. 3, pp. 316-322.

<http://dx.doi.org/10.3145/epi.2009.may.09>

**Rubio-Lacoba, María** (2005). "Nuevos tiempos para la documentación informativa en el periodismo digital: viejas y nuevas funciones del servicio de documentación digital". *Comunicación y sociedad*, v. 18, n. 1, pp. 153-168.

<http://eprints.rclis.org/16327>

**Rubio-Lacoba, María** (2010). "Documentalistas de prensa, ¿cuál es vuestro oficio?". *El profesional de la información*, v. 19, n. 6, pp. 645-651.

<http://dx.doi.org/10.3145/epi.2010.nov.11>

**Sánchez-Vigil, Juan-Miguel; Marcos-Recio, Juan-Carlos; Olivera-Zaldua, María** (2009). "Influencia de la fotografía digital en los departamentos de documentación de prensa". *El profesional de la información*, v. 18, n. 3, pp. 278-283.

<http://dx.doi.org/10.3145/epi.2009.may.04>

**Torres-Rodríguez, Vanessa-Berenice** (2012). *Documentación periodística*. México: Red Tercer Milenio. ISBN: 978 607 733 042 4

## Colección EPI Scholar

### Libros científicos de Información, Documentación y Comunicación



#### Gestión de datos de investigación de Fabiano Couto Corrêa

Vivimos momentos de auténtica expectación por el papel que el acceso a grandes cantidades de datos de investigaciones puede llegar a jugar en el presente y el futuro inmediato en el desarrollo del conocimiento en todas las disciplinas. Para ello es absolutamente clave la existencia de procedimientos adecuados en la gestión de los datos de investigación.

En este libro se presentan diversas recomendaciones para ello. Las recomendaciones están dirigidas en primer lugar a investigadores. Y como los bibliotecarios tienen un papel cada vez más importante en estos procesos, se propone también un enfoque adicional: "ayudarles a ayudar a los investigadores". Así, ambos colectivos pueden encontrar en estas páginas una guía tremendamente útil para la gestión de los datos de investigación.

Del prefacio de Fernanda Peset: "Aunque gestionar el conocimiento en el ámbito de las ciencias es extremadamente complejo, en este momento enfrentamos la oportunidad de convertir en realidad el acceso a los resultados de la investigación en su más amplia acepción. Y el libro que tienen en las manos, les ayudará a comprender e involucrarse en las diferentes dimensiones de la gestión de los datos de investigación".

**Couto Corrêa, Fabiano** (2016). *Gestión de datos de investigación*. Barcelona: El profesional de la información, Editorial UOC, colección EPI Scholar n. 6, 164 pp. ISBN: 978 84 9116 430 2

**Información:** Isabel Olea

[epi.iolea@gmail.com](mailto:epi.iolea@gmail.com)

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

# PREVALENCIA Y CAPACIDAD DE RECONOCIMIENTO DEL PLAGIO ACADÉMICO ENTRE EL ALUMNADO DEL ÁREA DE ECONOMÍA

## Prevalence and ability to recognize academic plagiarism among university students in economics

Rubén Comas-Forgas y Jaume Sureda-Negre



**Rubén Comas-Forgas** es doctor europeo en ciencias de la educación por la *Universidad de las Islas Baleares* donde actualmente es profesor adscrito al *Departamento de Pedagogía Aplicada y Psicología de la Educación*. Sus líneas de trabajo se centran en la integridad académica y la documentación científica en ciencias de la educación. Ha sido investigador/profesor visitante en diversos centros como: *University of East Anglia* (Reino Unido), *Liverpool John Moores University* (Reino Unido), *Panteion University of Social and Political Sciences* (Grecia), y *Universidad Autónoma de Yucatán* (México).

<http://orcid.org/0000-0002-8885-753X>

[rubencomas@uib.es](mailto:rubencomas@uib.es)



**Jaume Sureda-Negre** es catedrático de universidad y director del *Departamento de Pedagogía Aplicada y Psicología de la Educación* de la *Universidad de las Islas Baleares*. Ha ocupado durante varios años cargos de gestión académica (vicerrectorados y dirección de institutos de investigación). Ha dirigido numerosos proyectos de investigación competitivos tanto a nivel nacional como internacional. Sus líneas de trabajo se centran en la integridad académica y la documentación científica en ciencias de la educación.

<http://orcid.org/0000-0003-2689-5626>

[sureda.negre@gmail.com](mailto:sureda.negre@gmail.com)

*Universidad de las Islas Baleares*

Campus UIB, Edificio Beatriu Pinos, 16. 01722 Palma de Mallorca, España

### Resumen

Se estudia la relación existente entre el conocimiento de lo que supone plagio académico y el conocimiento de las normas de citación, y la comisión de plagio entre alumnado de último curso del área de Economía en una universidad de tamaño medio española. Para ello se contó con la participación de una muestra (n=325) de alumnos y alumnas a quienes se les administró un cuestionario elaborado *ad-hoc* para este estudio. Esta investigación pone de manifiesto los déficits que tiene el alumnado participante para determinar cuándo un texto está plagiado y qué es una cita y una referencia. Los resultados sugieren la inexistencia de relación entre el reconocimiento de lo que supone un plagio y la prevalencia en su comisión. El estudio aporta datos relevantes acerca del alumnado que comete plagio de forma accidental y aquél que lo hace intencionalmente.

### Palabras clave

Competencias informacionales; Plagio académico; Integridad académica; Evaluación de competencias; Normas de citación.

### Abstract

The relationship between knowledge of what constitutes academic plagiarism, knowledge of citation norms, and academic plagiarism perpetration among students in their final year of Economics at an average size Spanish university is analyzed. The research counted on the participation of a sample (n=325) of students who responded to a questionnaire prepared *ad-hoc* for this study. This research highlights the deficits that participating students had in determining when a text is plagiarized and in recognizing a citation or a reference. The results suggest an absence in students' ability to recognize plagiarism relates directly to the commission of it. The study provides data on students who commit plagiarism accidentally and those who do it intentionally.

### Keywords

Information competencies; Academic plagiarism; Academic integrity; Competencies assessment; Citation norms.

Artículo recibido el 26-10-2015

Aceptación definitiva: 16-03-2016

**Comas-Forgas, Rubén; Sureda-Negre, Jaume (2016).** "Prevalencia y capacidad de reconocimiento del plagio académico entre el alumnado del área de economía". *El profesional de la información*, v. 25, n. 4, pp. 616-622.

<http://dx.doi.org/10.3145/epi.2016.jul.11>

## 1. Introducción

La comunicación académica debe regirse, entre otros, por los principios de honestidad, veracidad, rigor, calidad, probidad e integridad. La contravención de dichos valores supone la comisión de prácticas deshonestas y fraudulentas. Engañar, falsificar, defraudar, mentir, adulterar, ocultar, confundir, etc., son verbos no conjugables en la transmisión e intercambio de conocimiento en el campo educativo.

A pesar de ello son numerosas las señales e indicios que advierten que las prácticas deshonestas están ciertamente asentadas en las universidades (Hallak; Poisson, 2002; Poisson, 2010). Y no sólo eso, el problema es todavía más alarmante ya que parece existir cierto acuerdo en cuanto a que los usos y praxis fraudulentas van en aumento en lugar de retroceder (Hallak; Poisson, 2007; Christakis; Christakis, 2012).

La modalidad de fraude en las actividades de evaluación más extendida entre el alumnado es, sin duda alguna, la comisión de plagio académico. Por ejemplo, centrándonos en España, según los escasos estudios llevados a cabo sobre la materia, más de la mitad de los universitarios españoles (cifras en torno al 60%) manifiesta haber incidido en alguna modalidad de plagio académico a la hora de redactar trabajos para asignaturas de sus estudios (Comas-Forgas, 2009; Comas-Forgas; Sureda-Negre; Salva-Mut, 2010).

Y este hecho, en bastantes ocasiones, se ha relacionado con un déficit de competencias informacionales entre el alumnado -por ejemplo: desconocimiento de normativas de citación y elaboración de referencias, incapacidad para reconocer lo que supone plagio de lo que no (Moon, 2002; Guedes; Gomes-Filho, 2015, DeGeeter *et al.*, 2014)- lo que genera situaciones en las que el alumnado incurre en plagio de forma inconsciente o accidental (Park, 2003). A la vez, se ha demostrado que una adecuada formación en competencias informacionales actúa como factor de prevención y reducción de los casos de plagio académico entre alumnado (Hurlbert; Savidge; Laudenslager, 2003; Humphreys, 2013).

Ante esta situación conviene poner de manifiesto que unas pocas universidades españolas, desde sus servicios de biblioteca y documentación, ofrecen recursos formativos/informativos al alumnado acerca de cómo eludir la comisión de plagio en la presentación de trabajos académicos (Domínguez-Aroca, 2012). Además, algunos centros universitarios españoles han introducido en sus planes de estudio asignaturas centradas en mejorar las competencias informacionales del alumno (búsqueda y localización de información, valoración de la información, citar y referenciar las fuentes, etc.). Sea como fuere, todavía no disponemos de datos y evidencias para conocer el impacto de estas iniciativas en la mejora de las competencias del alumnado para producir trabajos académicos de calidad.

Las cuestiones de investigación (CI) que se intentan responder con esta propuesta son:

CI1: ¿Son capaces los alumnos de último curso del área de economía de reconocer un texto plagiado y diferenciarlo de un texto elaborado con arreglo a los principios y estándares académicos?

CI2: ¿Son capaces los alumnos de último curso del área de economía de reconocer qué es una cita y una referencia bibliográfica?

CI3: ¿Cuál es el nivel de prevalencia en la comisión de plagio académico entre el alumnado de último curso del área de economía?

C31: ¿Existen diferencias significativas entre el alumnado de último curso del área de economía en cuanto a la frecuencia de comisión de plagio académico atendiendo al género?

C32: ¿Existen diferencias significativas entre el alumnado de último curso del área de economía en cuanto a la frecuencia de comisión de plagio académico atendiendo a su capacidad para reconocerlo?

C33: ¿Existen diferencias significativas en cuanto a la frecuencia de comisión de plagio académico atendiendo al conocimiento que posee el alumno de lo que son las citas y las referencias bibliográficas?

## 2. Material y método

### 2.1. Población y muestra

En el estudio han participado 325 estudiantes de cuarto curso, de los años 2013-2014 y 2014-2015, del grado en administración y dirección de empresas (GADE) y del grado de economía (GECO) de la *Universidad de las Islas Baleares*. Ello representa una muestra con un margen de error, calculado para el total de alumnado matriculado en cuarto curso de las dos titulaciones en ambos años académicos (n=492), de  $\pm 3,17\%$  para un intervalo de confianza estimado del 95% bajo la condición más desfavorable de  $p=q=0,50$ . El alumnado de estas titulaciones no ha contado en su formación con ninguna asignatura estrictamente dedicada a mejorar sus competencias informacionales.

La recogida de datos se realizó, de manera individual y anónima en situaciones de aula con la presencia del equipo de investigadores, en dos tandas: enero de 2014 (alumnado del curso 2013-14) y enero de 2015 (alumnado del curso 2014-15). Ningún alumno rehusó participar en el estudio; sin embargo, pese a que se obtuvo un total de 329 encuestas, la muestra se contrajo hasta 325 sujetos, dado que 4 cuestionarios fueron invalidados por alguno de estos motivos:

- multiplicidad de respuesta en ítems de respuesta única;
- cuestionarios contestados en un porcentaje inferior al

- 50% de los items;
- ininteligibilidad de las respuestas dadas por los encuestados.

En relación a las características de los sujetos:

- el 48,9% fueron mujeres y el 51,1% hombres;
- accedieron a la universidad por distintas vías: 92,6% a través de la Prueba de acceso a la universidad (PAU o Selectividad), 1,9% a través de pruebas de acceso a mayores de 25 años, 4% a través de una titulación de FP de grado superior, y un 1,5% a través de otras titulaciones;
- en cuanto a la distribución por edades, éstas varían entre los 21 y los 48, situándose la media en 23,7 años (con una desviación típica de 3,3).

## 2.2. Fuente de datos y variables

En la recolección de datos se empleó como instrumento el *Cuestionario sobre plagio académico entre el alumnado de GADE y GECCO*, diseñado expresamente para el presente estudio y fundamentado en:

- análisis de la bibliografía existente sobre la materia;
- adaptación de diversos items de los cuestionarios validados de **Roig** (1997), **Comas-Forgas** (2009) y **Morey** (2011).

Este tipo de encuestas son los instrumentos más utilizados para el estudio de la integridad académica, presentando un aceptable nivel de fiabilidad (**Cizek**, 1999; **Comas-Forgas**, 2009). Constaba de 10 preguntas:

- cinco referidas a características del alumnado (género, edad, vía de acceso a los estudios, nota de acceso a la carrera y media de expediente académico);
- una sobre la formación recibida en elaboración y presentación de trabajos académicos;
- dos centradas en analizar el nivel de conocimientos respecto a las citas académicas y las referencias bibliográficas;
- una para determinar el nivel de conocimiento que tienen de lo que supone un texto plagiado; y
- una referida a la comisión de diversas prácticas constitutivas de plagio por parte del alumnado.

Los resultados que se publican en el presente trabajo se centran en el análisis de tres de las variables abordadas en el cuestionario y la posterior asociación entre éstas (V3 con V1, V2 y V4):

V1: Capacidad del alumnado para determinar cuándo un texto está plagiado.

V2: Conocimiento de la normativa de citación y elaboración de referencias bibliográficas.

V3: Frecuencia e índice de perpetración de plagio académico.

V4: Género.

Como indicador de medición de la V1, se adaptó y simplificó el cuestionario *Plagiarism knowledge survey (PKS)* de **Roig** (1997). Al igual que Roig, se presentaron a los participantes en el estudio diversos textos (4 en total). El primero era un fragmento extraído literalmente de la *Wikipedia* sobre la revolución francesa; se hacía saber a los participantes que se trataba un texto original y se les indicaba cuál era su fuente. El texto original se había transformado, por los investigadores, en otros tres textos:

1) Texto A: representaba un parafraseado del texto original con ausencia de cita;

2) Texto B: era un plagio literal del texto original ya que sólo se cambió el orden de las frases y no se citó la fuente;

3) Texto C: se trataba de un uso correcto de la información original, ya que incluían citas directas y parafraseados con citación de la fuente primaria.

Los participantes debían contestar, una vez leídos los cuatro textos, las siguientes preguntas:

- cuál de los tres fragmentos derivados del original supone un plagio;
- cuál de los tres fragmentos derivados del original te parece más adecuado para entregar en un trabajo académico;
- cuál de los tres fragmentos te parece menos adecuado para entregar en un trabajo académico.

Las posibilidades de respuesta para cada pregunta eran: "Texto A", "Texto B", "Texto C" y "No lo sé". De cara a poder establecer un índice de reconocimiento de plagio (en adelante, IRP), se codificaron las respuestas (3 puntos por cada respuesta correcta y 1 punto para las respuestas incorrectas) y se realizó el sumatorio para cada participante de la puntuación obtenida en cada pregunta:

- los alumnos con tres aciertos obtenían una puntuación de 9;
- los que contestaron acertadamente dos de las tres preguntas tenían una puntuación de 7;
- los que contestaron bien una de las tres preguntas sumaban 5 puntos;
- los que no contestaron correctamente ninguna de las preguntas de la V1 obtuvieron 3 puntos.

Para medir la V2 los investigadores emplearon dos preguntas adaptadas de **Morey** (2011):

a) una pregunta en la que se presentaba una definición de lo que es una cita y los participantes tenían 5 opciones de respuesta para contestar a qué hacía referencia la definición presentada ("Cita", "Nota a pie de página", "Referencia bibliográfica", "Bibliografía" y "No lo sé");

b) una pregunta en la que se presentaba una referencia bibliográfica de un libro de un autor, elaborada según la normativa APA y los participantes debían contestar qué era, disponiendo de cinco opciones: "Nota a pie de página", "Paráfrasis bibliográfica", "Cita", "Referencia bibliográfica" y "No lo sé".

Con el fin de generar un índice de conocimiento de las citas y referencias bibliográficas (en adelante ICCRB) se codificaron las respuestas dadas (3 puntos por cada respuesta correcta y 1 punto las respuestas incorrectas) y se realizó el sumatorio para cada participante de la puntuación de cada pregunta:

- los alumnos con dos aciertos obtenían una puntuación de 6;
- los que contestaron acertadamente una de las dos preguntas tenían una puntuación de 4;
- los que no contestaron correctamente ninguna de las preguntas obtuvieron 2 puntos.

La V3 se fundamenta en una escala creada por **Comas-Forgas** (2009) y mide las respuestas emitidas acerca de la

frecuencia en la realización de 7 acciones constitutivas de plagio (planteadas de manera independiente):

- Acción 1: entregar un trabajo completo, realizado por el propio alumno, que ya había sido entregado en cursos o convocatorias anteriores.
- Acción 2: entregar un trabajo realizado por otro alumno que ya había sido entregado en cursos o convocatorias anteriores (para la misma asignatura o para otra).
- Acción 3: copiar de páginas web fragmentos de textos y –sin citar– pegarlos directamente en un documento –en el cual hay una parte de texto escrita por el propio alumno– y entregarlo como trabajo de una asignatura.
- Acción 4: bajar o descargar un trabajo completo de internet y entregarlo, sin modificarlo, como trabajo propio de una asignatura.
- Acción 5: copiar fragmentos de fuentes impresas (libros, enciclopedias, periódicos, artículos de revista, etc.) y añadirlos –sin citar– como partes de un trabajo propio de una asignatura.
- Acción 6: hacer íntegramente un trabajo a partir de fragmentos copiados literalmente de webs (sin que ninguna parte del trabajo haya sido realmente escrita por el alumno/a).
- Acción 7: copiar partes de trabajos entregados durante cursos anteriores y, sin citar, usarlos como apartados de un trabajo nuevo.

Para cada acción, los participantes debían señalar la frecuencia con que habían efectuado este tipo de prácticas a lo largo del curso académico anterior al momento de realización del cuestionario. Podían escoger una de estas cinco posibilidades: “Nunca”, “Entre 1 y 2 veces”, “Entre 3 y 5 veces”, “Entre 6 y 10 veces” o “Más de 10 veces”. Las preguntas de la V3 se han recodificado para conseguir el índice de comisión de plagio académico (en adelante, ICPA) a partir del sumatorio de respuestas de cada alumno estableciéndose una escala comprendida entre 7 y 35 (donde 7 significa no haber plagiado y 35 representa haber plagiado muy frecuentemente).

### 2.3. Análisis de los datos

Para cada variable de categoría analizada se ha calculado la frecuencia y el porcentaje, mientras que en las variables de escala se ha calculado también la media, varianza y desviación típica. Seguidamente, con el objetivo de establecer potenciales asociaciones entre el ICPA y los índices IRP e ICCRB, se ha proyectado el análisis estadístico a partir de la comparación de medias, obtenida mediante la aplicación

del análisis de la varianza anova. El análisis de la significación estadística en la asociación entre el género y los índices ICPA, IRP e ICCRB se ha efectuado mediante la prueba t-student para muestras independientes.

Para el conjunto de análisis llevados a cabo en el presente estudio se ha utilizado el paquete estadístico *Statistical Package for the Social Sciences (SPSS, v. 19.0)*.

## 3. Resultados

### 3.1. Reconocimiento de plagio

Se recuerda al lector que los participantes debían leer tres fragmentos derivados de un texto original, extraído literalmente de *Wikipedia*, y determinar cuál de ellos era un plagio, cuál era la mejor opción para entregar como trabajo académico y cuál la peor opción. Los resultados obtenidos sugieren que algo más de tres cuartas partes (77,8%) del alumnado participante es capaz de determinar cuál es el texto plagiado: texto “B” (tabla 1).

El 72,3% acertadamente responde que el mejor fragmento para presentar en un trabajo académico es el “C”.

El 68% es capaz de determinar, con acierto, que el peor fragmento para entregar como trabajo académico es el “B”.

La recodificación y posterior sumatorio de los datos obtenidos en esta pregunta –explicado en la metodología– posibilitan generar un IRP que arroja los siguientes resultados:

- casi la mitad de la muestra (49,2%) acertó en las tres preguntas: IRP alto;
- un 27,1% respondió correctamente dos preguntas: IRP medio;
- un 14,5% respondió correctamente una de las tres preguntas: IRP bajo;
- un 8,6% no contestó acertadamente ninguna de las preguntas: IRP Nulo.

Se ha realizado la prueba t-student para analizar la relación estadística entre el IRP y el género, y los resultados obtenidos demuestran la inexistencia de dicha asociación (significación bilateral = 0,487).

### 3.2. Conocimiento de las citas y referencias bibliográficas

El 64,9% del alumnado encuestado fue capaz de determinar que la definición aportada correspondía a una cita académica (tabla 2). El 91,5% supo determinar que la información que se le presentaba, en otra pregunta, era una referencia

Tabla 1. Reconocimiento de plagio en textos académicos

¿Cuál de los tres fragmentos representa un plagio del original? (%)		¿Cuál es el mejor fragmento para entregar como parte de un trabajo académico? (%)		¿Cuál es el peor fragmento para entregar como parte de un trabajo académico? (%)	
Texto A (Parafraseado sin cita)	7,7	Texto A (Parafraseado sin cita)	20,0	Texto A (Parafraseado sin cita)	17,2
Texto B (Plagio literal)	77,8	Texto B (Plagio literal)	6,5	Texto B (Plagio literal)	68,0
Texto C (Texto correctamente citado)	6,5	Texto C (Texto correctamente citado)	72,3	Texto C (Texto correctamente citado)	8,9
No lo sé	8,0	No lo sé	1,2	No lo sé	5,8

Tabla 2. Reconocimiento de citas y referencias bibliográficas en textos académicos

¿A qué concepto corresponde esta definición: “incorporación dentro del texto de conceptos, palabras o frases de un autor determinado identificando el linaje de este y el año de publicación para que cualquier lector pueda localizar la fuente utilizada”? (%)		Si en un trabajo académico encuentras: <b>Novo, M. (1988). La educación ambiental: bases éticas y metodológicas. Ginebra: Unesco. ¿Qué es? (%)</b>	
Referencia bibliográfica	23,8	Nota a pie de página	2,0
Cita	64,9	Referencia bibliográfica	91,5
Nota a pie de página	8,6	Paráfrasis bibliográfica	3,3
Bibliografía	2,0	Cita	0,0
No lo sé	0,7	No lo sé	3,3

bibliográfica redactada con arreglo a la normativa APA (tabla 2).

En cuanto al ICCRB:

- el 60,9% de la muestra contestó acertadamente las dos preguntas: ICCRB alto;
- el 34,4% acertó en una pregunta: ICCRB bajo;
- el 4,6% erró en las dos preguntas: ICCRB nulo.

En cuanto a la relación entre el ICCRB y el género, la prueba t-student demuestra la inexistencia de una relación estadísticamente significativa entre ambas variables (sig. bilateral = 0,326).

### 3.3. Comisión de plagio académico

Las prácticas más extendidas (tabla 3) son las calificadas en la bibliografía como “plagio collage” (Comas-Forgas, 2009), esto es, componer un trabajo académico a partir de la copia de diversos fragmentos de texto, bien sea de fuentes digitales bien sea de fuentes impresas, y su inserción en un documento en el que hay alguna parte de aportación propia, sin citar su origen. Entre las acciones menos repetidas, sobresale:

- descargar un trabajo completo de internet y entregarlo como propio;
- presentar un trabajo elaborado únicamente a partir de fragmentos copiados de internet;
- presentar un trabajo elaborado y previamente entregado en cursos o convocatorias anteriores.

Ahondando en la explotación de resultados, se ha efectuado el sumatorio de respuestas de las siete acciones analizadas con el fin de generar el ICPA. El resultado de la

Tabla 3. Frecuencia de respuesta para las 7 acciones constitutivas de plagio académico

Acción	Nunca (%)	1 a 2 veces (%)	3 a 5 veces (%)	6 a 10 veces (%)	Más de 10 veces (%)
1	80,7	17,7	1,6		
2	76,9	21,2	1,2	0,3	0,3
3	37,0	32,3	22,4	5,0	3,4
4	86,8	12,3	0,3	0,3	0,3
5	43,6	31,2	20,6	2,8	1,9
6	86,3	7,8	3,4	1,9	0,6
7	51,6	34,3	11,6	1,9	0,6

Tabla 4. Valores obtenidos mediante la prueba t para muestras independientes para el análisis de relaciones entre el ICPA y el género

	Género	N	$\bar{X}$	t	gl	Significación (bilateral)
ICPA	Mujer	162	10,18	1,420	311	0,157
	Hombre	151	10,72			

N=nº de alumnos de la muestra;  $\bar{X}$  =media; t=valor obtenido; gl=grados de libertad.

agrupación genera una media de respuesta de 10,46 con una desviación típica de 3,38 (siendo el valor mínimo 7 y el máximo 26).

### 3.4. Asociación entre el índice de comisión de plagio académico y el género

Los hombres presentan medias de perpetración más altas que las mujeres, aunque no se constata una relación estadísticamente significativa (tabla 4).

### 3.5. Asociación entre el índice de comisión de plagio académico y el índice de reconocimiento de plagio

Los alumnos con un bajo IRP son los que tienen un ICPA más alto, lo cual podría resultar esperable. Ahora bien, el segundo grupo con un ICPA mayor es el que conforman aquellos que han obtenido un IRP más alto; es decir, aquellos alumnos que saben reconocer lo que supone un texto plagiado presentan tasas de plagio bastante elevadas, si lo comparamos con el resto (tabla 5). Sea como sea, no existe una relación significativamente estadística entre ambas variables ya que el análisis de la varianza realizado (anova de un factor) arroja una significación de 0,186.

Tabla 5. Medias de ICPA asociadas a los niveles de IRP

IRP	N	$\bar{X}$ ICPA
Nulo	26	11,53
Bajo	43	9,95
Medio	88	10,12
Alto	156	10,62

$\bar{X}$  ICPA= media del ICPA

Tabla 6. Medias de ICPA asociadas a los niveles de ICCRB

ICCRB	N	$\bar{x}$ ICPA
Nulo	17	11,85
Bajo	104	10,20
Alto	185	9,91

$\bar{x}$  ICPA= media del ICPA

### 3.6. Asociación entre el índice de comisión de plagio académico y el índice de conocimiento de las citas y referencias bibliográficas

El alumnado que tiene un ICCRB más alto presenta unos niveles de ICPA más bajos y a la inversa. Los datos parecen tener bastante lógica: a mejor conocimiento de las citas y referencias, menor propensión a plagiar. Aun así, los resultados obtenidos no son estadísticamente significativos dado que el análisis realizado de la varianza (anova de un factor) arroja una significación de 0,337.

Los alumnos que saben reconocer lo que es un plagio presentan tasas de plagio bastante elevadas, si lo comparamos con el resto

## 4. Limitaciones, discusión y conclusiones

El trabajo aquí presentado adolece de ciertas limitaciones que los autores consideran necesario publicitar: por un lado, la muestra está compuesta exclusivamente por alumnado de dos titulaciones de un área determinada en su último curso de carrera y este hecho invalida la posibilidad de generalizar los resultados obtenidos a un espectro mayor de la población universitaria española. Aun así, las características de la muestra -que sean alumnos a punto de graduarse y salir al mercado laboral y/o continuar con una formación de postgrado- aporta datos altamente significativos acerca de qué competencias adquiere el alumnado de administración y dirección de empresas (GADE) y del grado de economía (GECO) en la universidad -en lo tocante a la elaboración de trabajos académicos desde una perspectiva ética y con arreglo a las normas de comunicación científica-, y da pistas de cuán preparados están para su entrada en el mercado laboral -o su continuidad en el ámbito académico- y desarrollarse profesionalmente de manera íntegra y honesta.

En segundo lugar, el hecho de basar el estudio en métodos de encuesta puede provocar cierto sesgo en los datos obtenidos por factores de deseabilidad social (tendencia a atribuirse cualidades y comportamientos socialmente deseables y rechazar los socialmente no deseables). Los autores son conscientes de que el fenómeno del plagio puede abordarse con otras técnicas, una de las cuales, implica el uso de programas de detección para calibrar con mayor precisión la extensión del fenómeno. Se ha descartado dicha opción porqué los programas que presentan unos niveles de fiabilidad aceptables, aunque sin ser infalibles (Weber-Wulff *et al.*, 2013), son de pago y el coste de la licencia para la muestra con que se trabajó, junto al hecho de ser ésta una

aproximación incipiente que se pretende ampliar y mejorar en el futuro, hacían inviable esta posibilidad.

Los resultados alcanzados por este trabajo ponen de manifiesto que una parte importante del alumnado de último curso de las carreras del área de Economía tiene serias dificultades para reconocer lo que supone un plagio en la composición escrita. Es sorprendente que, después de haber cursado al menos 7 semestres en la universidad, casi la mitad de la muestra con la que se trabajó no fuese capaz de contestar acertadamente las tres preguntas referidas al reconocimiento del plagio académico. Algo se debe estar haciendo mal desde las instituciones académicas, sobre todo, cuando estos datos son altamente parejos con los resultados obtenidos por estudios internacionales de análogas características (Roig, 1997; Moon, 2002; Guedes; Gomes-Filho, 2015).

En la misma línea, aunque con unos resultados un tanto más positivos que los anteriores, van las apreciaciones dirigidas a comentar el resultado del reconocimiento de las citas y referencias. Que 4 de cada 10 alumnos de último curso encuestados haya errado en al menos una de las dos preguntas de reconocimiento de las citas y las referencias bibliográficas da bastante qué pensar y lleva a los autores a cuestionarse cómo habrán hecho los trabajos académicos durante su trayectoria en la universidad y en qué disposición están estos alumnos para afrontar la elaboración del trabajo fin de grado (TFG) y, en un futuro muy próximo, su labor profesional y/o continuidad en otros estudios superiores.

Una parte importante del alumnado de último curso de las carreras del área de Economía tiene serias dificultades para reconocer lo que supone un plagio en la composición escrita

Los datos relativos a la comisión de plagio académico ponen de manifiesto que esta práctica está ampliamente extendida en las aulas universitarias: sólo el 22,4% de la muestra afirmó no haber cometido plagio durante el curso académico anterior a la administración del cuestionario. Esta cifra es muy parecida a la obtenida por trabajos nacionales centrados en alumnado universitario (Comas-Forgas, 2009) y también entre alumnado de educación secundaria obligatoria y bachillerato (Sureda-Negre; Comas-Forgas; Oliver-Trobat, 2015). El fenómeno del plagio académico no nace en la universidad, pero en ella tampoco parece que se corrija. Es interesante, en este punto, destacar el hecho que existe un 6,4% de los participantes que podemos calificar como de "plagiadores accidentales", siguiendo la clasificación de Park (2003). Se trata de alumnos que son incapaces de reconocer el plagio y que a la vez cometen acciones plagiarias. Este dato, sumado al 22,4% -porcentaje de "no plagiadores"- supone que el resto (71,2%) cometió al menos una acción constitutiva de plagio teniendo cierta o plena conciencia de lo que estaba haciendo. Son datos ciertamente alarmantes.

De manera general, a la luz del presente estudio, no se demuestra que exista una relación entre la capacidad para reconocer cuándo un texto está plagiado y la comisión de este tipo de falta. De la misma manera, no se puede concluir que

el hecho de saber reconocer qué es una cita y una referencia bibliográfica represente que se plagie menos. Ello invita a pensar que el conocimiento de lo que es y no es plagio y de las normas de citación no es suficiente para atajar el problema aunque a buen seguro ayuda.

Es imprescindible que desde las instituciones educativas se luche contra todo tipo de conducta deshonestas que vaya contra los valores de la probidad y la honradez. Los centros educativos y universidades deben incluir entre sus objetivos luchar contra esta plaga (Hexham, 2013) y formar a ciudadanos íntegros y honestos. No se puede expresar de forma más clara de lo que lo hizo Moreno (1999):

La escuela es el primer 'campo de prácticas' del fraude y la corrupción; el primer escenario público en el que se adquiere y desarrolla una serie de conductas y 'habilidades' que más adelante se utilizarán en las relaciones con otras instituciones sociales, a saber, el mercado laboral, la Seguridad Social, la Hacienda Pública, etc. Y, además, la escuela también es el primer espacio en el que se aprende a disculpar, justificar y hasta a legitimar tales conductas y habilidades" (p. 72).

## Bibliografía

- Christakis, Erika; Christakis, Nicholas** (2012). "Harvard cheating scandal: Is academic dishonesty on the rise". *Time ideas*, 4 Sept. <http://ideas.time.com/2012/09/04/harvard-cheating-scandal-is-academic-dishonesty-on-the-rise>
- Cizek, Gregory** (1999). *Cheating on tests: How to do it, detect it, and prevent it*. Routledge. ISBN: 978 0805831443
- Comas-Forgas, Rubén** (2009). *El ciberplagio y otras formas de deshonestidad académica entre el alumnado universitario*. Tesis doctoral (no publicada). Palma Mallorca, España: Universidad de las Islas Baleares.
- Comas-Forgas, Rubén; Sureda-Negre, Jaume; Salva-Mut, Francesca** (2010). "Academic plagiarism prevalence among Spanish undergraduate students: an exploratory analysis". *Biochemia medica*, v. 20, n. 3, pp. 301-306. <http://dx.doi.org/10.11613/bm.2010.038>
- DeGeeter, Michelle; Harris, Kira; Kehr, Heather; Ford, Carolyn; Lane, Daniel; Nuzum, Donald; Compton, Cinthya; Gibson, Withney** (2014). "Pharmacy students' ability to identify plagiarism after an educational intervention". *American journal of pharmaceutical education*, v. 78, n. 2. <http://dx.doi.org/10.5688/ajpe78233>
- Domínguez-Aroca, María-Isabel** (2012). "Lucha contra el plagio desde las bibliotecas universitarias". *El profesional de la información*, v. 21, n. 5, pp. 498-503. <http://dx.doi.org/10.3145/epi.2012.sep.08>
- Guedes, Diego-Oliveira; Gomes-Filho, Douglas-Leonardo** (2015). "Percepção de plágio acadêmico entre estudantes do curso de odontologia". *Revista bioética*, v. 23, n. 1, pp. 139-148. <http://dx.doi.org/10.1590/1983-80422015231054>
- Hallak, Jacques; Poisson, Muriel** (2002). *Ethics and corruption in education*. Paris: Unesco. <http://www.unesco.org/iiep/PDF/Forum15.pdf>
- Hallak, Jacques; Poisson, Muriel** (2007). "Fraude académico, acreditación y garantía de calidad: lecciones aprendidas del pasado y retos para el futuro". En: Sanyal, Bikas; Tres, Joaquín (ed.). *La educación superior en el mundo - 2007. Acreditación para la garantía de la calidad: ¿Qué está en juego?* España: Mundi-Prensa. ISBN: 84 8476 303 X <http://goo.gl/LMcpUJ>
- Hexham, Irving** (2013). *The plague of plagiarism: Academic plagiarism defined*. Irving Hexham (ResearchGate profile). <https://goo.gl/vXODSy>
- Humphreys, Louise** (2013). "The internet as a challenge to HE: Implications for plagiarism & information literacy" (poster). *Innovative practice in higher education*, v. 1, n. 3. <http://journals.staffs.ac.uk/index.php/ipihe/article/viewFile/40/85>
- Hurlbert, Janet-McNeil; Savidge, Cathleen; Laudenslager, Georgia** (2003). "Process-based assignments: How promoting information literacy prevents plagiarism". *College & undergraduate libraries*, v. 10, n. 1, pp. 39-51.
- Moon, Young-in** (2002). "Korean university students' awareness of plagiarism in summary writings". *Language research*, v. 38, n. 4, pp. 1349-1365. <http://s-space.snu.ac.kr/bitstream/10371/90765/1/12.%202225589.pdf>
- Moreno, Juan-Manuel** (1999). "Con trampa y con cartón. El fraude en la educación, o cómo la corrupción también se aprende". *Cuadernos de pedagogía*, n. 283, pp. 71-77. [http://cedoc.infod.edu.ar/upload/Con\\_trampa.....PDF](http://cedoc.infod.edu.ar/upload/Con_trampa.....PDF)
- Morey, Mercè** (2011). *Anàlisi de l'alfabetització informacional entre l'alumnat de la Universitat de les Illes Balears*. Tesis doctoral (no publicada). Palma Mallorca, España: Universidad de las Islas Baleares.
- Park, Chris** (2003). "In other (people's) words: Plagiarism by university students - literature and lessons". *Assessment & evaluation in higher education*, v. 28, n. 5, pp. 471-488. [http://www.lancaster.ac.uk/staff/gyaccp/caeh\\_28\\_5\\_02lores.pdf](http://www.lancaster.ac.uk/staff/gyaccp/caeh_28_5_02lores.pdf) <http://dx.doi.org/10.1080/02602930301677>
- Poisson, Muriel** (2010). *Corruption and education*. Paris: International Institute for Educational Planning, Unesco. ISBN: 978 92 803 1342
- Roig, Miguel** (1997). "Can undergraduate students determine whether text has been plagiarised?". *Psychological record*, v. 47, n. 1, pp. 113-122. <http://goo.gl/vFMNvH>
- Sureda-Negre, Jaume; Comas-Forgas, Rubén; Oliver-Trobat, Miquel-Francesc** (2015). "Plagio académico entre alumnado de secundaria y bachillerato: diferencias en cuanto al género y la procrastinación". *Comunicar*, v. 22, n. 44, pp. 103-111. <http://dx.doi.org/10.3916/c44-2015-11>
- Weber-Wulff, Deborah; Möller, Christopher; Touras, Janis; Zinke, Elin** (2013). "Plagiarism detection software test (2013)". *Abgerufen am*, n. 12. <http://plagiat.htw-berlin.de/wp-content/uploads/Testbericht-2013-color.pdf>

# ANÁLISIS

## TÉCNICAS *BIG DATA*: ANÁLISIS DE TEXTOS A GRAN ESCALA PARA LA INVESTIGACIÓN CIENTÍFICA Y PERIODÍSTICA

Big data techniques: Large-scale text analysis for scientific and journalistic research

Carlos Arcila-Calderón, Eduar Barbosa-Caro y Francisco Cabezuelo-Lorenzo



**Carlos Arcila-Calderón** es profesor de la *Universidad de Salamanca* (España), miembro del *Observatorio de Contenidos Audiovisuales (OCA)* y editor de la revista *Disertaciones*. Es doctor europeo en comunicación, cambio social y desarrollo por la *Universidad Complutense de Madrid* y *Máster en Periodismo* por la *Universidad Rey Juan Carlos*. Con anterioridad desarrolló su carrera en la *Universidad del Rosario* y en la *Universidad del Norte* de Colombia, y en la *Universidad de Los Andes* y *Universidad Católica Andrés Bello* de Venezuela.  
<http://orcid.org/0000-0002-2636-2849>

*Universidad de Salamanca, Facultad de Ciencias Sociales  
Campus Miguel de Unamuno, Edificio FES  
Av. Francisco Tomás y Valiente, s/n. 37071 Salamanca, España  
carcila@usal.es*



**Eduar Barbosa-Caro** es periodista e investigador universitario. Cuenta con un *Máster en Comunicación* de la *Universidad del Norte* (Colombia). Forma parte del equipo *Colciencias* para el *Grupo de Investigación en Comunicación y Cultura*. Es el editor adjunto del *Anuario Electrónico de Estudios en Comunicación Social Disertaciones* de la *Universidad del Rosario* (Colombia).  
<http://orcid.org/0000-0001-9009-5581>

*Universidad del Norte  
Vía Puerto de Colombia, km 5. Barranquilla, Colombia  
eduarbarbosacc@gmail.com*



**Francisco Cabezuelo-Lorenzo** es profesor en el campus de Segovia de la *Universidad de Valladolid*. Es licenciado en periodismo por la *Universidad Complutense de Madrid (UCM)* y licenciado en publicidad y relaciones públicas por la *Universidad Camilo José Cela (UCJC)*. Está acreditado como profesor titular de universidad y cuenta con un sexenio (2007-2012) reconocido por la *Cneai*. Ha participado en varios programas competitivos de I+D+i de convocatorias autonómicas, estatales y europeas.  
<http://orcid.org/0000-0002-9380-3552>

*Universidad de Valladolid, Facultad de Ciencias Sociales, Jurídicas y de la Comunicación  
Plaza de la Universidad, 1. 40005 Segovia, España  
cabezuelo@hmca.uva.es*

### Resumen

Este trabajo conceptualiza el término *big data* y describe su importancia en el campo de la investigación científica en ciencias sociales y en las prácticas periodísticas. Se explican técnicas de análisis de datos textuales a gran escala como el análisis automatizado de contenidos, la minería de datos (*data mining*), el aprendizaje automatizado (*machine learning*), el modelamiento de temas (*topic modeling*) y el análisis de sentimientos (*sentiment analysis*), que pueden servir para la generación de conocimiento en ciencias sociales y de noticias en periodismo. Se expone cuál es la infraestructura necesaria para el análisis de *big data* a través del despliegue de centros de cómputo distribuido y se valora el uso de las principales herramientas para la obtención de información a través de software comerciales y de paquetes de programación como *Python* o *R*.

Artículo recibido el 30-03-2016  
Aceptación definitiva: 28-06-2016

## Palabras clave

Datos; *Big data*; Minería de datos; Aprendizaje automático; Modelamiento de temas; Análisis de sentimientos.

## Abstract

This paper conceptualizes the term big data and describes its relevance in social research and journalistic practices. We explain large-scale text analysis techniques such as automated content analysis, data mining, machine learning, topic modeling, and sentiment analysis, which may help scientific discovery in social sciences and news production in journalism. We explain the required e-infrastructure for big data analysis with the use of cloud computing and we assess the use of the main packages and libraries for information retrieval and analysis in commercial software and programming languages such as *Python* or *R*.

## Keywords

Data; Big data; Data mining; Machine learning; Topic modeling; Sentiment analysis.

**Arcila-Calderón, Carlos; Barbosa-Caro, Eduar; Cabezuelo-Lorenzo, Francisco** (2016). "Técnicas *big data*: análisis de textos a gran escala para la investigación científica y periodística". *El profesional de la información*, v. 25, n. 4, pp. 623-631.

<http://dx.doi.org/10.3145/epi.2016.jul.12>

## 1. Introducción

Existe un creciente interés tanto científico como periodístico por la explotación de las grandes cantidades de datos textuales disponibles en internet gracias al uso masivo de los denominados medios sociales (*Facebook*, *Twitter*, blogs, etc.) y de otras fuentes textuales de información (medios de comunicación online, webs oficiales, libros electrónicos, documentos financieros, etc.). Un buen ejemplo de este interés es el caso conocido como *Panama papers* o papeles de Panamá, en el que se han usado técnicas de ciencia de datos (Woodie, 2016) para revelar a la opinión pública fraude fiscal y financiero por parte de personajes importantes (jefes de estado, empresarios, políticos, etc.).

Sin embargo, tanto en los campos de las humanidades digitales, la comunicación e información (Verbeke et al., 2014), como del periodismo de datos, existe poca claridad y consenso sobre el concepto de *big data* y sobre las técnicas para análisis de textos a gran escala. Este artículo tiene como objetivo sintetizar los principales enfoques existentes sobre *big data* y describir los principales métodos computacionales que científicos sociales y periodistas tienen a su disposición para la explotación y el análisis de información.

Big data se refiere a volúmenes masivos y complejos de información estructurada y no estructurada que requiere de métodos computacionales para extraer conocimiento

A pesar de que se ha intentado vincular el concepto de *big data* sólo con el tamaño de los datos, en términos de *terabytes* o *petabytes* (por ejemplo, en los papeles de Panamá se usaron 2,6 terabytes), esta dimensión es insuficiente para caracterizarlo. El concepto de *big data* se refiere fundamentalmente a volúmenes masivos y complejos de información tanto estructurada como no estructurada, que es recogida durante cierto período de tiempo y que requiere de métodos computacionales para extraer conocimiento.

Otros conceptos importantes ligados al estudio de los *big data* también aluden a su intencionalidad y utilidad (Murphy; Barton; 2014). El objetivo principal en la generación de datos no contempla generalmente la posibilidad de ser combinados con otros, pues cuando se reúnen grandes cantidades de datos para una finalidad específica, éstos suelen perderse en un mar de información sin pensar en usos secundarios. Por ello se suele sacar el mayor provecho de los datos recogidos sólo a partir de su reutilización básica, su fusión interna y el hallazgo de combinaciones dos por uno (Mayer-Schönberger; Cukier, 2013), en donde hasta los desechos digitales pueden ser objeto de estudio. Es decir, se realiza una explotación al máximo de los recursos recogidos, pero luego no se suelen reutilizar.

Existen tres retos asociados al fenómeno *big data* (Nunan; Di-Domenico, 2013) que científicos sociales y periodistas de datos deben tener en cuenta:

- problemas tecnológicos asociados al almacenamiento, seguridad y análisis de los siempre crecientes volúmenes de datos;
- valor comercial que puede ser añadido a través de la generación de *insights* más efectivos;
- impactos sociales, particularmente las implicaciones para la privacidad personal.

Desde un punto de vista académico, estos retos están vinculados a su vez a tres cambios de paradigma:

- mayor importancia de la disponibilidad y acceso de los datos;
- aceptación de niveles de imprecisión y desorden en los datos;
- centrarse más en las correlaciones, en vez de buscar constantemente la causalidad (Mayer-Schönberger; Cukier, 2013).

Estos cambios, junto con las conceptualizaciones mencionadas, demuestran la inmensa potencialidad que tiene el trabajo con grandes cantidades de datos, pero también dejan ver los problemas tanto técnicos como conceptuales que aún quedan por resolver.

## 2. Métodos computacionales para el análisis de *big data*

Una vez recogida una importante cantidad de datos textuales (estructurados, semi-estructurados o sin estructura) por medio de procedimientos que van desde la recolección manual (texto por texto) y su digitalización, hasta los más automatizados (como *web scrapping*), y construida una base de datos (relacional, no relacional u orientada a grafos), son necesarios métodos computacionales para realizar un análisis de datos y obtener cierto conocimiento o al menos información relevante y novedosa para la sociedad.

En el caso citado de los papeles de Panamá, se utilizó reconocimiento óptico de caracteres (OCR, por sus siglas en inglés: *optical character recognition*) para la digitalización de 11,5 millones de documentos que contenían el registro de cuatro décadas de negocios de la firma *Mossack Fonseca*. Para realizar búsquedas flexibles a gran escala entre estos documentos no estructurados se utilizó *Apache Solr*, a través de una interface más amigable para los periodistas conocida como *Blacklight Project*. Estos documentos fueron estructurados después en un esquema de relaciones (tipo nodo-arista) para crear una base de datos orientada a grafos usando la tecnología *Neo4j*, lo que finalmente permitió hacer uso de técnicas de análisis de *big data* para encontrar las relaciones entre individuos y datos financieros que destaparon el escándalo mundial.

Las técnicas de ciencia de datos se pueden aplicar, como en el ejemplo anterior, a datos previamente recogidos y trabajados, pero también se puede realizar análisis a gran escala con datos en tiempo real (en *streaming*), lo cual amplía las posibilidades de los científicos.

A continuación se describen y analizan algunas de las técnicas y programas más representativos utilizados en el análisis de grandes cantidades de datos textuales a través de una breve explicación conceptual-metodológica de cada uno.

### 2.1. Análisis automatizado de contenido

El análisis de contenido es un ejercicio analítico cuyo objetivo es obtener información de cierto conjunto de datos, generalmente textos o grabaciones (Leetaru, 2012; Krippendorff, 2004). Históricamente, el análisis de contenido se ha servido de otras técnicas que mejoran su alcance y se ha venido aplicando en marcos de investigación cuantitativos, cualitativos y mixtos, mientras “emplea un amplio rango de técnicas analíticas para generar descubrimientos y ponerlos en contexto” (White; Marsh, 2006, p. 22).

A través de los métodos computacionales y del análisis de contenido automatizado (ACA) se vencen limitaciones que tenían los análisis de contenido tradicionales. Además de mayores muestras y mejor codificación, la confiabilidad alcanzada a través de la tecnología disminuye notablemente los sesgos que puedan desviar la interpretación, con lo que podemos replicar los estudios de manera más acertada y a distintas escalas.

No obstante, estas consideraciones generan argumentos a favor y en contra sobre la fiabilidad y validez de dichos análisis computarizados. Harwood y Garry (2003) recalcan que al no cumplirse estándares de validez, la generalización de

los resultados puede quedar en tela de juicio, mientras que desde otras perspectivas la búsqueda instantánea entre los datos y el incremento en la amplitud de los estudios se configuran como variables que justifican y dan valor al uso de las nuevas tecnologías (West, 2001).

Este tipo de análisis automatizados impulsan trabajos de investigación cada vez más diversos. Por ejemplo, Cheng *et al.* (2008, p. 2) destacan el incremento en el uso de “la lingüística computacional [...] aplicada a dominios como los de la captura de datos de inteligencia, traducción con máquinas, análisis de contenido automatizado y la indexación y recuperación de bases de datos completas”.

Estas técnicas son fundamentales para estudiar todo tipo de datos a nuestro alcance, incluyendo los miles de millones de mensajes publicados por medios digitales y redes sociales. Específicamente, se han realizado algunas aplicaciones que permiten la aplicación del análisis de contenido automatizado de forma rápida e intuitiva, entre ellas las más difundidas son *Linguistic inquiry and word count (LIWC)*, *Hamlet*, *WordStat* y *QDAMiner*. Sin embargo existen cada vez más softwares para el análisis del lenguaje natural para entornos de programación como *Python* o *R*. En el caso de *Python*, destaca una serie de librerías bajo el nombre *NLTK*, y en *R* encontramos la librería *ReadMe*.

En función del tamaño de la información y de las necesidades de los científicos de datos, se pueden generar algoritmos específicos (basados en librerías o con funciones nativas) ejecutados sobre entornos de programación, en especial *Python*. Estos algoritmos o scripts se suelen desarrollar también para análisis a gran escala mediante computación distribuida (ordenadores conectados usualmente en la nube) en entornos como *Hadoop*, *Flink* o *Spark*, requiriendo también programar en otros lenguajes como *Java* o *Scala*. Todas estas tecnologías utilizan la filosofía *Map-Reduce* para distribuir las tareas de análisis en diferentes nodos (*Map*) y luego juntar los resultados en un único archivo (*Reduce*). Además de las *grids* académicas, existen grandes compañías comerciales que proporcionan servicios de computación en la nube (eludiendo muchos problemas técnicos para el usuario final) como *Amazon Web Services (AWS)*, *Oracle Cloud Computing* o *Microsoft Azure*.

Un ejemplo de un análisis de contenido automatizado a gran escala es el conteo típico de palabras o frecuencias de aparición de un término en conjuntos de datos que se encuentran almacenados y que no pueden ser procesados por los programas comerciales (ejemplo: un *dataset* de 100 GB). Para hacer este conteo se debe escribir un pequeño script *Map-Reduce*, por ejemplo en *Python*, que *tokenice* cada palabra por medio de la fórmula “clave, valor”, es decir, “palabra\_X, 1”. Para ejecutar este script se debe desplegar un cluster de instancias (nodos *master* y *esclavos*) conectadas en la nube en donde se debe subir el archivo con los datos (en un formato también distribuido como *HDFS*, *Hadoop distributed file system*) y el algoritmo que permitirá el análisis. Este centro de cómputo en la nube, usando por ejemplo *Hadoop*, permitiría distribuir las tareas de análisis de forma equilibrada entre los nodos, paralelizando el análisis que en un solo ordenador hubiese sido imposible. Cuando son completadas las tareas de los nodos, se realiza

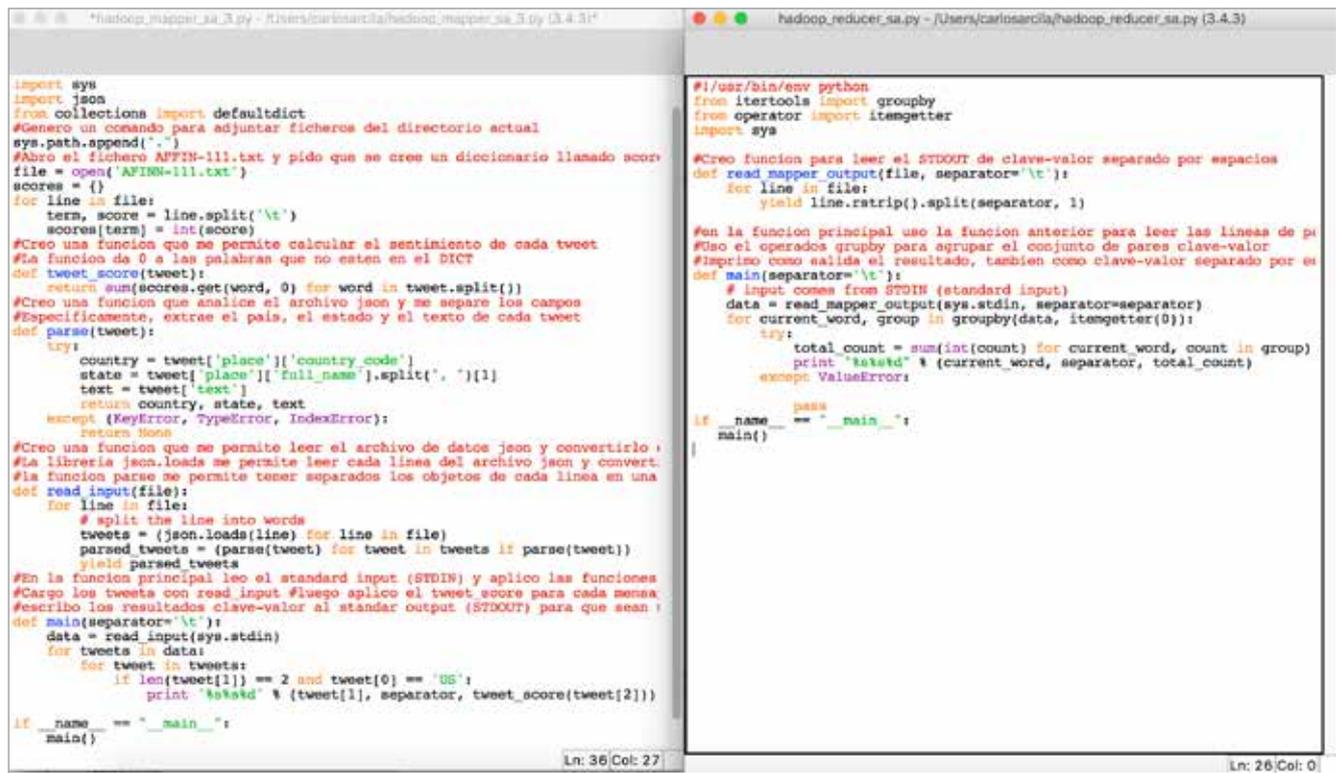


Figura 1. Algoritmo en Python para el análisis de sentimientos a textos a gran escala, utilizando MapReduce para ser desplegado sobre Hadoop

un proceso de resumen de los datos en forma “clave, valor” generados, mediante procedimientos sencillos de suma (SUM) o agrupamiento (GROUP BY). En el caso anterior obtendríamos por ejemplo: “palabra\_X, 35”, indicando que palabra\_X tuvo una frecuencia de 35 apariciones.

### 2.2. Análisis de sentimiento automatizado

Una de las técnicas aplicadas a grandes cantidades de datos y de mayor interés para científicos sociales y periodistas es probablemente el *sentiment analysis* o análisis de sentimiento. Su objetivo se centra en analizar el vocabulario de un texto con el fin de determinar sus cargas emocionales, haciendo uso de un ordenador que a través de *lexicons* procese, reconozca y evalúe dichos sentimientos (Leetaru, 2012), y así saber si los mensajes contienen emociones positivas, negativas o neutras en su estructura (Feldman, 2013). *Opinion mining* y *sentiment analysis* son dos cuestiones distintas (Kechaou; Ben Ammar; Alimi, 2013). *Opinion mining* se dirige a la detección de la polaridad, y *sentiment analysis* al reconocimiento de emociones, pero debido a que la identificación de sentimientos es a menudo explotada para la detección de la polaridad, los dos campos se suelen utilizar como sinónimos (Cambria et al., 2013).

Destacan estudios tradicionales como el de Turney (2002), que aplica el análisis de sentimiento a reseñas (*reviews*) para clasificarlas en recomendadas o no recomendadas; o trabajos como los de Meena y Prabhakar (2007) que se centran en extraer sentimiento de frases u oraciones. Otros como Cai et al. (2010) llegan incluso a combinar las técnicas de *sentiment analysis* y *topic modeling* para extraer resultados más concretos sobre estas características (sentimiento y tema) y sus relaciones.

Existen múltiples fuentes de datos a los cuales se puede apli-

car análisis de sentimiento, entre las que destacan los blogs, sitios especializados en reseñas, conjuntos de datos ya diseñados y sitios de *microblogging* como Twitter (Vinodhini; Chandrasekaran, 2012). Éste último se ha convertido en el principal reto de los científicos sociales y periodistas, debido a la enorme cantidad de información semi-estructurada en formato *JSON* (*javascript object notation*) que es posible obtener tanto del *streaming* (flujo en directo) como del archivo histórico a partir del uso de las APIs (*Steaming* y *REST*) que Twitter ofrece a sus usuarios de forma gratuita. En el mercado existe un gran número de programas comerciales para el análisis de sentimiento (*MeaningCloud*, *Semantria*, *WordStat*, etc.), aunque la mayoría sólo permite análisis de pequeñas cantidades de datos en servidores remotos u ordenadores locales.

“ El *sentiment analysis* analiza el vocabulario de un texto para determinar sus cargas emocionales ”

Un ejemplo de análisis de sentimiento a gran escala es el despliegue de un cluster en *Spark* cuya fuente de datos sea el *streaming* de Twitter y que permita monitorizar en tiempo real el tono de los mensajes que se están emitiendo con una etiqueta o *hashtag* (como #*AtentadosParís*), clasificando estos mensajes por zona geográfica para determinar el impacto en varias partes del mundo. Para ello se puede hacer uso de diccionarios de sentimientos, entre los que se encuentran el *Afinn-111* que está disponible tanto en inglés como en castellano y otros idiomas, y que da una valoración a cada palabra (“*love* = +3”; “*war* = -2”). El algoritmo de calificación funciona *tokenizando* palabras con la estructura “clave: valor” descrita en los párrafos anteriores, pero en vez de contar la aparición

de una palabra, el script debe asignar a cada mensaje un valor a partir de la suma aritmética de los sentimientos detectados (+3 -2 = +1). Esto permite realizar posteriores operaciones (distribuidas usando *MapReduce*) de agrupación (como países con sentimientos más negativos hacia el *hashtag*) y de cruces estadísticos más sofisticados.

### 2.3. Data mining

Implica la extracción de conocimiento a partir de datos masivos y las relaciones subyacentes que pueden existir entre ellos. El *data mining* se originó en 1990 a medida que la tecnología relacional de bases de datos maduró y los procesos de negocio crecieron en automatización (Dhar, 2013, p. 67), fomentando la creación de software orientado a aprovechar los datos sobre comportamiento y transacciones, para predecir y planear de manera más acertada. Siguiendo la línea de Han, Kamber y Pei (2006), el *knowledge discovery from data* (término que se ha usado a la par de *data mining*) se puede dividir en siete fases:

- limpieza de datos
- integración de los datos
- selección de datos
- transformación de los datos
- minería de datos
- evaluación de patrones
- presentación del conocimiento.

Desde esta perspectiva, se identifica al *data mining* como sólo uno de varios momentos, si bien de suma importancia, para el conocimiento a partir de los datos, lo que no resta trascendencia a su posición como un instrumento de análisis eficiente de grandes datos. Para Hand, Mannila y Smyth (2001, p. 6) en la minería de datos podemos encontrar datos

observacionales que se relacionan con el hecho de que

“la minería de datos típicamente trata con datos que ya han sido recopilados para algún propósito distinto al del análisis de minería de datos”.

Kalina (2013) estima que no se debería concebir la extrapolación de los descubrimientos particulares como finalidad primordial, pues cada conjunto de información habla de ese corpus en particular. Pero debemos tener cuidado y evitar pensar que este tipo de técnicas reemplazan totalmente nuestra labor como investigadores.

La minería de datos extrae conocimiento a partir de datos masivos y de las relaciones subyacentes que pueden existir entre ellos

Si bien los científicos sociales y los periodistas de datos hacen uso frecuente de software estadístico comercial y de acceso libre (como *Statistical Package for the Social Sciences*, *SPSS*, y su versión libre *PSPP*), para el análisis de datos a gran escala la mayoría de estos paquetes son insuficientes. Existen productos comerciales como *MatLab* que son escalables (permite su ejecución en clusters y nubes), sin embargo, desde la limpieza de datos hasta la visualización final los científicos de datos prefieren la utilización de lenguajes de software libre como *R* o *Python*.

En *R*, el lenguaje de programación estadístico en abierto más extendido, existen cientos de funciones nativas (sin necesidad de librerías adicionales) que permiten el análisis y minería de datos. Se han creado librerías específicas que facilitan y agilizan la minería de datos como *car*, *Hmisc*, *ggplot2*, *dplyr* o *tidyr*. In-

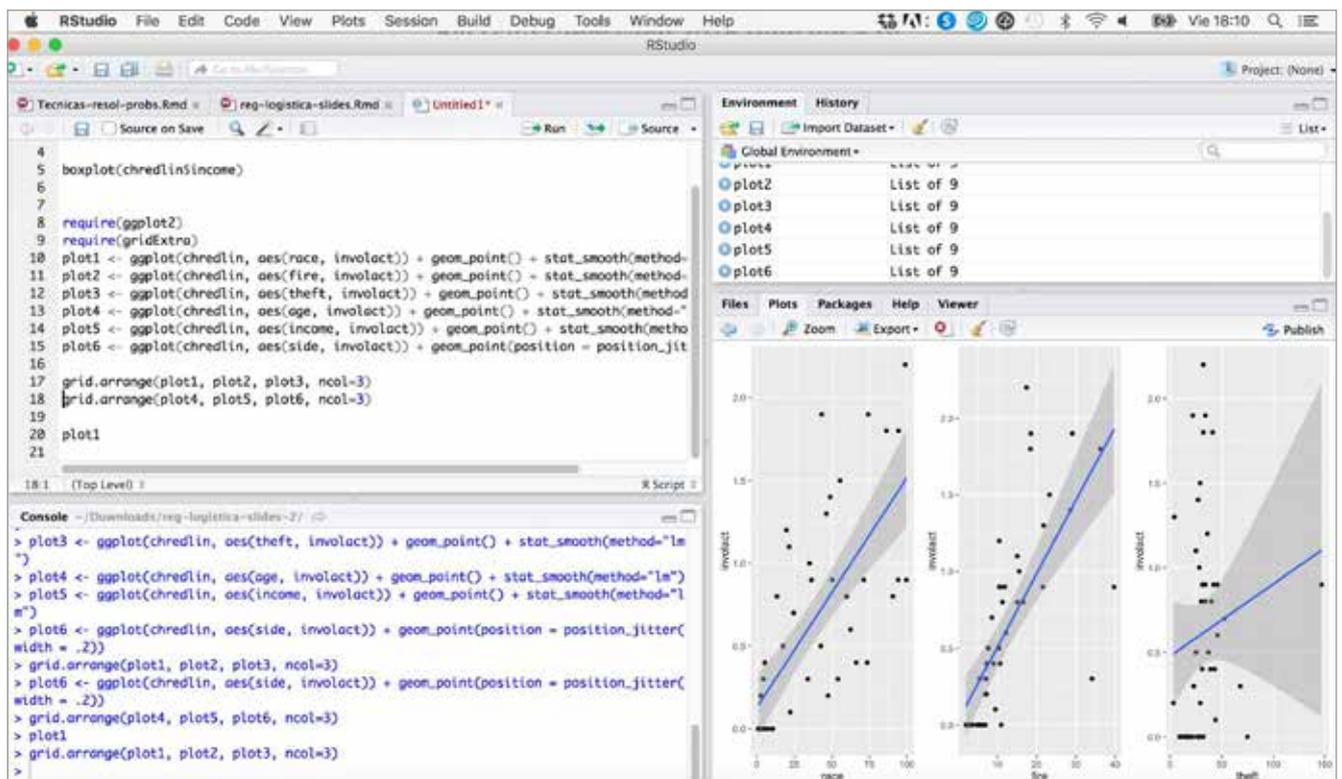


Figura 2. Minería y visualización de datos en *R*, usando *R Studio*

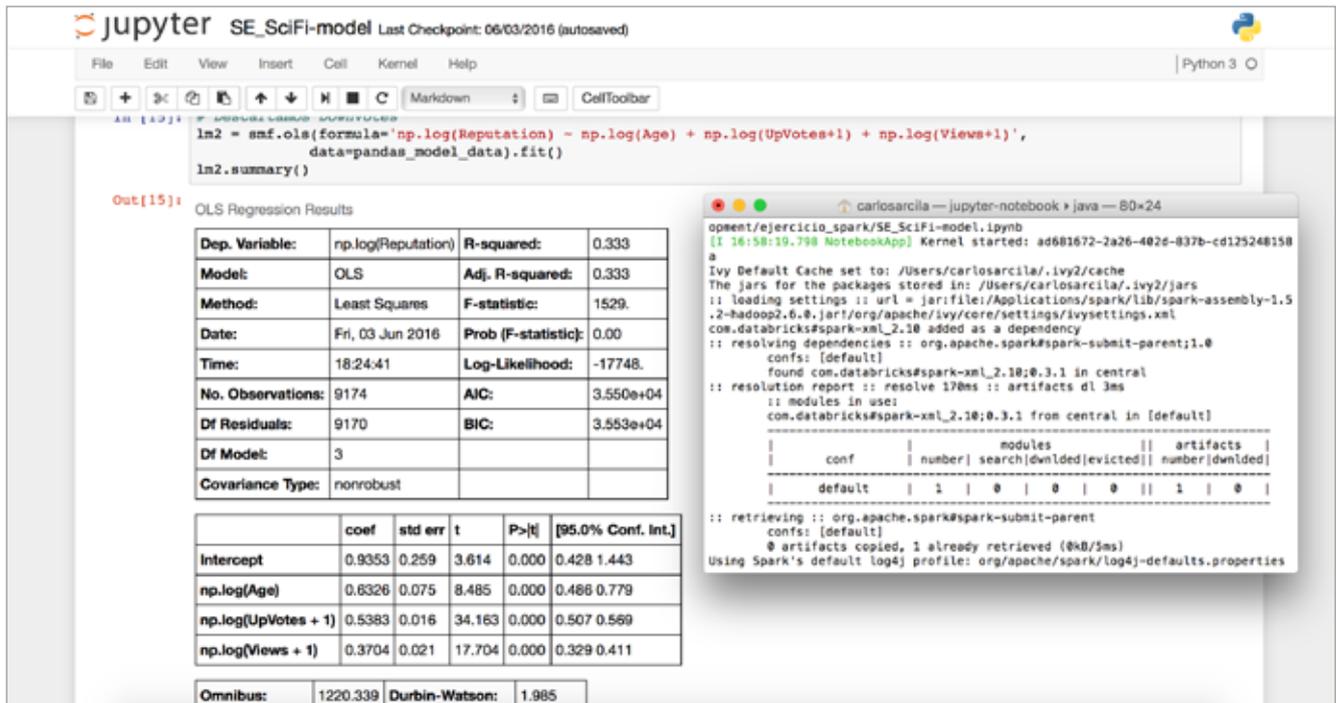


Figura 3. Aplicación de regresión lineal de forma distribuida y en streaming con Spark, utilizando Jupyter como iNotebook de Python

cluso se incluyen librerías que proveen la visualización de grafos y el análisis de redes como *igraph*. Una de las limitaciones de la minería de datos de R es la paralelización de procesos (el cómputo distribuido en diferentes ordenadores), aunque en las últimas versiones de la plataforma Spark se incluye un módulo (aún limitado) de R que permite paralelizar los análisis.

Por lo anterior, gran parte de la computación científica requerida para la minería de datos a gran escala se sigue diseñando en scripts para Java, Scala o Python. Este último es probablemente el más usado entre los científicos y periodistas de datos, ya que es un lenguaje de programación interpretado, lo que en cierta medida facilita la sintaxis y la ejecución de las funciones. Además existe un sinnúmero de librerías en Python que facilitan y maximizan funciones típicas y avanzadas de minerías de datos, entre las que destacan Pandas, Numpy, Matplotlib y SciPy. Python se entiende además en general bastante bien con los principales desarrollos de computación distribuida, y se ha convertido en un estándar dentro de muchas comunidades científicas.

### 2.4. Machine learning

Es un concepto derivado de la propia minería de datos que se refiere al diseño de programas o algoritmos que pueden aprender reglas a partir de datos, adaptarse a cambios y mejorar el rendimiento con la experiencia (Blum, 2003). Como campo multidisciplinar en donde confluye la estadística y la complejidad computacional (Mitchell, 1997), esta técnica reduce tiempos y costos. También obtiene resultados fiables a través del aprendizaje que realiza la máquina al agregársele parámetros y configuraciones específicas para cada estudio. Un ejemplo de machine learning es la clasificación automática de correo electrónico. Para su funcionamiento tenemos en primera instancia unos recursos de texto ya clasificados (ejemplo: correo spam vs. correo no spam) que son cargados al sistema de análisis (datos de entrenamien-

to o training), ya sea a través de una interfaz o una línea de códigos. Este sistema permite entonces generar un conocimiento basado en el corpus introducido previamente, lo que se convierte en un algoritmo con el cual la máquina aprende las reglas subyacentes en dichos documentos (ejemplo: aparición de términos como “lotería”, etc.). Tras este paso, estas reglas o patrones son ingresados nuevamente en el sistema de análisis y usados sobre una nueva muestra también clasificada (muestra de prueba o testing) para mejorar progresivamente los resultados y su precisión, forjando un análisis cada vez más robusto.

El machine learning, usado fundamentalmente para la clasificación y la predicción, se ha aplicado en áreas tan diferentes como las búsquedas en internet y el diseño de medicamentos (Domingos, 2012), además de en situaciones puntuales que estimulan el uso de esta técnica, entre las que se puede mencionar, además del volumen de los datos: falta de expertos para resolver un problema a partir de datos, imposibilidad de exponer claramente las reglas de análisis de datos, alta velocidad con que cambia un conjunto de datos, y labores de personalización de grandes conjuntos de información (Dietterich, 2003). Este último caso ha sido utilizado para situaciones en las que no existe un algoritmo único, como por ejemplo uno que permita diferenciar automáticamente correos electrónicos no deseados de los legítimos (Alpaydin, 2010), tal como se explica en el párrafo anterior.

Se puede dividir el machine learning en dos grandes grupos (Murphy, 2012):

- aprendizaje supervisado o predictivo, en donde la máquina aprende no sólo de los propios datos finales (inputs) sino que es posible darle modelos o datos adicionales ya categorizados (outputs) para que el aprendizaje sea mucho más fiable;
- aprendizaje no supervisado o descriptivo en el que sólo se

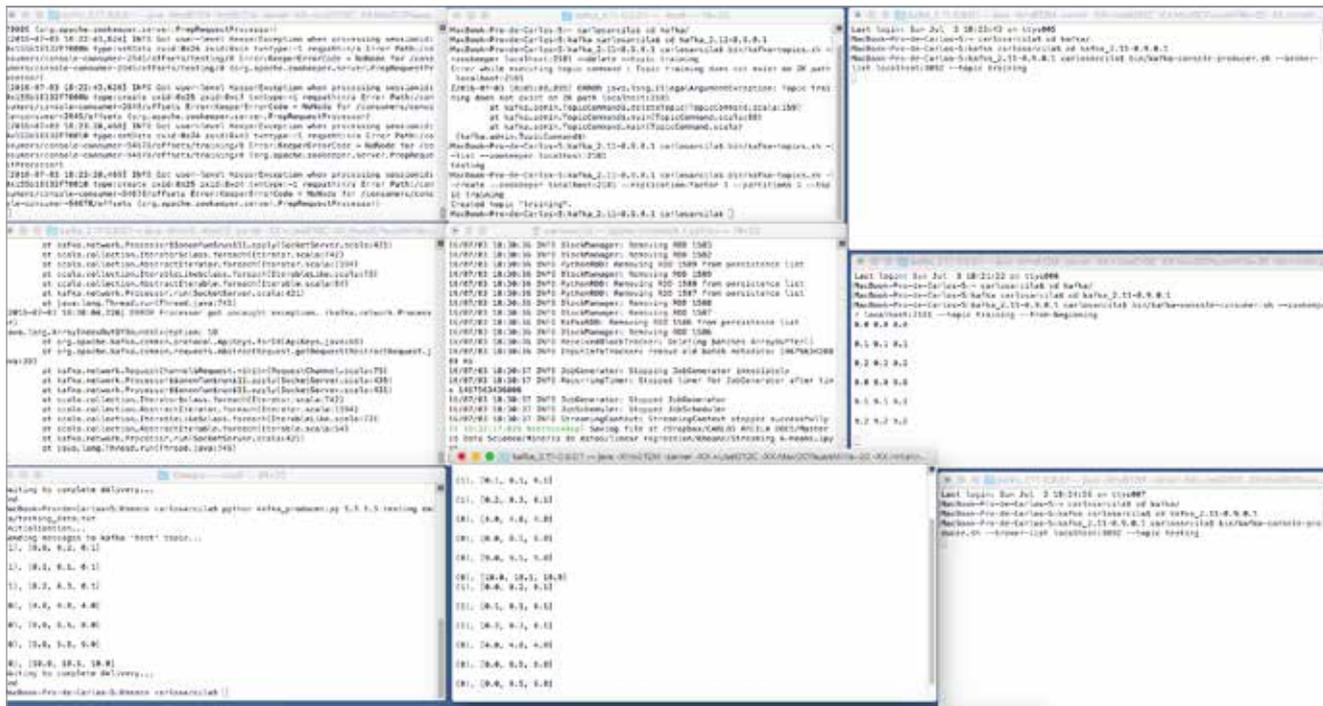


Figura 4. Aplicación del algoritmo K-Means para creación de clusters y predicción usando datos en streaming (flujo en tiempo real), utilizando Spark, Kafka y Zookeeper

dan los *inputs* a la máquina para que encuentre patrones interesantes a partir de los datos.

Desde el punto de vista del algoritmo utilizado (Kelleher; MacNamee; D’Arcy, 2015), las técnicas de *machine learning* se pueden dividir en aprendizaje basado en:

- información: árboles de decisión con algoritmos como *ID3*, métodos de ensamblado como *boosting* o *bagging*, y bosques aleatorios o *random forests*;
- similitud: análisis de cluster no jerárquico con el algoritmo *K-Means*, análisis de cluster jerárquico con extensiones *kernel* usando máquinas de vectores soporte (SVM, por sus siglas en inglés: *support vector machines*) y redes neuronales;
- probabilidad: modelo *Naive Bayes*;
- error: regresión lineal múltiple con método del gradiente.

En la investigación científica de medios podemos mencionar estudios que utilizan algunos de estos algoritmos como los de Pennacchiotti y Popescu (2011), quienes trabajan en las inmediaciones de los *social media* y el *machine learning* para detectar atributos como la inclinación política, la etnia o afinidades de negocio, o el trabajo de Téllez-Valero, Montes y Villaseñor-Pineda (2009) que aporta conocimientos metodológicos para recopilar y analizar datos sobre reportes noticiosos a través del *machine learning*.

**2.4.1. Aprendizaje supervisado**

Las aplicaciones de aprendizaje supervisado requieren algoritmos especializados que detecten patrones en los datos. Estos algoritmos pueden implementarse en lenguajes de programación como *Python*, pero al igual que en el análisis de contenido automatizado, si se aplican sobre grandes cantidades de datos requieren plataformas distribuidas para el procesamiento en paralelo. Para superar las dificultades que implica el desarrollo de código y el despliegue de cen-

tros de cómputo en la nube, ha prosperado una serie de servicios comerciales que permiten el aprendizaje automático de manera mucho más sencilla. Entre los más extendidos y de relativa facilidad de uso para científicos sociales y periodistas, encontramos la plataforma AWS que incluye un módulo llamado *Amazon Machine Learning (AML)* que incorpora tanto asistentes como software de visualización, o los servicios de la empresa *Databrick* que basan sus servicios de computación en la nube exclusivamente en *Spark*.

Para el aprendizaje supervisado, modelos como el de predicción basado en máquinas de vectores soporte (SVM) que pasan nuestros datos a un espacio multidimensional, permiten crear algoritmos potentes a partir de datos existentes (un ejemplo: conjuntos de noticias ya separadas por tema) para crear patrones que permitan categorizar automáticamente nuevos conjuntos de textos. Esto es muy útil para la clasificación automática de noticias. Lo mismo sucede para la generación de predicciones de comportamiento de la opinión pública a partir del uso de perfiles creados con encuestas históricas, respondiendo preguntas como cuál es la probabilidad de que un conjunto de ciudadanos que votan al partido X aprueben o desapruében un tema de agenda emergente.

**2.4.2. Aprendizaje no supervisado y topic modeling**

A diferencia del aprendizaje supervisado, el no supervisado utiliza procedimientos inductivos, extrayendo conocimiento sólo de los datos, como en el caso del análisis de clusters para clasificación. Una de las aplicaciones específicas más útiles para científicos sociales y periodistas es el modelado de temas o *topic modeling*, que comprende la extracción de temáticas a partir de cuerpos de documentos cuya envergadura vence nuestras competencias para obtener manualmente temas, relaciones temporales y patrones en la clasificación (Arora et al., 2013). Esta meto-

dología parte de los mismos datos para obtener los temas (nombrados a posteriori por el investigador) en los que luego serán agrupados los documentos (o colecciones de éstos). Para llevar a cabo esta tarea se seleccionan automáticamente palabras del corpus que aparecen frecuentemente, lo que indicaría que podrían pertenecer o no a cierto tema y, observando su presencia en los documentos, podemos buscarlos y clasificarlos sin intervención humana. Ello diversifica el uso que se puede dar a esta técnica y, por ende, los resultados que arroja. Se usa un diccionario de “lista de parada” o *stop list*, cuya función es decirle a un algoritmo qué palabras no deben ser tomadas en cuenta para la creación de clusters.

Existen varios paquetes informáticos para el modelamiento de temas. *Mallet* es uno de los más difundidos, puede descargarse gratuitamente y permite la clasificación de documentos, el *clustering* y la extracción de información (McCallum, 2002). Uno de los modelos más simples de *topic modeling* es el *Latent Dirichlet Allocation (LDA)*, que cuenta con dos principios fundamentales:

- patrones implícitos, y
- conjuntos de términos que podríamos llamar *temas* (Blei, 2012).

También existen interfaces como *Stanford topic modeling toolbox* para realizar modelado de temas que, a diferencia de *Mallet*, proveen de un entorno gráfico más amigable para llevar a cabo los procesos sin necesidad de conocer la materia en su totalidad. Estos programas, sin embargo, no son escalables, por lo que para analizar datos a gran escala se debe desarrollar código o contratar servicios comerciales, como los de *AWS* o *Microsoft Azure*.

### 3. Conclusiones

Las grandes cantidades de datos fluyen a través de nuevos canales constituyéndolos en una fuente valiosa de información. Esto da lugar a nuevos retos para las ciencias sociales y el periodismo en lo que a capacidad de procesamiento y análisis se refiere. Si bien es cierto que esta imbricación entre los métodos computacionales y otras disciplinas supone cambios en el quehacer científico, los *big data* y las herramientas relacionadas invitan a repensar las lógicas de investigación social y del propio periodismo desde una perspectiva más amplia, donde se desdibujan aún más los límites entre los campos de estudio y de obtención de información. Las nuevas lógicas implican la necesidad de construir equipos interdisciplinarios y centros de análisis de *big data* en las universidades y centros de investigación, que faciliten el desarrollo de proyectos de investigación para explotar el enorme potencial de análisis de estas fuentes para las ciencias sociales y el periodismo.

A partir de los conceptos, teorías y metodologías que se han revisado en este texto, se observa que se necesita mayor profundización (tanto en la teoría como en la práctica), para hacer este campo más accesible (menos requerimientos técnicos y/o facilidad de uso de las plataformas de cómputo distribuido) y que pueda aportar mayor conocimiento en ciencias sociales y en la investigación periodística.

### 4. Bibliografía

**Alpaydin, Ethem** (2010). *Introduction to machine learning*. Cambridge/London: The MIT Press. ISBN 978 0262012430

**Arora, Sanjeev; Ge, Rong; Halpern, Yoni; Mimno, David; Moitra, Ankur; Sontag, David; Wu, Yichen; Zhu, Michael** (2013). “A practical algorithm for topic modeling with provable guarantees”. En: *30<sup>th</sup> Intl conf on machine learning*. pp. 280-288.  
<http://jmlr.org/proceedings/papers/v28/arora13.html>

**Blei, David M.** (2012). “Topic modeling and digital Humanities”. *Journal of digital humanities*, v. 2, n. 1, pp. 8-11.  
<http://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei>

**Blum, Avrim** (2003). “Machine learning theory”. En: *FOCS 2003 Procs of the 44<sup>th</sup> Annual IEEE Symposium on foundations of computer science*. Washington DC: IEEE Computer Society, pp. 2-4. ISBN: 0 7695 2040 5

**Cai, Keke; Spangler, Scott; Chen, Ying; Zhang, Li** (2010). “Leveraging sentiment analysis for topic detection”. En: *IEEE/WIC/ACM International Conference on Web Intelligence and Agent Systems: An International Journal*, pp. 265-271.  
<http://www.csce.uark.edu/~sgauch/5013NLP/S13/hw/Chris.pdf>  
<http://dx.doi.org/10.1109/WIIAT.2008.188>

**Cambria, Erick; Schuller, Björn; Liu, Bing; Wang, Haixun; Havasi, Catherine** (2013). “Knowledge-based approaches to concept-level sentiment analysis”. *IEEE intelligent systems*, v. 28, n. 2, pp. 12-14.  
<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6547971>  
<http://dx.doi.org/10.1109/MIS.2013.45>

**Cheng, An-Shou; Fleischmann, Kenneth; Wang, Ping; Oard, Douglas** (2008). “Advancing social science research by applying computational linguistics”. En: *Procs of the American Society for Information Science and Technology*, v. 45, n. 1, pp. 1-12.  
[http://www.asis.org/Conferences/AM08/proceedings/posters/55\\_poster.pdf](http://www.asis.org/Conferences/AM08/proceedings/posters/55_poster.pdf)

**Dhar, Vasant** (2013). “Data science and prediction”. *Communications of the ACM*, v. 56, n. 12, pp. 64-73.  
<https://archive.nyu.edu/bitstream/2451/31553/2/Dhar-DataScience.pdf>  
<http://dx.doi.org/10.1145/2500499>

**Dietterich, Thomas** (2003). “Machine learning”. *Nature encyclopedia of cognitive science*. London: Macmillan.  
<http://eecs.oregonstate.edu/~tgd/publications/nature-ecs-machine-learning.ps.gz>

**Domingos, Pedro** (2012). “A few useful things to know about machine learning”. *Communications of the ACM*, v. 55, n. 10, pp. 78-87.  
<http://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>  
<http://dx.doi.org/10.1145/2347736.2347755>

**Feldman, Ronen** (2013). “Techniques and applications for sentiment analysis”. *Communications of the ACM*, v. 56, n. 4, pp. 82-89.  
<http://dx.doi.org/10.1145/2436256.2436274>

- Han, Jiawei; Kamber, Micheline; Pei, Jian** (2006). *Data mining. Concepts and techniques*. San Francisco: Morgan Kaufmann Publishers. ISBN: 978 0123814791  
<http://goo.gl/5zTYb6>
- Hand, David; Mannila, Heikki; Smyth, Padhraic** (2001). *Principles of data mining*. Cambridge: MIT Press. ISBN: 978 0262082907  
[ftp://gamma.sbin.org/pub/doc/books/Principles\\_of\\_Data\\_Mining.pdf](ftp://gamma.sbin.org/pub/doc/books/Principles_of_Data_Mining.pdf)
- Harwood, Tracy; Garry, Tony** (2003). "An overview of content analysis". *The marketing review*, v. 3, pp. 479-498.  
<http://dx.doi.org/10.1362/146934703771910080>
- Kalina, Jan** (2013). "Highly robust methods in data mining". *Serbian journal of management*, v. 8, n. 1, pp. 9-24.  
[http://www.sjm06.com/SJM%20ISSN1452-4864/8\\_1\\_2013\\_May\\_1\\_132/8\\_1\\_2013\\_9-24.pdf](http://www.sjm06.com/SJM%20ISSN1452-4864/8_1_2013_May_1_132/8_1_2013_9-24.pdf)  
<http://dx.doi.org/10.5937/sjm8-3226>
- Kechaou, Zied; Ben-Ammar, Mohammed; Alimi, Adel** (2013). "A multi-agent based system for sentiment analysis of user-generated content". *International journal on artificial intelligence tools*, v. 22, n. 2, pp. 1-28.  
<http://dx.doi.org/10.1142/S0218213013500048>
- Kelleher, John D.; MacNamee, Brian; D'Arcy, Aoife** (2015). *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. Londres: MIT Press. ISBN: 978 0262029445
- Krippendorff, Klaus**. (2004). *Content analysis. An introduction to its methodology*. Los Angeles: Sage Publications. ISBN: 978 0761915454
- Leetaru, Kalev-Hannes** (2011). *Data mining methods for the content analyst: An introduction to the computational analysis of informational center*. New York: Routledge. ISBN: 978 0415895149
- Mayer-Schönberger, Viktor; Cukier, Kenneth** (2013). *Big data. La revolución de los datos masivos*. Madrid: Turner. ISBN: 978 8415832102
- McCallum, Andrew-Kachites** (2002). *Mallet: A machine learning for language toolkit*.  
<http://mallet.cs.umass.edu>
- Meena, Arun; Prabhakar, T. V.** (2007). Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis. En: Amati, Giambattista; Carpineto, Claudio; Romano, Giovanni (eds.). *Advances in information retrieval. 29th European conf on IR research (ECIR)*, April 2-5, 2007, Rome, Italy, pp. 573-580.  
[http://dx.doi.org/10.1007/978-3-540-71496-5\\_53](http://dx.doi.org/10.1007/978-3-540-71496-5_53)
- Mitchell, Tom** (1997). *Machine learning*. New York: McGraw-Hill. ISBN: 978 0070428072  
[http://personal.disco.unimib.it/Vanneschi/McGrawHill\\_-\\_Machine\\_Learning\\_-\\_Tom\\_Mitchell.pdf](http://personal.disco.unimib.it/Vanneschi/McGrawHill_-_Machine_Learning_-_Tom_Mitchell.pdf)
- Murphy, Kevin** (2012). *Machine learning. A probabilistic perspective*. Cambridge/London: The MIT Press. ISBN: 978 0262018029
- Murphy, Michael; Barton, John** (2014). "From a sea of data to actionable insights: Big data and what it means for lawyers". *Intellectual property & technology law journal*, v. 26, n. 3, pp. 8-17.  
<http://www.pillsburylaw.com/publications/from-a-sea-of-data-to-actionable-insights>
- Nunan, Dan; Di-Domenico, Maria-Laura** (2013). "Market research and the ethics of big data". *International journal of market research*, v. 55, n. 4, pp. 505-520.  
<http://dx.doi.org/10.2501/IJMR-2013-015>
- Pennacchiotti, Marco; Popescu, Ana-Maria** (2011). "A machine learning approach to Twitter user classification". En: *Procs of the 5th Intl conf on weblogs and social media*. Menlo Park, California: The Association for the Advancement of Artificial Intelligence Press.  
<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/2886/3262>
- Téllez-Valero, Alberto; Montes, Manuel; Villaseñor-Pineda, Luis** (2009). "Using machine learning for extracting information from natural disaster news reports". *Computación y sistemas*, v. 13, n. 1, pp. 33-44.  
<http://www.scielo.org.mx/pdf/cys/v13n1/v13n1a4.pdf>
- Turney, Peter** (2002). "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews". En: *Procs of the 40th Annual meeting of the Association for Computational Linguistics (ACL)*, Philadelphia, July 2002, pp. 417-424.  
<http://www.aclweb.org/anthology/P02-1053.pdf>
- Verbeke, Mathias; Berendt, Bettina; D'Haenens, Leen; Opgenhaffen, Michaël** (2014). "When two disciplines meet, data mining for communication science". En: *64th Annual meeting of International Communication Association (ICA) conf*. Seattle, USA.  
<https://lirias.kuleuven.be/handle/123456789/436424>
- Vinodhini, Gopalakrishnan; Chandrasekaran, Ramaswamy M.** (2012). "Sentiment analysis and opinion mining: A survey". *International journal of advanced research in computer science and software engineering*, v. 2, n. 6, pp. 282-292.  
[http://www.ijarcse.com/docs/papers/June2012/Volume\\_2\\_issue\\_6/V2I600263.pdf](http://www.ijarcse.com/docs/papers/June2012/Volume_2_issue_6/V2I600263.pdf)
- West, Mark** (2001). *Theory, method, and practice in computer content analysis*. Westport, Connecticut: Ablex Publishing. ISBN: 978 1567505030
- White, Marilyn-Domas; Marsh, Emiliy** (2006). "Content analysis: A flexible methodology". *Library trends*, v. 55, n.1, pp. 22-45.  
<https://www.ideals.illinois.edu/bitstream/handle/2142/3670/whitemarch551.pdf?sequence=2>  
<http://dx.doi.org/10.1353/lib.2006.0053>
- Woody, Alex** (2016). "Inside the Panama papers: How cloud analytics made it all possible". *Datanami*, 7 April.  
<http://www.datanami.com/2016/04/07/inside-panama-papers-cloud-analytics-made-possible>

# CÓMO ANALIZAR EL IMPACTO DE LOS DATOS DE INVESTIGACIÓN CON MÉTRICAS: MODELOS Y SERVICIOS

## How to analyze the impact of research data with metrics: Models and services

**Antonia Ferrer-Sapena, Enrique-Alfonso Sánchez-Pérez, Rafael Aleixandre-Benavent y Fernanda Peset**



**Antonia Ferrer-Sapena** es licenciada en geografía e historia contemporánea por la *Universidad de Valencia (UV)*, y doctora en técnicas y métodos de información y documentación por la *Universidad Politécnica de Valencia (UPV)*. Es directora académica del *Master en Gestión de la Información MUGI* y directora de la *Cátedra de Transparencia y Participación de la Conselleria de Transparencia, Responsabilidad Social, Participación y Cooperación*. Es profesora titular de la *UPV* y miembro del *Grupo ThinkEPI: estrategia y prospectiva de la información*. Su interés se centra en la investigación sobre datos abiertos y sus implicaciones en la empresa, la administración y la ciencia que aborda el proyecto de I+D del *Ministerio de Economía y Competitividad Datasea, datos abiertos de investigación, open research data*. Editora en España del repositorio de acceso abierto *e-Lis, E-prints in library & information science*.

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

*Universitat Politècnica de València, Departament de Comunicació Audiovisual, Documentació i Història de l'Art  
Camino de Vera, s/n. 46022 Valencia, España  
anfersa@upv.es*



**Enrique-Alfonso Sánchez-Pérez** es catedrático del *Departamento de Matemática Aplicada* de la *Universidad Politécnica de Valencia (UPV)* y miembro del *Instituto Universitario de Matemática Pura y Aplicada* de la misma universidad. Es licenciado en matemáticas, ciencias químicas, filosofía, y ciencias de la educación, y doctor en matemáticas por la *UPV*. Es profesor de la *Escuela de Ingenieros de Caminos, Canales y Puertos* de la *UPV*. Su investigación se centra en el análisis matemático, particularmente en temas de análisis funcional, como la teoría de operadores, la integración vectorial y la topología. Ha participado en numerosos proyectos de investigación en estos temas y también en colaboración con otros grupos científicos (física y biomedicina, principalmente), aplicando técnicas matemáticas avanzadas de tratamiento de señales. Ha publicado un centenar de artículos de investigación en revistas científicas de prestigio, principalmente de matemáticas.

<http://orcid.org/0000-0001-8854-3154>

*Universitat Politècnica de València, Institut Universitari de Matemàtica Pura i Aplicada  
Camino de Vera, s/n. 46022 Valencia, España  
easancpe@mat.upv.es*



**Rafael Aleixandre-Benavent** es científico titular del *Consejo Superior de Investigaciones Científicas (CSIC)* y catedrático acreditado de biblioteconomía y documentación. Es doctor en medicina, especialista en documentación médica por la *Universitat de València (UV)* y codirector de la *Unidad de Información e Investigación Social y Sanitaria (Uisys)* de la *UV*. Ha coordinado el proyecto *Open-DataScience, Centro de recursos para la preservación y gestión de datos abiertos de investigación (Odasci)* y participa en la *Red española sobre datos de investigación en abierto-Maredata*. Sus principales líneas de trabajo se centran en la evaluación de la investigación y de las publicaciones científicas, y en los estudios sobre el acceso abierto a los datos científicos.

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

*Ingenio (CSIC-UPV). UISYS (Universitat de València - CSIC)  
Plaza Cisneros, 4. 46003 Valencia, España  
rafael.aleixandre@uv.es*



**Fernanda Peset** es profesora titular de la *Universidad Politécnica de Valencia (UPV), Departamento de Comunicación Audiovisual, Documentación e Historia del Arte*. Trabajó en el *Servicio de Información Bibliográfica* de la *Universidad de Valencia* hasta 1999. Doctor por la *Universidad de Murcia* en 2002, es coordinadora del programa de doctorado *Industrias de la comunicación y culturales* de la *UPV*. Es evaluadora para revistas científicas y organismos nacionales de acreditación y evaluación. Su docencia y publicaciones se orientan a la comunicación científica, el acceso abierto y la implantación del protocolo OAI-PMH, normalización de la información, descripción de documentos, sistemas de documentación de museos, y datos de investigación abiertos y enlazados. Participa en proyectos como *IraLIS* y red *Maredata*, así como en el gobierno de *E-LIS* y del *Grupo Ciepi*, el inventario *ODiSEA* o *transparencyscience.es*. Dirige los proyectos de I+D *Datasea* y *Datasea Extended*.  
<http://orcid.org/0000-0003-3706-6532>

*Universitat Politècnica de València, Departament de Comunicació Audiovisual, Documentació i Història de l'Art*  
Camino de Vera, s/n. 46022 Valencia, España  
[mpesetm@upv.es](mailto:mpesetm@upv.es)

## Resumen

Revisión de las vías de publicación de datos de investigación, las métricas de evaluación de la publicación y reutilización de datos, y los servicios existentes para medir la reutilización de los datos de investigación. Los datos abiertos de investigación aún no se encuentran incluidos en los indicadores de evaluación de la actividad científica. Éste es uno de los motivos por lo que los investigadores no los tienen incluidos en sus rutinas de trabajo. Igualmente, sus métricas se encuentran en sus primeros pasos. Es necesario incrementar los estudios para que estas métricas sean válidas para los sistemas de evaluación.

## Palabras clave

Investigación; Datos; Datos de investigación; Métricas; *Data citation index*; Publicación científica; Publicación de datos de investigación.

## Abstract

This article is a summary of the different pathways to publish research data, of metrics for evaluating publishing and reuse, and of the existing services for measuring the reuse of research data. It is not yet possible to accurately analyze the metrics for open research data and this is one of the reasons why, in general, researchers do not include them in the assessment of their work. In order to further the use of these tools, it is necessary to increase the number of studies about these evaluation metrics.

## Keywords

Research; Data; Research data; Metrics; *Data citation index*; Scientific publication; Publishing open research data.

**Ferrer-Sapena, Antonia; Sánchez-Pérez, Enrique-Alfonso; Aleixandre-Benavent, Rafael; Peset, Fernanda (2016).** "Cómo analizar el impacto de los datos de investigación con métricas: modelos y servicios". *El profesional de la información*, v. 25, n. 4, pp. 632-641.

<http://dx.doi.org/10.3145/epi.2016.jul.13>

## Financiación

Este trabajo se ha desarrollado gracias a la financiación parcial del Plan Nacional de I+D+i del *Mineco: Opendatascience* CS02012-39632-C02-02.

## 1. Introducción

En el contexto social actual existe un sentimiento generalizado entre la comunidad científica internacional de la necesidad de compartir los datos de investigación. Son conocidas las ventajas que aporta (**Tenopir et al.**, 2011), pero no siempre es aconsejable dejar en abierto todos los datos que se generan en una investigación. Debemos tener en cuenta, tal y como señala el grupo de trabajo *RDA (Research Data Alliance) (RDA, sin fecha)* que no todos los datos son susceptibles de ser compartidos. Existen áreas de conocimiento donde es mayor el coste de ponerlos en abierto que los beneficios que puede aportar su reutilización. Para la selección de los datos a compartir hay que tener en cuenta la facilidad en la repetición del experimento o prueba para la selección de los datos.

Es rentable hacerlo en aquellas áreas donde se producen datos derivados de grandes volúmenes de experimentación, existen datos derivados de series acumuladas y se emplean modelos generados utilizando el enfoque de sistemas.

En ciencias de la vida, este intercambio de datos es necesario para el beneficio de la comunidad científica, y lo mismo ocurre en campos como genética o cinética enzimática, donde se recomienda la reutilización, ya que en ellos se tiene mayor facilidad para la repetición de los experimentos o pruebas (*RDA, s.f.*). Compartir los datos de investigación ahorra costes e incrementa el valor de la investigación, pero no es una tarea fácil. Requiere un alto esfuerzo que consiste en una inversión en tiempo y recursos por parte de los investigadores, aunque en general el balance entre costes y beneficios es positivo.

El uso de métricas de datos puede crear nuevos incentivos que apoyen su intercambio e incrementar la velocidad en la difusión de la información en numerosas disciplinas. Las métricas ayudan a estudiar cómo se evalúa el resultado de un solo investigador o de un grupo, contribuyendo a una definición más amplia de los logros y la reputación en el ecosistema científico, más allá de la publicación en artículos científicos. El uso de métricas contribuirá a concienciar a las instituciones e infraestructuras para promover la utilización de datos de una manera responsable y que permita la promoción del bien público global (Lin et al., 2014).

Uno de los supuestos más importantes que pueden potenciar el crecimiento de las métricas de datos es que el intercambio de éstos sea una fuente de potencial reconocimiento científico para los investigadores. El cuidado y difusión adecuada de los datos se puede considerar una actividad científica que tenga cabida en la evaluación de la investigación y que sirva por tanto como méritos para la contratación, promoción, obtención de fondos, etc.

« No todos los datos son susceptibles de ser compartidos; hay áreas donde es mayor el coste de ponerlos en abierto que los beneficios de su reutilización »

El reconocimiento de los artículos publicados proviene generalmente del número de citas recibidas. Varios estudios muestran que los trabajos que comparten los datos reciben mayor número de citas que los estudios similares que no lo hacen (Piwowar; Day; Fridsma, 2007; Piwowar; Vision, 2013) y son una forma de fomentar la erudición responsable (Mooney; Newton, 2012). Aunque muchos expertos señalan las ventajas que conllevaría el uso de métricas de datos, su implantación en la evaluación de la actividad científica también puede contribuir a pervertir el sistema, ya que es probable que algunos investigadores prioricen esta práctica para conseguir el máximo número de puntos y lograr el reconocimiento frente a invertir en investigaciones innovadoras (Ball, 2015).

Hasta ahora el uso de indicadores bibliométricos ha sido clave en la evaluación de la actividad científica. El Factor de impacto de los *Journal Citation Reports* de Thomson Reuters, el cómputo de citas del *Science Citation Index* o el Índice h (índice de Hirsch), han sido utilizados por agencias de financiación, de evaluación y universidades para ponderar el resultado académico de instituciones y científicos. La idea básica de la bibliometría es evaluar el impacto que tienen las publicaciones científicas dentro de la comunidad. A pesar de las numerosas críticas recibidas, los indicadores basados en las citas han sido adoptados mayoritariamente en la evaluación de la ciencia (Aleixandre-Benavent; Valde-rrama-Zurián; González-Alcaide, 2007; RDA-WDS, s. f.). Por este motivo se tiende a buscar indicadores similares para valorar el impacto de los datos, lo que evitaría las posibles reticencias de los productores de datos para compartirlos.

Aunque todavía no hay métricas adecuadas para medir el impacto de los datos de investigación, existe una crecien-

te preocupación para que se creen. Uno de los proyectos que ha abordado el desarrollo de estas métricas es *Making data count: A data metrics pilot project* (Lin et al., 2014) en el que han participado *The California Digital Library*, *PLoS*, y *DataONE*. En este proyecto se están estudiando posibles métricas relacionadas con la actividad de los datos, creando mecanismos de testeo automático y un prototipo flexible para la gestión de la vida de los datos (*data life model* – DLM).

## 2. Estándares de reutilización de datos de investigación

Uno de los principales problemas de los investigadores es cómo encontrar los datos que necesitan, ya que aún no se ha elaborado ningún buscador (Lin et al., 2014). En algunos casos los datos se pueden localizar en la Red, mientras que en otros están incluidos en el artículo al que están asociados (Brase; Farquhar, 2011), o depositados en repositorios específicos de datos (Ferrer-Sapena et al., 2016).

Para que los investigadores puedan encontrar de forma exacta y precisa los datos que necesitan, es necesario que éstos vayan acompañados de los metadatos adecuados. Los metadatos deben proporcionar todos los detalles sobre el origen y la manipulación que han experimentado para evitar un uso inadecuado o una interpretación errónea. Además, es necesario utilizar estándares en los formatos que faciliten que el intercambio se produzca de manera eficaz.

Los métodos aplicados para el intercambio y reutilización pueden tener diferentes enfoques. Cada investigación debe disponer de su propia estrategia en función de para quién deban estar disponibles. En general, existen dos formas de hacer posible el intercambio:

- depositando los datos en repositorios, ya sea el de la propia institución o de terceros cuando la organización no dispone de ellos;
- depositándolos en las webs de las revistas.

El intercambio directo de datos entre investigadores suele producirse cuando no existen repositorios específicos en sus áreas temáticas. Asimismo, puede ser adecuado el intercambio entre comunidades cerradas cuando existan condicionantes éticos o de confidencialidad que así lo aconsejen.

« Uno de los principales problemas de los investigadores es cómo encontrar los datos que necesitan, ya que aún no se ha implementado ningún buscador »

El *Biotechnology and Biological Sciences Research Council* (Bbsrc, 2007) indica que en su campo de conocimiento es recomendable guardar los datos de investigación por un período de diez años. Aconsejan que sean compartidos de manera inmediata una vez sean publicados los primeros resultados y, si se liberan previamente, las normas éticas recomiendan que no se publique ningún estudio hasta que los propios autores de los datos hayan publicado algunos resultados de estos datos recogidos. La escala de tiempo para

liberarlos depende de cada comunidad científica. El *Bbsrc* propone los siguientes plazos para liberarlos en función de las áreas temáticas de biotecnología (*Bbsrc*, 2007):

“en el área de cristalografía (*Protein Data Bank*), se ha acordado un plazo máximo de 12 meses desde la publicación del primer documento sobre una estructura; y en los estudios sobre secuenciación (base de datos de secuencias de nucleótidos *EMBL*), los datos pueden ser retenidos hasta la publicación de los resultados, pero no después”.

La reutilización de determinados tipos de datos puede conllevar problemas derivados de la propiedad intelectual y de la comercialización potencial a partir de las ideas patentables generadas con el nuevo conocimiento. Sin embargo, el *Bbsrc* considera que este tipo de negocio no debe demorar el intercambio de datos ni excluirlo. De la misma manera considera que los datos que provengan del estudio de grandes magnitudes deben ser puestos en abierto de manera inmediata a medida que se encuentren disponibles o se publiquen.

### 3. Modelos de publicación de datos

De acuerdo con **Parsons y Fox** (2013), se pueden señalar los siguientes cinco modelos:

#### Modelo basado en la publicación científica

Este enfoque es semejante al de la publicación científica, es decir, su métrica está basada en el recuento de citas. El conjunto de datos debería estar bien descrito y con un cierto nivel de control de la calidad o de revisión por pares.

Aunque es uno de los modelos más maduros, cuenta con numerosas limitaciones, ya que no tiene claramente definido qué se entiende por publicación de datos (**Parsons; Fox**, 2013). Este modelo también se está analizando en el *Bibliometrics Working Group* de la *Research Data Alliance (RDA)* (**Callaghan; Lehnert**, s.f.), donde se trabaja en el fomento de más y mejores citas de datos para incrementar su disponibilidad y calidad. Al igual que en la publicación científica, para que la publicación de datos sea eficaz, los datos deben satisfacer los siguientes criterios:

- persistencia
- longevidad
- sostenibilidad
- calidad.

Otros problemas se relacionan con la granularidad, control de las versiones, y los aspectos legales.

#### Big iron

Este enfoque proviene de la cultura de ingeniería y, por lo general se ocupa de grandes volúmenes de datos que son relativamente homogéneos, bien definidos y dinámicos. Necesita de una gran infraestructura, sofisticada y bien controlada. Involucra a centros de supercomputación y sus interfaces son especializadas. Dispone de estándares de metadatos y suele utilizar estructuras de datos relacionales y jerárquicas con esquemas organizativos. Algunas instituciones que aplican este modelo son la *NASA* y la *Agencia Europea del Espacio (ESA)*.

Tabla 1. Modelos de publicación de datos y principales características.

	<b>Modelo basado en la publicación científica</b>	<b>Big iron</b>	<b>Science support</b>	<b>Map making</b>	<b>Linked data</b>
Analogía	Publicación académica	Producción industrial	Producción artesanal del trabajo	Cartografía	Creación en la WWW
Características de los datos	Pequeño volumen, diversas formas, escalas y temas,	Alto volumen y más homogéneos en la forma	Pequeños y diversos	Geoespaciales, características y atributos	Entidades iguales nombradas de distintas maneras
Modelos organizativos de los datos	Jerárquico o relacional	Jerárquico	Geoespacial, jerárquico y relacional	Geoespacial y relacional	Basados en mapas y datos enlazados
Fin primario	Calidad de los datos, certificación y preservación	Rendimiento y acceso manejable	Síntesis de los datos y reproducibilidad	Visualización basada en mapas e intercomparación	Interoperabilidad e interconexión
Normas	Cita de datos	Formatos de datos, control de versiones	Procesos locales	Sistemas de referencia de coordenadas, transformaciones espaciales	Ontologías
Ejemplos en la ciencia	<i>Pangea</i> , repositorios universitarios <i>Eosdis</i>	<i>Eosdis</i>	<i>LTER</i>	<i>Inspire</i> , <i>Geodata.gov</i>	<i>Integrated Ocean Drilling Program - IODP</i> , <i>MyGrid</i> , <i>Linked Open Government Data</i>
Terminología	Datos de autor, editor, citación de datos	Productor de datos, nivel de procesamiento, versión de lanzamiento	Colector de datos, personal de apoyo	Fuente de datos, capas	Proveedor de datos, nombre, enlace, recurso
Contexto cultural	Bibliotecas y grupos de investigación universitarios (por ejemplo, la <i>NSF</i> )	Ingeniería de sistemas y gestión de proyectos (por ejemplo, la <i>NASA</i> , el <i>Departamento de Defensa</i> )	Basado en el lugar de investigación (por ejemplo, <i>NSF</i> )	Uso y manejo de la tierra (por ejemplo, <i>United States Geological Survey -USGS</i> , e <i>Integrated Ocean Drilling Program - IODP</i> )	Aplicaciones informáticas y comerciales (por ejemplo, <i>National Science Foundation - NSF</i> y <i>Common Information Sharing Environment - CISE W3C</i> )

Fuente: **Parsons y Fox** (2013)

### Science support

Se denomina así la estructura de apoyo operativo que se encuentra en un centro de investigación o en un laboratorio. Dado que se trata de estaciones o centros que realizan investigaciones de larga duración, la gestión de los datos se considera una función necesaria de la infraestructura para el apoyo a la investigación. Este modelo lo aplican organizaciones como la red *Long Term Ecological Research (LTER)*.

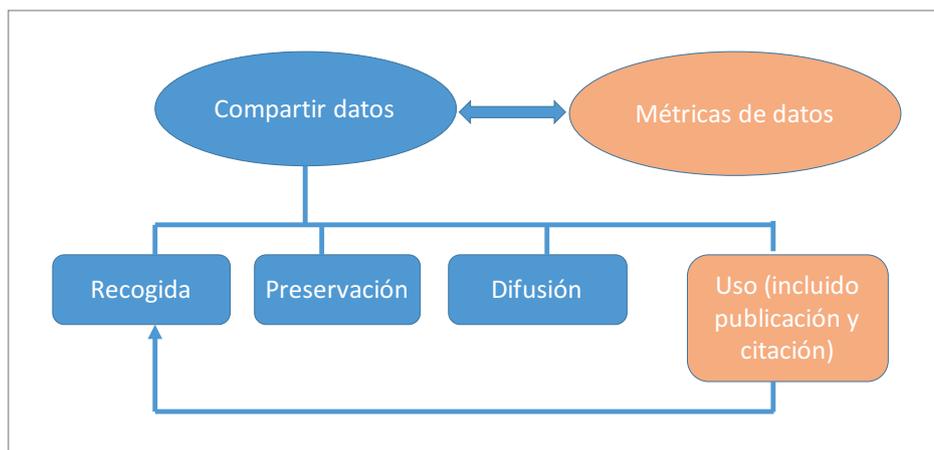


Figura 1. Esquema de las relaciones entre intercambio de datos y sus métricas. Fuente: Costas et al., 2013

### Map making

Son las infraestructuras de datos espaciales y los sistemas de información geográfica (SIG) asociados. La realización de un mapa puede ser visto como un subconjunto de la publicación de datos, pero en este contexto se publica un mapa o atlas.

### Linked data

Se basa en la web de los datos y se fundamenta en el principio del diseño de la web semántica. El objetivo de este enfoque es más la interoperabilidad y la capitalización de la web interconectada que la preservación, la curación o la calidad. Aquí el énfasis se pone en la representación de los datos.

## 4. Situación actual

En la actividad académica se han valorado hasta el momento únicamente los artículos, pero no la calidad de los datos. Los datos pueden ir unidos o no a los artículos pero éstos deben encontrarse debidamente estructurados. Son heterogéneos y varían desde pequeñas unidades, como tablas individuales, a grandes conjuntos de datos. Este hecho dificulta su cita, que puede hacerse del conjunto o de una parte, lo que depende también del campo científico del que se trate.

Es necesario disponer de un marco abierto para la validación de los datos y poder compararlos, así como permitir la implementación de programas que agreguen valor a los datos en bruto. Las métricas deben estar abiertas tanto para los artículos como para los conjuntos de datos y permitir la reutilización en la mayor medida posible.

Existen formas tradicionales para medir el impacto de los datos como *Data Citation Index* de Thomson Reuters. En esta base de datos no se incluye la utilización y reutilización de los conjuntos de datos, sino únicamente su cita, por lo que su alcance tiene limitaciones. También tiene un retraso significativo de las métricas que proporciona debido a que los datos citados dependen de la publicación del artículo.

En el estudio *The value of research data. Metrics for datasets from a cultural and technical point of view* (Costas et al., 2013), en el que se exponen las métricas disponibles, las percepciones de los usuarios sobre las mismas, los problemas existentes y las posibles soluciones, se indica que para identificar claramente las métricas de datos se debe precisar la terminología. Así “métrica de datos” comprende todo lo relacionado con la publicación de los datos y sus citas, incluyendo también algunos indicadores altmétricos. Tanto la publicación como la cita pueden ser consideradas como signos indicativos de uso de los datos. El uso puede generar nuevos datos que pueden alimentarlos de nuevo. En la figura 1 se muestran las relaciones entre el intercambio de datos y sus métricas.

En el proceso de publicación de los datos y sus métricas se encuentran implicados varios agentes. En la figura 2 se pueden ver los más representativos.

Cada agente tiene, por motivos distintos, interés en conocer las métricas de datos:

- las organizaciones financiadoras para conocer el impacto de lo que han financiado;
- las estructuras de investigación para potenciar el acceso y la preservación;
- los investigadores, como una medida más del reconocimiento e impacto de su trabajo;



Figura 2. Agentes implicados en la publicación de datos. Fuente: Costas et al., 2013

- los centros de datos, para estandarizar el almacenamiento y crear metadatos armonizados a nivel mundial, efectuar un seguimiento de la reutilización y promover buenas prácticas, ofrecer datos sobre la curación y recomendar la cita a los conjuntos de datos;
- los editores, para hacer frente a los datos que se adjuntan a los artículos, facilitar políticas y normas para la cita y métrica de los datos;
- las bibliotecas, para hacer los datos accesibles e identificables y establecer una coordinación entre los investigadores y los centros de datos;
- las bases de datos de publicaciones, para unir las publicaciones con las citas de los datos y habilitar indicadores y recuento de citas.

Otra revisión interesante es la guía publicada en 2015 por el *Digital Curation Center (DCC)*, *How to measure the impact of research data* (Ball; Duke, 2015), que quiere servir de referencia para medir el impacto de los datos de investigación de las instituciones. Ofrece una panorámica general de los conceptos clave necesarios para medir el impacto, así como los servicios y programas disponibles.

Como los investigadores únicamente obtienen reconocimiento por los artículos publicados, no invierten tiempo en otras actividades que pueden beneficiar a largo plazo su campo de estudio

Otro grupo que está abordando esta problemática junto al *Bibliometrics Working Group* de *Research Data Alliance* (Callaghan; Lehnert, s.f.) es el *Grupo de Trabajo de Datos del Consortium Advancing Standards in Research Administration Information (Casrai)*, que está trabajando en proyectos de gestión de datos de investigación, abordando los temas de terminología, vocabularios e indicadores. Los grupos de interés detectados, al igual que en otros estudios, son: editores, organismos de financiación, administradores de las universidades y repositorios de datos.

Otros dos estudios a destacar son *Joint declaration of data citation principles* (Martone, 2014) y *Data citation implementation pilot (DCIP)*, s.f.). Los principios de la *Joint declaration of data citation principles* reconocen la doble necesidad de crear prácticas de citación que sean comprensibles para los investigadores y para los ordenadores. No pretenden crear estándares para implementaciones específicas, sino alertar a las comunidades de la necesidad realizar prácticas y aplicaciones que incorporen estos principios. Por su parte, el objetivo principal del *DCIP* es proporcionar coordinación entre editores, repositorios y servicios de metadatos y de identificadores para abordar el problema de los investigadores que están empezando a citar datos, así como guías y grupos de consulta que ayuden a lograr este propósito.

## 5. Modelos de métricas de datos

Como se ha dicho antes, uno de los problemas es que si los investigadores únicamente obtienen reconocimiento y recompensa por los artículos publicados, eso les disuade de invertir tiempo en otras actividades que pueden beneficiar

a largo plazo su campo de estudio. Ésta es una de las razones por las que se están examinando otras métricas para evaluar el impacto de la investigación que vayan más allá del mero recuento de citas.

Los investigadores pueden tener impacto en su disciplina no únicamente por las citas que han recibido sus artículos, sino también por otros reconocimientos a su labor como (Ball; Duke, 2015):

- reutilización de los datos que comparten;
- recomendación de la lectura de sus publicaciones;
- invitaciones a intervenir en los medios de comunicación;
- uso que se hace del software que han escrito.

A la hora de medir el impacto se debe tener en cuenta el alcance de lo que se quiere medir. Las métricas pueden servir para indicar o sugerir el impacto que ha tenido un conjunto de datos o *dataset*, pero esta medida puede que no sea del todo fiable, dado el elevado número de factores que intervienen en la consideración del impacto. Por ello, al igual que ocurre en la interpretación del impacto de las publicaciones, es necesaria la interpretación humana para evaluar la calidad de los *datasets*, ya que no siempre el hecho de que se cite mucho significa que sea bueno, sino que puede ser citado por lo contrario. Así, que un *dataset* sea muy mencionado en los medios sociales puede ser debido a que tiene un impacto positivo o negativo en su disciplina (Ball, 2015).

Existe muy poca bibliografía acerca de métricas de datos, hecho ya señalado por Costas *et al.* (2013), que indican que todavía no existen indicadores de uso. A continuación se presentan posibles métricas basadas en modelos existentes que pueden ser útiles para el desarrollo de las de datos. La tabla 2 muestra las métricas y su relación con el actual modelo de publicación científica.

Como puede observarse, uno de los modelos se basa en la publicación y cita de datos. Aunque es uno de los modelos más maduros (Ball; Duke, 2015), plantea numerosos problemas, sobre todo debidos a las estructuras necesarias para la revisión por pares que garanticen la calidad de los *datasets* y la cita formal de los conjuntos de datos (Lawrence *et al.*, 2011). Hay que recordar que la cita es uno de los principales elementos de reconocimiento académico y es un elemento clave para la identificación, recuperación, replicación y verificación de los datos en los que se basan los estudios publicados (Mooney; Newton, 2012).

Como hemos señalado, un aspecto muy importante en el depósito en repositorios es que se garantice la calidad de los datos depositados. Sin embargo, existen repositorios donde no existe un protocolo de evaluación de la calidad de los datos que incluyen (Brase *et al.*, 2009). Esta evaluación puede llegar a ser muy importante cuando el usuario potencial no tiene la suficiente experiencia y capacidad para evaluarlos. Se ha propuesto la revisión por pares como sistema de control de calidad de los conjuntos de datos (Lawrence *et al.*, 2011); no obstante, esta revisión humana tradicional puede ser apropiada para ciertos conjuntos de datos, pero es demasiado lenta para evaluar la avalancha actual de datos.

Este modelo tiene, como se ha comentado, numerosas limitaciones que todavía no han sido resueltas (Costas *et al.*,

Tabla 2. Escenarios de métricas de modelos de publicación de datos (comparativa con el actual modelo de publicación científica)

Tipos de métricas	Herramientas disponibles con las posibles métricas de datos	Dimensiones de las métricas	Modelos			
			Publicación científica	Publicación de datos	Publicación de datos	
					Publicación de datos solos	Revistas de publicación de datos
Publicación de datos y métricas basadas en citas	<i>Data Citation Index (Web of Science)</i>	Tamaño dependiente	Sí	Difícil	Sí	Sí
	<i>Google Scholar</i>	Tamaño independiente				
	<i>Scopus</i>	Rendimiento promedio directo	Sí	No	Sí	Sí
	<i>Microsoft Academic Search DataCite</i>	Rendimiento basado en la fuente	Sí	No	Sí	Sí
Métricas basadas en altmétricas	<i>ImpactStory Twitter Facebook</i>	Indicadores social media	Sí	Sí	Sí	Sí
	<i>Mendeley CiteULike</i>	Número de lectores	Sí	No	Sí	Sí
	<i>Repositorios Data Journals</i>	Descargas, número de lecturas (Métricas <i>DUI - data usage index</i> )	Sí	Difícil	Sí	Sí

Fuente: **Costas et al.**, 2013

2013). Por un lado el concepto de cita de datos no se encuentra estandarizado y aceptado. Por otro, no se encuentra generalizada la cita directa a los conjuntos de datos, de manera que en algunas citas se describen los conjuntos de datos y sus colecciones sin sacar conclusiones científicas, mientras que en otros casos los datos se publican en una sección especial de una revista o en una revista de datos, como *Journal of open archaeology data* (Ball; Duke, 2015). En muchas disciplinas predomina el enfoque de citar el artículo que hace uso de los datos, confiando en que el artículo indique si se han compartido los datos y cómo se ha hecho. Generalmente no es posible sin un esfuerzo manual considerable, identificar las citas a estos artículos que deben contarse a través de su presencia en el texto. En estas disciplinas por lo tanto, el número de citas es de poca ayuda como indicador del impacto de los datos y por ello deben encontrarse métricas alternativas.

“ El concepto de cita de datos no se encuentra estandarizado y aceptado ”

No existe aún una teoría bibliométrica asentada sobre el valor de la cita de los datos que sea útil para la evaluación. Además, existen importantes limitaciones tecnológicas que restringen el desarrollo de la publicación y cita de datos y por consiguiente sus métricas. Estas limitaciones se deben a la incompatibilidad entre los softwares y los repositorios, las estructuras de los datos, su almacenamiento y su gestión.

Aunque este modelo es el más desarrollado, tal y cómo se ha comentado, no aporta grandes avances en este campo; ni

siquiera han sido implementados por parte de infraestructuras tan conocidas como *DataCite*, aunque se encuentran trabajando en métricas y modelos para el reconocimiento de la publicación y *data sharing*.

Las métricas que se establecen para este modelo son semejantes a las de la bibliometría tradicional y tienen en cuenta el impacto de cada uno de los autores, de las instituciones a las que pertenecen, del proyecto que desarrollan y de la institución que financió la investigación. También se pueden sacar conclusiones sobre el impacto de la entidad frente a sus competidores.

**Ball** (2015) distingue tres tipos de métricas basadas en citas:

- número total de citas de un investigador o de un grupo;
- índice h;
- rendimiento promedio: promedio del número de citas en relación al promedio de la población analizada.

## 6. Altmétricas

De acuerdo con la *NISO* (2016b):

“Altmétricas es un término amplio que incluye la recogida, creación y uso de múltiples formas de evaluación obtenidas a partir de la actividad y la relación entre los diversos actores y la producción científica en el ecosistema de la investigación”.

En las métricas tradicionales no se tiene en cuenta el comportamiento del lector online, la interacción con los contenidos en la Red o las referencias en los medios sociales. Tampoco contemplan las nuevas formas de producción científica, como los conjuntos de datos publicados en repositorios, los algoritmos, las estructuras moleculares o

el software que se comparte en *GitHub*. Estas nuevas producciones son difíciles de evaluar a través de las métricas tradicionales, pues en ellas no existe una cultura de la cita (*NISO*, 2016b).

Los indicadores alométricos pueden aplicarse a los repositorios de datos, ya que tienen en cuenta el número de visualizaciones, descargas, menciones en las redes sociales y otros indicadores que pueden adaptarse bien.

Existe una preocupación acerca de si las técnicas para citar y medir el impacto de los artículos de revistas científicas pueden no ser las más apropiadas para medir el impacto de otros tipos de productos de la investigación, software o datos.

Dentro del apartado destinado a las alométricas, el *Grupo de Trabajo B* de la *NISO* estudia los productos de investigación no tradicionales y sus identificadores. Han publicado recomendaciones dirigidas a instituciones, administradores de repositorios, organizaciones de investigación internacional y agencias financiadoras sobre cómo citar datos y sobre las métricas de datos. Entre otras, se indica que éstas deberían estar disponibles tan pronto como sea posible, deberán estar relacionadas con la cita, usar identificadores persistentes, metadatos y aportar una página de destino (*NISO*, 2016a).

A pesar de las críticas del sistema basado en el cómputo de citas, se han utilizado métricas basadas en él (**Ball**, 2015), como:

- indicadores de las redes sociales: menciones o *likes*;
- contabilización de lectores;
- descargas y visualizaciones, es decir, el número de veces que se accede a un registro.

Por otra parte, las estadísticas de uso no se correlacionan directamente con las citas, ya que reflejan diferentes comportamientos de lectura que pueden no tener nada que ver con la cita a un trabajo de investigación. Se ha observado que el número de lectores de *Mendeley* parece correlacionarse mejor con el número de citas que recibe un trabajo; sin embargo, la correlación no es lo suficientemente fuerte como elemento de predicción de futuras citas.

Entre las principales críticas de las alométricas deben mencionarse las siguientes:

- el concepto no está claramente definido y el término tiene significados distintos según quien lo trate;
- se están convirtiendo en objeto de muchos estudios, pero de momento no en una alternativa;
- se encuentran más bien relacionadas con las métricas de los social media.

A pesar de las críticas, las alométricas se consideran un elemento clave para las métricas de los datos de investigación, aunque aún no se pueden observar en la mayoría de webs de los principales repositorios de datos de investigación.

## 7. Servicios de medida de impacto

Existen muy pocos servicios de medida del impacto de los datos de investigación. Citaremos el *Data Citation Index* y los repositorios de datos.

### **Data Citation Index**

En octubre de 2012, *Thomson Reuters* lanzó el *Data Citation Index (DCI)* como parte de los servicios de *la Web of Science*. El *DCI* recoge las citas de los datos de los artículos indexados por la *Web of Science*. La información que contiene proviene de los repositorios de datos y de las prácticas de citación de datos, las cuales son inconsistentes en muchos casos (**Robinson-García; Jiménez-Contreras; Torres-Salinas**, 2015). En la actualidad el *DCI* se basa en la información ofrecida por el repositorio de datos con respecto a publicaciones en las que se cita el conjunto de datos o el estudio de los datos. En cada registro el *DCI* muestra el número de veces que el objeto ha sido citado en la *Web of Science*. El *DCI* admite tres tipos de tipos de documentos:

- *data sets*
- *data studies*
- repositorios.

Como con los artículos, existen diferencias entre las áreas de conocimiento: las citas son mayores en ingeniería y tecnología que en ciencias sociales, arte y humanidades, aunque en éstas el estudio de datos es importante. Las citas de datos se encuentran más extendidas en disciplinas como la cristalografía y la genómica. Los *datasets* que contiene el *Data Citation Index* se encuentran muy sesgados hacia las ciencias puras y experimentales. El 94% de ellos provienen del 75% de las bases de datos *Gene Expression Omnibus*, *UniProt Knowledgebase*, *Pangaea* and *U.S. Census Bureau Tiger/Line Shapefiles* (**Torres-Salinas; Martín-Martín; Fuente-Gutiérrez**, 2014).

El *DCI* sigue un proceso de selección con el fin de mantener ciertos estándares de calidad. Los criterios seguidos para la indexación de los repositorios incluyen factores como:

- normas de publicación;
- contenido editorial;
- diversidad internacional de la autoría;
- procedencia geográfica;
- citas recibidas.

### **Repositorios de datos**

Aportan poca información sobre métricas de los *datasets* depositados. Se han consultado algunos y únicamente ofrecen información sobre visualizaciones, descargas y en algún caso estadísticas de carácter general. Por ejemplo:

- *Figshare*: visualizaciones y descargas;
- *Dryad*: descargas;
- *Open Aire*: información estadística general sobre visualizaciones, bancos de datos más citados, publicaciones por tipo, país, idioma y proveedor.

Otros productos que dan información sobre métricas de datos (**Ball; Duke**, 2015) son *ResearchGate* e *ImpactStory*, que además proporcionan información sobre las publicaciones y sobre *datasets* estadísticos. *PLoS* por su parte aporta estadísticas de los artículos que ha publicado, ofreciendo tanto información sobre citas en *Web of Science* y *Scopus*, como sobre visualizaciones y descargas, comentarios y calificaciones en los medios sociales, en algunos blogs o en *Wikipedia*.

Otro producto de estas características es *PlumX*, que fue adquirido por *Ebsco* a principios de 2014. Lo produce *Plum Analytics*, el principal proveedor de altmétricas de investigación. Ofrece una imagen completa del impacto de la investigación, incluyendo conjuntos de datos y códigos fuente, así como las publicaciones tradicionales. El producto va más dirigido a instituciones que a personas. Ofrece métricas de cinco categorías:

- uso: contabiliza el número de veces que un recurso ha sido visto o descargado, enlaces recibidos, clics, usuarios que contribuyen en *GitHub*, bibliotecas que tienen una copia;
- captura: veces que el recurso ha sido marcado como de interés (por ejemplo, un libro marcado en *Delicious*, añadido a una biblioteca en *Mendeley*, seguido, enlazado o visto en *GitHub*);
- menciones: posts en blogs, comentarios hechos (en *Facebook*, *SlideShare*, *YouTube*, etc.), o recibidos (en *Amazon* o *GoodReads*);
- medios de comunicación social: veces que se ha recomendado el recurso (por ejemplo “me gusta” en *Facebook*, “+1” en *Google+*, votos en *Reddit*, tweets...);
- citas: número de citas que ha recibido en la *WoS*, *Scopus* y otras bases de datos.

Es necesario encontrar métricas que permitan salir del círculo vicioso en el que nos encontramos

## 8. Conclusiones

Actualmente aún es escasa la reutilización de datos en todas las áreas de conocimiento, y las métricas todavía no se encuentran lo suficientemente desarrolladas.

Un problema con el que se enfrenta la evolución de las métricas basadas en la cita de datos es la escasa conciencia que existe para incluir su cita en las publicaciones. La citación se incrementará si se reconoce en la evaluación de la actividad científica o si los editores ven un modelo de negocio en ello. Puede influir también favorablemente el que los organismos financiadores lo tengan en cuenta a la hora de financiar proyectos o como elemento obligatorio para incrementar su impacto y transferencia.

Otros problemas que hay que resolver a la hora de proponer métricas es el gran número de bancos de datos existentes y su dispersión (más de 1.300 indexados por *re3data*), y la necesidad de elaborar estándares y protocolos de interoperabilidad entre los distintos actores.

Los indicadores altmétricos, que podrían ser adecuados para medir el impacto de los *datasets*, no se encuentran lo suficientemente desarrollados y aceptados por la comunidad académica.

Es necesario encontrar métricas adecuadas que permitan salir del círculo vicioso en el que nos encontramos, en el que los investigadores no comparten sus datos porque esta práctica no está reconocida y, por lo tanto, no les vale la pena invertir tiempo. Implementar métricas es difícil, y sin métricas cuesta estimular el intercambio de datos (Ball, 2015).

## 9. Bibliografía

**Aleixandre-Benavent, Rafael; Valderrama-Zurián, Juan-Carlos; González-Alcaide, Gregorio** (2007). “El factor de impacto de las revistas científicas: limitaciones e indicadores alternativos”. *El profesional de la información*, v. 16, n. 1, pp. 4-11.  
<http://eprints.rclis.org/9489>

**Ball, Alex** (2015). “Better data-level metrics: Quality and impact”. En: *Workshop held as part of the BioMedBridges Symposium ‘Open bridges for life science data’ at the Wellcome Genome Campus*, Hinxton, UK, 17-18 November 2015, pp. 17-18.  
<http://opus.bath.ac.uk/48224>  
<http://dx.doi.org/10.7490/f1000research.1111029.1>

**Ball, Alex; Duke, Monica** (2015). “How to track the impact of research data with metrics”. *DCC how-to guides & checklists*. Edinburgh: Digital Curation Centre.  
<http://www.dcc.ac.uk/resources/how-guides>  
<http://www.dcc.ac.uk/resources/how-guides/track-data-impact-metrics#sthash.U85HBEIb.dpuf>

**Bbsrc** (2007) “Bbsrc data sharing policy: version 1.2”. *Bbsrc, bioscience for the future*.  
<http://www.bbsrc.ac.uk/datasharing>

**Brase, Jan; Farquhar, Adam** (2011). “Access to research data”. *D-lib magazine*, v. 17, n. 1-2.  
<http://dx.doi.org/10.1045/january2011-brase>

**Brase, Jan; Farquhar, Adam; Gastl, Angela; Gruttenmeier, Herbert; Heijne, Maria; Heller, Alfred; Piquet, Arlette; Rombouts, Jeroen; Sandfaer, Mogens; Sens, Irena** (2009). “Approach for a joint global registration agency for research data”. *Information services and use*, v. 29, n. 1, pp. 13-27.  
<http://orbit.dtu.dk/files/4261926/Publication%20DataCite%20J.B..pdf>  
<http://dx.doi.org/10.3233/ISU-2009-0595>

**Callaghan, Sarah; Lehnert, Kerstin** (s.f.). *Bibliometrics working group. Publishing Data Bibliometrics*.  
[https://rd-alliance.org/sites/default/files/attachment/1\\_PublishingDataBibliometrics\\_SarahCallaghan.pdf](https://rd-alliance.org/sites/default/files/attachment/1_PublishingDataBibliometrics_SarahCallaghan.pdf)

**Costas, Rodrigo; Meijer, Ingeborg; Zahedi, Zohreh; Wouters, Paul** (2013). *The value of research data. Metrics for datasets from a cultural and technical point of view*. Copenhagen: Knowledge Exchange.  
<http://www.knowledge-exchange.info/datametrics>

**DCIP** (s.f.). *Data Citation Implementation Pilot*  
<https://www.force11.org/group/dcip>

**Ferrer-Sapena, Antonia; Aleixandre-Benavent, Rafael; Peset, Fernanda; Vidal-Infer, Antonio; Alonso-Arroyo, Adolfo** (2016). “Los científicos ante los datos abiertos. Red Maredata”. En: *II Encuentro de datos abiertos de la Universidad de Alicante*.  
<http://es.slideshare.net/uadatos/presentacion-alicante-59894628>

**Lawrence, Bryan; Jones, Catherine; Matthews, Brian; Pepper, Sam; Callaghan, Sarah** (2011). “Citation and peer review of data: Moving towards formal data publication”. *The international journal of digital curation*, v. 6, n. 2, pp. 4-37.

<http://dx.doi.org/10.2218/ijdc.v6i2.205>

**Lin, Jennifer; Cruse, Patricia; Fenner, Martin; Strasser, Carly** (2014). *Making data count: A data metrics pilot project*. University of California. California Digital Library. <http://escholarship.org/uc/item/9kf081vf>

**Martone, Maryann** (ed.) (2014). *Data Citation Synthesis Group: Joint declaration of data citation principles*. San Diego CA: Force11. <https://www.force11.org/group/joint-declaration-data-citation-principles-final>

**Mooney, Hailey; Newton, Mark P.** (2012). "The anatomy of a data citation: Discovery, reuse, and credit". *Journal of librarianship and scholarly communication*, v. 1, n. 1, eP1035. <http://dx.doi.org/10.7710/2162-3309.1035>

**NISO** (2016a). *Alternative outputs in scholarly communications: Data metrics*. National Information Standards Organization. <http://goo.gl/BHe7qq>

**NISO** (2016b). *Altmetrics definitions and use cases*. National Information Standards Organization. <http://goo.gl/WcsBHR>

**Parsons, Mark; Fox, Peter** (2013). "Is data publication the right metaphor?". *Data science journal*, v. 12, pp. WDS32–WDS46. [https://www.jstage.jst.go.jp/article/dsj/12/0/12\\_WDS-042/\\_pdf](https://www.jstage.jst.go.jp/article/dsj/12/0/12_WDS-042/_pdf)

**Piowar, Heather A.; Day, Roger S.; Fridsma, Douglas B.** (2007). "Sharing detailed research data is associated with increased citation rate". *PLoS one*, 21 March.

<http://dx.doi.org/10.1371/journal.pone.0000308>

**Piowar, Heather A.; Vision, Todd J.** (2013). "Data reuse and the open data citation advantage". *PeerJ*, 1:e175. <https://doi.org/10.7717/peerj.175>

**RDA** (s.f.). "Data re-use, share your experiences". *Research Data Alliance*. <https://rd-alliance.org/groups/data-re-use-share-your-experiences.html>

**RDA-WDS** (s.f.). *Cost recovery for data centres Working Group, Case statement*. RDA-WDS Publishing Data Interest Group. Bibliometrics Working Group. [https://rd-alliance.org/sites/default/files/case\\_statement/RDA\\_WDS\\_IG\\_Publishing\\_Costs.pdf](https://rd-alliance.org/sites/default/files/case_statement/RDA_WDS_IG_Publishing_Costs.pdf)

**Robinson-García, Nicolás; Jiménez-Contreras, Evaristo; Torres-Salinas, Daniel** (2015). "Analyzing data citation practices using the data citation index". *Journal of the Association for Information Science and Technology*. <https://arxiv.org/abs/1501.06285> <http://dx.doi.org/10.1002/asi.23529>

**Tenopir, Carol; Allard, Suzie; Douglass, Kimberly; Aydinoglu, Arsev U.; Wu, Lei; Read, Eleanor; Manoff, Maribeth; Frame, Mike** (2011). "Data sharing by scientists: Practices and perceptions". *PLoS one*, v. 6, n. 6, p. e21101. <http://dx.plos.org/10.1371/journal.pone.0021101>

**Torres-Salinas, Daniel; Martín-Martín, Alberto; Fuente-Gutiérrez, Enrique** (2014). "Analysis of the coverage of the data citation index - Thomson Reuters: disciplines, document types and repositories". *Revista española de documentación científica*, v. 37, n. 1, e036. <http://dx.doi.org/10.3989/redc.2014.1.1114>

# Inforàrea

Ayudamos a tu organización en la transformación digital y el gobierno de la información



- \* Consultoría estratégica en gestión y gobierno de la información
- \* Gestión documental y "records management"
- \* Gestión de contenidos, intranets corporativas y entornos de colaboración
- \* Estudios especializados

Clientes satisfechos, cientos de empresas nacionales e internacionales y más de 30 años de experiencia son la mejor garantía de nuestra reputación.

Para más información consulta [www.Inforarea.es](http://www.Inforarea.es)

# USO DE DATOS EN LA CREATIVIDAD PUBLICITARIA: EL CASO DE ART, COPY & CODE DE GOOGLE

Use of data in advertising creativity: The case of Google's Art, Copy & Code

David Selva-Ruiz y Lucía Caro-Castaño



**David Selva-Ruiz**, doctor por la *Universidad de Sevilla*, es profesor del *Departamento de Marketing y Comunicación* y coordinador de *Marketing e Imagen Institucional* de la *Universidad de Cádiz*. Imparte clases sobre creatividad publicitaria y nuevas herramientas de comunicación comercial. Su investigación se centra en las nuevas tendencias y aplicaciones en comunicación y la cultura popular, contando con decenas de publicaciones en revistas científicas y libros académicos. Puede destacarse su libro *El videoclip. Comunicación comercial en la industria musical* (Alfar, 2014).  
<http://orcid.org/0000-0002-9251-0045>

[david.selva@uca.es](mailto:david.selva@uca.es)



**Lucía Caro-Castaño** es licenciada y doctora en publicidad y RRPP por la *Universidad de Cádiz (UCA)*. Ha sido profesora asociada en la *Universidad de Sevilla (US)* y en la actualidad imparte docencia en el *Departamento de Marketing y Comunicación* de la *UCA*. Miembro del grupo de investigación *Métodos, Análisis y Estrategias de la Comunicación Empresarial e Institucional (US)*. Investiga sobre nuevas prácticas y estrategias de comunicación online entre usuarios y marcas comerciales. Ha trabajado como responsable de comunicación en diversas empresas.  
<http://orcid.org/0000-0003-2720-1534>

[lucia.caro@uca.es](mailto:lucia.caro@uca.es)

*Universidad de Cádiz, Facultad de Ciencias Sociales y de la Comunicación*  
Av. de la Universidad, s/n. 11405 Jerez de la Frontera (Cádiz), España

## Resumen

La publicidad ha empleado tradicionalmente la investigación como una forma de conocer al consumidor, testar sus campañas, etc. Así, el uso de datos está relativamente asentado en la planificación estratégica publicitaria. En cambio, las posibilidades creativas de los datos están aún por explotar plenamente, máxime teniendo en cuenta que brindan oportunidades que la publicidad viene demandando, como una tendencia a la personalización y a la inmediatez. Este trabajo analiza el uso de datos en la creatividad publicitaria a través del caso del proyecto *Art, Copy & Code* de *Google* y de cuatro de sus campañas. El análisis revela numerosos rasgos comunes, como la aportación de valor al usuario o el empleo conjunto de diversas tecnologías *open web* y varios dispositivos. Todo ello en el marco de procedimientos y estructuras asentados en el ámbito publicitario pero que se ven beneficiados de la aportación de los datos.

## Palabras clave

Datos; Publicidad; Creatividad publicitaria; Comunicación digital; Marketing; Grandes datos; *Big data*; Datos abiertos; *Open data*; Publicidad online; Publicidad personalizada; *Google*.

## Abstract

Advertising has traditionally used research as a way to know the consumer, test its campaigns, etc. Thus, data usage is already part of the strategic plan. However, the creative possibilities of data are yet to be fully exploited, especially those opportunities that provide personalization and immediacy. This paper analyzes data usage in advertising creativity through the case of *Google's Art, Copy & Code* project and four of its campaigns. This analysis reveals many features that would add value for the user including joint use of various open web technologies and different devices, all within the framework of procedures and structures that are well-established in the advertising field.

## Keywords

Data; Advertising; Advertising creativity; Digital communication; Marketing; Big data; Open data; Online advertising; Personalized advertising; *Google*.

Artículo recibido el 02-04-2016  
Aceptación definitiva: 12-05-2016

Selva-Ruiz, David; Caro-Castaño, Lucía (2016). "Uso de datos en creatividad publicitaria: el caso de *Art, Copy & Code* de Google". *El profesional de la información*, v. 25, n. 4, pp. 642-651.

<http://dx.doi.org/10.3145/epi.2016.jul.14>

## 1. Introducción

En el mercado publicitario *Google* ha sido hasta la fecha sinónimo de publicidad funcional: anuncios por palabras en su buscador y publicidad contextual en su red de espacios publicitarios. Aún hoy, pese a la diversificación de los formatos (*apps* de marca, publicidad nativa, etc.) la inversión en búsquedas pagadas mantiene el grueso de la inversión online en España —47,4% en 2015 (*IAB Spain*, 2015)—. Sin embargo, muchas marcas necesitan no sólo estar presentes en el momento de la evaluación de compra, sino también tener una presencia social (*Rayport*, 2013) que les permita vincularse con su público objetivo. *Art, Copy & Code* es un proyecto de *Google* destinado a fomentar el uso de la tecnología y en concreto de los datos en el ámbito publicitario y mercadotécnico, especialmente en las labores creativas. El objetivo último es posicionar a *Google* como una compañía clave en el sector de la publicidad, así como ubicar la gestión de los datos como un elemento tan importante como la redacción creativa y la dirección de arte en el día a día de la creatividad publicitaria en el entorno digital.

## 2. Objetivos y metodología

El presente trabajo tiene como objetivo examinar el uso de datos como recurso creativo en *Art, Copy & Code*, mediante el análisis de los elementos comunes a las campañas de ese proyecto de *Google*. Ello permitirá conocer las características de esta tendencia publicitaria cada vez más presente en el panorama actual de la comunicación.

Hay que puntualizar que este no es un trabajo sobre *Google* en sentido estricto. Se elige el caso de *Art, Copy & Code* como un proyecto paradigmático, en el que *Google* trata de abanderar lo que hoy es una tendencia creciente en publicidad, siendo un agente interesado, como *Facebook* y muchas otras empresas. El uso de datos es una opción más en el catálogo de recursos de los departamentos creativos de las agencias, y *Google* aspira a convertirse en socio prioritario de las agencias en este aprovechamiento de la tecnología y los datos.

Se lleva a cabo una revisión de los cambios producidos en la publicidad en relación con el uso de datos, y un análisis cualitativo de *Art, Copy & Code* como caso paradigmático de esta tendencia publicitaria, y de cuatro de las cinco campañas que se exhiben en la web del proyecto.

<http://www.artcopycode.com>

Se ha excluido la quinta campaña (*District Voices*, de *Lionsgate*) dado que no se basa en el empleo de datos y escapa por tanto al objeto de estudio de este trabajo. Las cuatro campañas analizadas son:

- *SmileDrive* (Volkswagen)
- *Kisses* (Burberry)
- *Phenomenal Shot* (Nike)
- *Madden Giferator* (EA Sports).

## 3. Uso de datos en creatividad publicitaria

### 3.1. El mensaje publicitario como contenido

Transformar la publicidad en contenido elegido por el consumidor, eliminando su carácter intrusivo, ha sido un anhelo clásico de los profesionales de la publicidad. Para ello, a lo largo de la historia de esta disciplina, los equipos creativos han apelado al humor, al placer estético o a la emoción como recursos para atraer la atención del espectador (*López*, 2007). La llegada de la *world wide web* abrió un nuevo campo para la comunicación comercial y la exploración de nuevas formas de trasladar los mensajes de marca buscando superar el carácter intrusivo de los formatos tradicionales —dinámica *push* (spots, cuñas, gráficas, etc.)—, y potenciando la aparición de formatos que no interrumpiesen al usuario —bajo una dinámica *pull* (publicidad nativa, *advergaming*, *apps*, etc.)—.

Sin embargo, durante la primera época de internet, anunciantes y agencias se dedicaron a replicar el modelo intrusivo de los medios de comunicación de masas. El *banner*, en sus múltiples formatos (robapáginas, *interstitials*, *layers*, etc.), se convirtió inicialmente en la herramienta central de la publicidad online, aunque desde el principio los anunciantes expresaron su decepción por los bajos ratios de éxito del formato y la escasa información que ofrecían sobre el consumidor (*Turow*, 2011), un problema que persiste (*Roca*, 2014). El carácter intrusivo que perciben los usuarios en formatos como el *banner* (*Van-Doorn; Hoekstra*, 2013), así como el incremento de la publicidad nativa (*Couldry; Turow*, 2014), donde los mensajes promocionados adoptan las pautas formales de los contenidos informativos de la plataforma (los *tweets* patrocinados, por ejemplo), son factores que contextualizan que el principal problema en internet para los españoles siga siendo que encuentran "demasiada publicidad" —59,4%— (*AIMC*, 2016).

Google aún la mayor red de espacios publicitarios online con un abanico de servicios al usuario que le permite conocer tanto su grafo social, como su comportamiento de consumo

Afortunadamente para anunciantes y publicitarios, en los comienzos del siglo XXI ha emergido una floreciente industria para la captación y análisis de grandes cantidades de datos (*big data*) sobre comportamiento de consumidor, así como un mayor conocimiento del modo en que los usuarios conciben este metamedio (*Manovich*, 2008) y el tipo de experiencias (*Ashley; Tuten*, 2015; *Rayport*, 2013; *Martí; Muñoz*, 2008) y relaciones (*Sheehan; Morrison*, 2009) que pueden ofrecer las marcas en estos entornos.

### 3.2. La personalización como nueva lógica del medio

Consumir contenido en la Red implica al mismo tiempo producir datos (Deuze, 2006), dado que el actual diseño de la misma se caracteriza por la “dataficación” (*datafication*) de magnitudes nunca antes medidas sobre el comportamiento de los usuarios (Van-Dijck; Poell, 2013). Esta nueva fuente de datos permite la monitorización de las opiniones que los consumidores vierten sobre productos, marcas y campañas publicitarias.

Al convertir a todo usuario de la web 2.0 en un productor de contenido, la Red entró en un proceso de crecimiento exponencial de la información que, unido a la fragmentación de las audiencias y al crecimiento del consumo multipantalla (Nielsen, 2015), ha convertido la atención en un bien cada vez más escaso. En una encuesta realizada por Nielsen (2015) en 60 países, la consultora detectaba que:

- el 47% de los encuestados utiliza los medios sociales mientras ve televisión en directo;
- el 58% usa internet para buscar contenidos relativos a lo que está viendo en televisión.

La personalización se perfila como la lógica mediática específica de internet al ofrecer a cada sujeto información relevante en relación con su propio consumo

La respuesta de compañías como *Google*, *Amazon* o *Facebook* a la necesidad de captar el interés de los usuarios ha sido la personalización de los contenidos, una adaptación que realizan de acuerdo con sus decisiones de consumo previas y/o con las elecciones de usuarios con intereses y comportamientos semejantes. Así, la personalización se perfila en la actualidad como la lógica mediática específica de internet (Couldry; Turow, 2014), al ofrecer a cada sujeto información relevante en relación con su propio consumo.

Los mensajes publicitarios personalizados son posibles gracias al refinamiento de los sistemas de intercambio y sincronización de *cookies* que permiten a las plataformas compartir estos archivos para rastrear los movimientos de un mismo usuario a través de la Web, aunque no se haya dado de alta (Nesamoney, 2015). Así, estas corporaciones pueden construir perfiles de consumidor cada vez más sofisticados que podrán alquilar a los anunciantes. A este empleo de los datos para mejorar las microsegmentaciones de perfiles debe añadirse la realización de pretests de optimización para conocer el éxito de las diferentes versiones de un mismo anuncio, lo que permite incrementar la eficacia de las campañas sobre la marcha (Nichols, 2013).

### 3.3. Datos contextuales como fuente de relevancia para la publicidad

Sheehan y Morrison (2009) plantearon la emergencia de una “cultura de la confluencia” en la web social, donde tendencias diversas tienden a fusionarse bajo la lógica del remix y el *mashup*. Esta cultura del remix ofrecería una nueva vía para resolver el desafío de la involucración que afronta

ba la publicidad online: reinventar el modelo de los mensajes masivos y ayudar al consumidor a contar historias. Al desarrollo de este tipo de narrativa propia del medio digital ha contribuido de modo central la mejora en la interoperabilidad entre plataformas, así como la implementación de nuevos estándares de diseño web abiertos —html5, CSS y *Java Script*, frente a *Flash* y *ActionScript*— que reducen la dependencia de la instalación de *pluggins* y permiten la combinación de datos de varios servicios en tiempo real (Percival, 2013). De este modo se ha ido configurando un ecosistema digital que ofrece nuevas libertades expresivas a la publicidad frente al encorsetamiento de los formatos en medios tradicionales, al tiempo que permite formas de incorporar al usuario en el propio mensaje. En esta línea, el publicitario español y socio fundador de la agencia interactiva española *Double You*, Daniel Solana (2012), ha señalado que el actual desarrollo de la Web ha permitido la aparición de una creatividad líquida, donde el soporte es sólo el punto de partida de la vinculación entre consumidor y marca.

Pese a que Solana anunciaba hace ya varios años la entrada en una nueva era de la creatividad online, la incorporación de los datos de modo creativo en los anuncios y proyectos de marca ha comenzado a desarrollarse en fechas muy recientes. La categoría *Creative data*, por ejemplo, no ha aparecido en los *Cannes Lions* hasta la edición de 2015. En este contexto de reinención, la iniciativa *Art, Copy & Code* de *Google* es probablemente la apuesta más potente del sector, en tanto que la compañía aúna el poder mediático, gracias a su capacidad para atraer la atención de los usuarios con:

- a) la mayor red de espacios publicitarios (*Google Display Network*), que permite la puja en tiempo real (Papi-Gálvez, 2015) mediante la compra programática;
- b) el acceso al grafo social de estos (*Gmail*, *G+*, *Google Docs*, etc.);
- c) el rastreo de su comportamiento a través de los servicios que le presta (buscador, *Google Chrome*, *YouTube*, *Google Play*, etc.).

La adaptación en tiempo real del anuncio al contexto del consumidor mediante diferentes servicios de información es una estrategia eficaz para atraer su atención

Otra forma de incrementar la relevancia potencial de los mensajes persuasivos es la incorporación de información contextual en tiempo real (Nesamoney, 2015; Rayport, 2013; Caro-Castaño; Selva-Ruiz, 2011). De este modo, por ejemplo, la adaptación del anuncio a la temperatura de la ciudad desde la que el usuario está conectado, la incorporación de la música que ha escuchado recientemente en una radio online o la aparición de la celebridad a la que sigue en *Twitter*, serían decisiones creativas que ayudan a atraer su atención. La incorporación de estos datos de contexto permite que el usuario se sienta interpelado, sin llegar a ser informaciones tan personales que deriven en una sensación

de pérdida de privacidad (Tucker, 2014), como sucede con la incorporación de datos de transacciones previas (Van-Doorn; Hoekstra, 2013).

Las ideas creativas y su desarrollo se apoyan en las posibilidades prestadas por la tecnología y en concreto por los datos

## 4. El caso de *Art, Copy & Code* de Google

### 4.1. El proyecto: introducción y contextualización

*Art, Copy & Code* es una iniciativa de Google a escala mundial en el campo publicitario, que se puso en marcha en 2013 y pervive en la actualidad. Como compañía especializada en productos y servicios de tecnología e internet, Google presta gran atención a la publicidad como elemento clave en su exitoso modelo de negocio. Por esta razón no debe sorprender que intente aliarse con este sector, y en concreto con el subsector de la creatividad publicitaria. El propio título de la iniciativa es ya una declaración de intenciones en toda regla. Desde los años 60 del siglo XX, el trabajo creativo tiene su núcleo en duplas creativas, sumando un *art director* o director de arte y un *copy* (*copywriter*) o redactor creativo. Esta estructura de trabajo, propuesta por Bill Bernbach (Castellblanque, 2006), figura clave en la historia de la publicidad, simboliza para Google una revolución creativa en el universo publicitario. Al insertar la palabra *Code* está ubicándose como un tercer elemento clave en la publicidad del futuro:

“Este proyecto fue inspirado por la revolución creativa de la década de 1960, cuando todo cambió simplemente asociando a directores de arte y redactores creativos. Había nacido la idea de un equipo creativo formado por ‘arte y copy’. Hoy en día estamos en medio de una segunda revolución creativa, impulsada por la tecnología. El ‘código’ está siendo agregado al núcleo del proceso creativo. Nuestros proyectos de *Art, Copy & Code* están orientados a proporcionar inspiración para todos aquellos que dan forma al futuro de la publicidad” (Google, 2016).

La novedad real de estas campañas se encuentra, no en el uso de datos para investigar al consumidor o evaluar resultados, aspectos que tocarían con otros ámbitos de la publicidad (planificación estratégica, planificación de medios, etc.), sino en el uso de los datos como activo para la creatividad. Son las propias ideas creativas y su implementación las que se apoyan en las posibilidades prestadas por la tecnología, y en concreto, por los datos.

El proyecto constituye un paso más allá en una estrategia que viene de atrás. Cabe resaltar el proyecto *Re: Brief*, de 2012. La propuesta, basada en la alianza con las marcas *Coca-Cola*, *Volvo*, *Alka-*

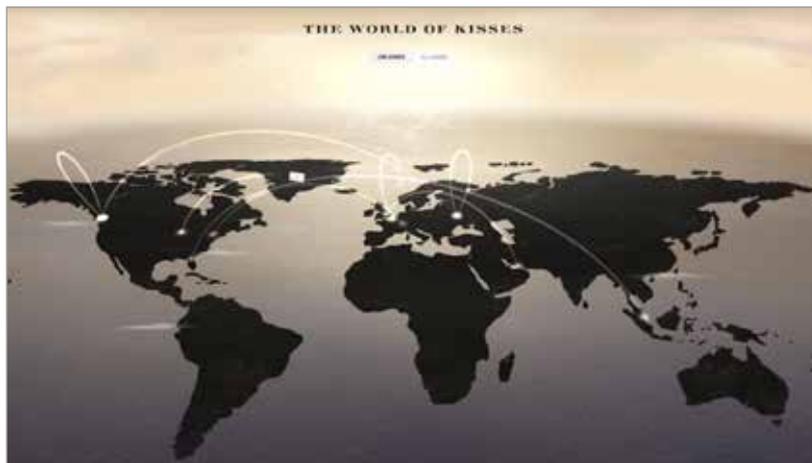
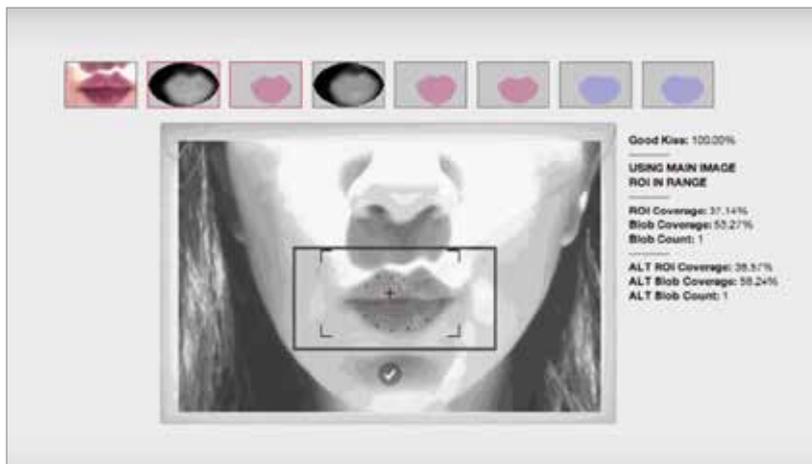
*Seltzer* y *Avis*, consistía en la actualización de una de sus campañas publicitarias icónicas. En colaboración con cada marca y con publicitarios que participaron en la creación de estas campañas, se adaptaban las campañas de publicidad convencional (basadas en anuncios convencionales en televisión, prensa, etc.) a las posibilidades brindadas por las nuevas tecnologías. Así por ejemplo, el mítico spot *Hilltop* (1971), de *Coca-Cola*, se reinventó con la colaboración del director de arte Harvey Gabor en una nueva campaña adaptada a las nuevas posibilidades prestadas por el medio digital, manteniendo el planteamiento estratégico y creativo de la campaña original pero añadiéndole personalización, participación del usuario, etc. y, en general, todos aquellos rasgos de la publicidad actual que eran inviables a principios de los años 70 del siglo XX. En concreto, el spot *Hilltop*, basado en la frase *I’d like to buy the world a Coke* y centrado en la celebración de la tolerancia y la unidad del género humano, se convierte con este proyecto en algo más. La nueva propuesta permite que los usuarios realmente “compren al mundo una *Coca-Cola*”, de modo que pueden grabar su mensaje con su móvil o su ordenador y enviarlo a cualquier lugar del mundo junto con una *Coca-Cola*; posteriormente, un usuario desconocido recibirá dicho mensaje mediante una máquina de *vending* diseñada al efecto, desde la cual podrá responder al mensaje. En paralelo los usuarios pueden ver, descargar y compartir películas personalizadas con el viaje del mensaje y la *Coca-Cola*. Así, lo que era un spot de carácter unidireccional se convierte en una acción compleja donde el usuario es partícipe activo y vive una experiencia personalizada. El lema *Advertising re-imagined* es bastante certero en cuanto al objetivo del proyecto, que coincide con el de *Art, Copy & Code*.

Se puede comprobar cómo Google intenta mostrar su propuesta como una nueva etapa en la historia de la publicidad, que resulta de una evolución natural propiciada por las nuevas posibilidades prestadas por la tecnología. *Art, Copy & Code* persigue, según Google:

“explorar nuevas formas en que la tecnología puede ayudar a construir marcas en un mundo digital. Desarrollamos campañas y experiencias que esperamos que la gente amará, recordará y compartirá” (Google, 2016).



SmileDrive, para Volkswagen



Kisses, para Burberry

## 4.2. Descripción de las campañas

### SmileDrive (Volkswagen)

Consiste en una aplicación orientada a convertir la conducción en una experiencia social. Las personas pasan muchas horas conduciendo, la mayoría de ellas en soledad. La aplicación les permite compartir momentos y estar conectados con los demás de un modo seguro (la aplicación se cierra automáticamente cuando se enciende el motor). Toma datos de tráfico, meteorología, duración, localizaciones, etc. y permite crear un registro de fotos, mapas, actualizaciones de estado, etc., que incluso pueden ser retransmitidos en directo (*SmileCast*) mediante *Google+*. Al finalizar la conducción, proporciona al usuario todo ese material y una puntuación (*SmileScore*). Consiste prácticamente en convertir la conducción en una experiencia lúdica.

Las cuatro campañas se basan en la colaboración entre productos y tecnologías, todos ellos clasificables como *open web technologies*

### Kisses (Burberry)

Se trata de una campaña orientada a conectar emocionalmente con los *millennials*. Para ello se recurre a los besos, a la posibilidad de enviarlos virtualmente entre personas de

todo el mundo y seguir su viaje. A través de una aplicación para móviles y tablets, se puede capturar el beso y enviarlo a quién se desee. Gracias a la conexión con los servicios de geolocalización de *Google*, se muestra visualmente el supuesto viaje de la carta con el beso entre el emisor y el receptor, incluyendo su vuelo por las calles de ambos.

### Phenomenal Shot (Nike)

Intenta conectar con el público que está viendo retransmisiones deportivas desde la televisión. Partiendo de que en la actualidad el público, mientras visiona un partido, busca estadísticas en internet, comenta las jugadas, celebra las victorias, etc., se generan y difunden anuncios en tiempo real basados en jugadas que han sucedido sólo segundos atrás en el partido que se emite en televisión. Con el propio navegador (sin necesidad de descargar ninguna aplicación), el público puede girar 360º en torno a la imagen hasta construir su propio anuncio *Nike*, que está siempre basado en un deportista vinculado a la marca.

### Madden Giferator (EA Sports)

Tiene algunos puntos en común con la anterior. Partiendo de nuevo de un uso del móvil como segunda pantalla mientras se visiona un partido de la *NFL* (la liga estadounidense de fútbol americano) en la tele-

visión, *Madden Giferator* crea una serie de gifs animados basados en lo que ha sucedido en el campo sólo segundos antes pero diseñados con imágenes propias del videojuego *Madden NFL 15*. Los gifs se muestran en tiempo real como anuncios, pero además el usuario puede editarlos mediante combinaciones de imágenes, frases y fondos.

## 4.3. Aspectos comunes en las campañas

De acuerdo con el análisis llevado a cabo, estas campañas se caracterizan por:

- importancia de los datos de cara a la ideación;
- personalización (automática) de la comunicación;
- vinculación con el *experiential marketing*;
- aportación de valor al usuario;
- mantenimiento de los conceptos básicos de la creatividad publicitaria;
- combinación entre comunicación *push* y *pull*;
- estructura de trabajo similar a la publicitaria (con el añadido de *Google*);
- cooperación entre distintos productos y tecnologías (*open web technologies*);
- uso de diversos dispositivos (con especial atención al móvil);
- vinculación con *insights* y conceptos cotidianos.

En todas las campañas analizadas, los datos tienen uno o varios roles de gran relevancia: conocimiento del consumidor, creatividad, segmentación en su difusión, evaluación

de resultados, etc. El potencial de los datos es consabido en una industria como la publicitaria, que invierte una gran cantidad de recursos en investigación (**González-Lobo; Prieto-del-Pino**, 2009). Sin embargo, se insiste mucho por parte de *Google* en el potencial de los datos no sólo para definir la estrategia o evaluar resultados, sino también para la creatividad y, en particular, para el *storytelling* (técnica consistente en contar historias). Es un aspecto de tanta importancia para la compañía que el propio título de la misma parte de los dos principales roles creativos tradicionales, que se combinan con lo que defienden que debe ser un nuevo perfil fundamental en el trabajo creativo. Así, recomiendan que el equipo creativo trabaje en un mismo espacio que los expertos en tecnología:

“Esto permite a los creativos y a los tecnólogos trasladar el trabajo de uno a otro a la perfección o trabajar en conjunto. Y a veces los problemas técnicos se pueden resolver con una solución creativa (por ejemplo, un poco de experiencia de usuario podría corregir un error técnico), o viceversa” (*Google*, 2016).

“*Google* acentúa el potencial de los datos no sólo para definir la estrategia o evaluar resultados, sino también para la creatividad y, en particular, para el *storytelling*”

El empleo de datos facilita además la personalización de la comunicación. Esta tendencia no es algo tan novedoso en publicidad y marketing (**Peppers; Rogers**, 1996) como otros aspectos del proyecto. A priori, que el usuario pueda seleccionar textos e imágenes y crear su propio gif con *Madden Giferator* no resulta tan innovador. Pero sí lo es la personalización llevada a cabo a través de los datos sin necesidad de una selección por parte del usuario. En *SmileDrive*, la propia aplicación va recopilando datos del viaje del usuario, de modo que el contenido personalizado que éste podrá ver, y que quizá incluso ha ido retransmitiendo en directo, no es seleccionado activamente por él. Incluso el vídeo promocional del propio proyecto *Art, Copy & Code* realiza algo parecido. Se trata de una pieza que cambia según el usuario y su contexto. Al lanzar el vídeo, el programa recopila datos sobre la hora, el lugar, la meteorología o los eventos que han tenido repercusión recientemente, configurándose automáticamente un vídeo que supone una experiencia personalizada.

Desde un punto de vista mercadotécnico, las

campañas de *Art, Copy & Code* se ubican en la órbita del *experiential marketing* (**Ashley; Tuten**, 2015; **Rayport**, 2013; **Martí; Muñoz**, 2008; **Cova; Salle**, 2008). Éste se centra en la creación de experiencias para el consumidor en torno a la marca: enviar besos a quien el usuario desee en cualquier lugar del mundo, compartir gifs animados para burlarse de los seguidores del equipo rival, convertir la conducción en un juego, etc. Evidentemente, algo que tienen en común las cuatro marcas es que son de alto valor añadido y pertenecen a categorías de producto propicias para este tipo de estrategias.

“El *experiential marketing* se centra en la creación de experiencias para el consumidor en torno a la marca”

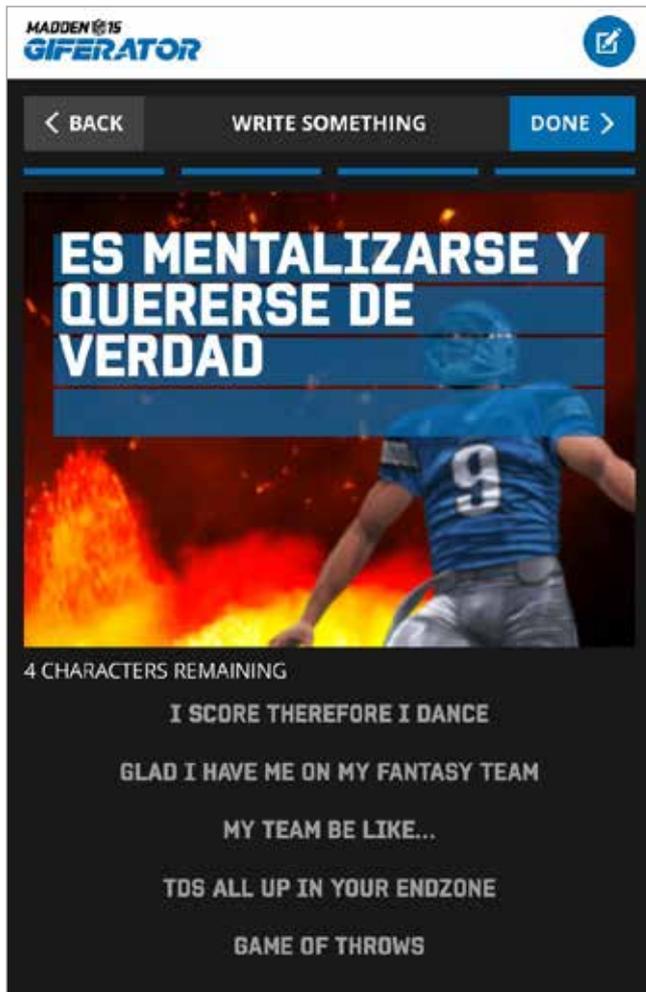
Es importante destacar que las cuatro campañas analizadas se basan en aportar valor al usuario ofreciéndole una experiencia personalizada y un contenido entretenido o útil. No en vano el propio término “campaña” comparte protagonismo de forma creciente con el de “proyecto” en el ámbito profesional de la publicidad, dada la disolución de fronteras entre comunicación y producto y la tendencia a que la comunicación no consista ya en una combinación de anuncios convencionales si no en aportar contenidos y valor al usuario con acciones que superan el alcance la publicidad convencional (o, dicho de otro modo, de los anuncios en medios tradicionales)<sup>1</sup>. Lo mismo sucede con las propuestas de *Art, Copy & Code*, que constituyen productos en sí mismos que el usuario puede consumir. Así lo defiende *Google* cuando recomienda pensar en el marketing como un producto:

“Complementa los anuncios mediante la creación de una plataforma como una aplicación o un producto que ofrezca utilidad y entretenimiento añadidos para los consumidores” (*Google*, 2016).

Una prueba de esta característica sería el hecho de que el 66% de los usuarios que descargan la aplicación de *SmileDrive* no son propietarios de un *Volkswagen* (*Google*, 2016).



Imagen personalizada generada por *Phenomenal Shot*, para Nike



Captura de la pantalla del teléfono móvil durante la personalización en *Madden Giferator*, para EA Sports.

Es importante destacar que pese a las novedades, todo queda encuadrado en publicidad. Cuando se observa la explicación de *Google* respecto a las cuatro campañas analizadas, se puede comprobar que se siguen planteamientos que, con ciertos añadidos, vienen de años atrás: un *brief*, un *insight*, una *big idea*, etc. Evidentemente no deja de ponerse de relieve cómo los datos ayudan en cada uno de esos pasos. Por ejemplo, los *big data* pueden ser una fuente muy útil de cara a la detección de *insights* sobre el consumidor. Para *Phenomenal Shot*, los datos de *Google* mostraban los picos en búsquedas después de grandes momentos durante las competiciones deportivas, y eso resultó clave para partir del *insight* de que los usuarios emplean de forma intensiva su móvil como segunda pantalla justo después de esos grandes momentos. Así, *Google* propone “abrazar los datos”:

**Insight:** Un aspecto oculto de la forma de pensar, sentir o actuar de los consumidores, útil para que la publicidad consiga conectar una marca con éstos y con sus experiencias cotidianas

“Los datos de búsquedas pueden ser una gran fuente de conocimiento acerca del tipo de contenido que los consumidores están buscando en grandes momentos culturales colectivos. El examen de estos datos puede ayudar, no sólo con lo que se crea, sino con cuándo se muestra al mundo” (*Google*, 2016).

Un aspecto de gran interés es que, pese a que las acciones tienen un marcado carácter de *branded content* y se basan en que el usuario elija acercarse a esos contenidos voluntariamente, las cuatro campañas de *Art, Copy & Code* emplean también publicidad intrusiva (*push*). La cuestión es sencilla: se espera que la gente acuda voluntariamente al contenido, pero para eso el usuario debe conocer ese contenido, de modo que la publicidad *push* opera aquí como una muestra del mismo. Cierto es que no se emplean métodos extremadamente intrusivos, gracias al empleo de módulos *Lightbox*. Esto supone la inserción de un *banner* que, cuando el público permanece 2 segundos con el cursor sobre él, permite ver un vídeo o un contenido interactivo determinado de un modo más protagonista. Así sucede por ejemplo en *SmileDrive*, en cuyos anuncios *Lightbox* el público podía ver una muestra del funcionamiento de la aplicación en los casos de microcelebridades (*blogueros, youtubers, etc.*).

“*Google* muestra su propuesta como una nueva etapa en la historia de la publicidad, que resulta de una evolución natural propiciada por las posibilidades de la tecnología”

Por otra parte, se produce una segmentación extrema e inmediata gracias al uso de *Google Display Network*, que constituye una red de webs con espacio publicitario servido a través de *Google Adwords*. La existencia de esta red propia de *Google* permite que los anuncios aparezcan en las webs y aplicaciones de modo inmediato y en función de las búsquedas realizadas por el usuario a través de *Google*. Su uso es clave para *Phenomenal Shot* y para *Madden Giferator*, dado que las campañas se basan en la inmediatez con respecto a lo que sucede en los encuentros deportivos que se están emitiendo en televisión. Todo ello permite observar que:

- incluso en campañas tan novedosas como esas sigue empleándose publicidad intrusiva, dado que lo *push* y lo *pull* se retroalimentan;
- incluso lo *push* evoluciona para acercarse a las nuevas tendencias del usuario hasta convertirse prácticamente en un modelo mixto.

En la cadena de trabajo la gran novedad es la introducción de *Google* como sujeto activo en el proceso. No obstante, si la cadena tradicional para el desarrollo de un spot televisivo, estaba compuesta principalmente por un anunciante, una agencia y una productora audiovisual, cabe realizar una analogía entre esta última y el papel de *Google* en tanto que desarrollador. En cualquier caso, la estructura de trabajo no cambia tanto como cabría esperar, ni desde luego *Google* suplanta el rol de la agencia publicitaria, sino que se trata de una colaboración. Por ejemplo, en el caso de *Kisses*, tiene lugar una colaboración entre *Burberry* (anunciante), *Grow*

(agencia digital) y *Google* (desarrollador). En otros casos, la estructura es más compleja, dadas las ramificaciones y complejidades técnicas de las campañas. Así, en *Phenomenal Shot* colaboran *Nike* (anunciante), *Wieden+Kennedy* (agencia publicitaria principal de la firma deportiva), *Grow* (agencia digital), *Mindshare* (agencia de medios) y *Google* (desarrollador).

Las cuatro campañas se basan en la colaboración entre productos y tecnologías, todos ellos clasificables como *open web technologies*. Exceptuando aquellos muy específicos de una campaña concreta, la mayoría de estas tecnologías son empleadas en varias campañas, y a su vez, cada campaña emplea un gran número de estas tecnologías (véase tabla 1). Por ejemplo en *Phenomenal Shot*, *html5* y *JavaScript* operan de forma conjunta con datos obtenidos del giroscopio y la brújula del teléfono para determinar la posición del deportista en la imagen 3D generada, de modo que el usuario puede modificar su visión de aquel moviendo su teléfono móvil.

Además de tecnologías, se tienen en cuenta también los diferentes dispositivos. Por ejemplo *Kisses* puede emplearse en ordenadores, tablets y móviles, de modo que la experiencia interactiva sea semejante al margen del dispositivo que se use habitualmente. En cualquier caso existe un convencimiento por parte de *Google* de la importancia presente y futura del teléfono móvil. Esto se observa especialmente en *Phenomenal Shot* y en *Madden Giferator*, donde se parte de que los telespectadores del fútbol emplean constantemente el móvil como segunda pantalla para buscar información y estadísticas, comentar las jugadas, burlarse de los rivales, etc. Aprovechan la actividad del usuario y se adaptan a ella, algo que explica *Google*:

“Al crear herramientas y contenido para la segunda pantalla, recuerda que eres la *cheerleader* -chica animadora-, no el *quarterback* –líder ofensivo en el fútbol americano-. En lugar de campañas totalmente nuevas, distrayendo la acción, haz ‘compañeros’: maneras de alimentar la pasión sin interponerte en el camino” (*Google*, 2016).

Esta visión no implica una merma de las posibilidades del móvil: al contrario, *Google* defiende su importancia hasta expresar que

“para las pantallas pequeñas, pienses en grande” (*Google*, 2016).

Es decir, no conviene concebir el móvil como una versión

Tabla 1. *Open web technologies* empleadas en las campañas de Art, Copy & Code

Tecnología	SmileDrive (Volkswagen)	Kisses (Burberry)	Phenomenal Shot (Nike)	Madden Giferator (EA Sports)
AdMob		X		
Android	X			
Bluetooth	X			
CSS			X	X
DoubleClick Rich Media			X	X
Gif Generation				X
Goo.gl URL Shortener			X	X
Google App Engine	X		X	X
Google Cloud SQL				X
Google Cloud Storage	X		X	X
Google Compute Engine			X	X
Google Display Network			X	X
Google Earth		X		
Google Maps	X	X		
Google Places		X		
Google+	X	X	X	X
GPS	X			
html5		X	X	X
Javascript			X	X
Lightbox ad units		X	X	
Lip Detection		X		
Mobile Gyroscope			X	
WebGL		X	X	
WebP		X		

en miniatura del escritorio del ordenador, sino que es factible expresar posibilidades del móvil sin siquiera requerir la descarga de una *app* específica. Van más allá, planteando incluso la conveniencia de diseñar pensando en las pantallas pequeñas para luego realizar la adaptación a pantallas grandes, y no al contrario.

En todas las innovaciones suelen detectarse ideas sencillas y cotidianas, pero impulsadas gracias a las posibilidades tecnológicas. Es claro el ejemplo de *Kisses*, que como afirma Christopher Bailey, CEO de *Burberry*, parte de una idea muy simple:

“mandas uno de tus besos personales a cualquiera, en cualquier lugar del mundo, y sigues el viaje de tu beso” (*Google*, 2016).

Luego la tecnología lo hará posible, pero la base es algo tan sencillo como enviar besos. Según Drew Ungvarsky, director creativo de la agencia *Grow*, el éxito de la campaña viene precisamente de la combinación de ideas nuevas y antiguas, tomando un tema universal e introduciendo la tecnología en él (*Google*, 2016). El consejo de *Google* en relación con la campaña *SmileDrive* es identificar lo que hace especial a la marca, y sólo después añadir interactividad para crear una experiencia digital en torno a ello.

## 5. Conclusiones

Como se ha podido comprobar en el caso de *Art, Copy & Code* de Google, las posibilidades de uso de datos para la creatividad publicitaria son enormes, hasta el punto de que al estar aún en un estadio incipiente, resultan difíciles de imaginar en toda su dimensión.

Aun así, hay ciertos aspectos que merece la pena hacer notar. La personalización es una tendencia ya asentada en comunicación; sin embargo, la posibilidad de captar datos del usuario y la realidad que le rodea en tiempo real permite una automatización de esta personalización que podría mejorar sensiblemente la eficacia de las campañas, además de permitir el desarrollo de propuestas creativas que hace pocos años habrían resultado inviables. Ello se lleva cabo mediante la cooperación necesaria entre diversos productos y tecnologías (que en el caso del proyecto analizado son siempre *open web technologies*) y múltiples dispositivos. En otras palabras, existe un *background* técnico de gran complejidad, pero con la pretensión de que el acceso del usuario a la experiencia sea fácil, cómodo y lúdico.

La posibilidad de captar datos del usuario y de la realidad que le rodea en tiempo real permite una automatización de la personalización que podría mejorar sensiblemente la eficacia de las campañas

Hay que destacar que nos movemos en el entorno del *experiential marketing* y en la aportación de valor al usuario. Esto, que puede ser considerado una virtud, nos revela también su posible restricción a marcas de alto valor añadido. Junto con los debates que se suelen generar en torno a la privacidad en el uso de datos, este aspecto constituye una de las principales limitaciones de este fenómeno.

En cualquier caso se comprueba que los datos vienen a sumarse a planteamientos consolidados en publicidad, y no a sustituirlos. La interrelación entre comunicación *push* y *pull*, así como la evolución de la primera, el trabajo en una estructura similar a la publicitaria, a la que Google y los técnicos vienen a sumarse y no a reemplazar, o la vinculación con conceptos arraigados en la práctica profesional de la publicidad, demuestran que no existe una ruptura sino una evolución.

Las nuevas posibilidades tecnológicas y en concreto el uso de los datos permiten nuevas opciones para conectar con un usuario cada vez más huidizo y crítico con la publicidad tradicional y, por tanto, constituyen un elemento de gran relevancia para la comunicación comercial y con un potencial aún por desarrollar en su plenitud.

## Nota

1. Pese a las diferencias conceptuales de ambas expresiones, campaña y proyecto se utilizan de forma casi indistinta en la práctica profesional de la publicidad. Ambas aludirían a iniciativas que son ideadas, planificadas y ejecutadas con un determinado objetivo de comunicación o publicitario, con el matiz de que la primera es la denominación tradicio-

nal, mientras que la segunda se considera más abierta a acciones que desbordan la publicidad convencional, es decir, los anuncios.

## 6. Bibliografía

AIMC (2016). *18º Informe navegantes en la Red. Encuesta AIMC a usuarios de internet*.

<http://www.aimc.es/-Navegantes-en-la-Red-.html>

Ashley, Christy; Tuten, Tracy (2015). "Creative strategies in social media marketing: An exploratory study of branded social content and consumer engagement". *Psychology & marketing*, v. 32, n. 1, pp. 15-27.

<http://dx.doi.org/10.1002/mar.20761>

Caro-Castaño, Lucía; Selva-Ruiz, David (2011). "Estrategias de convergencia y desintermediación en la difusión del videoclip mediante autocomunicación de masas". *Trípodos*, n. extra, pp. 155-163.

Castellblanque, Mariano (2006). *Perfiles profesionales de publicidad y ámbitos afines*. Barcelona: Editorial UOC. ISBN: 84 9788 507 4

Couldry, Nick; Turow, Joseph (2014). "Advertising, big data and the clearance of the public realm: Marketers' new approaches to the content subsidy". *International journal of communication*, n. 8, pp. 1710-1726.

<http://ijoc.org/index.php/ijoc/article/view/2166/1161>

Cova, Bernard; Salle, Robert (2008). "Marketing solutions in accordance with the SD logic: Co-creating value with customer network actors". *Industrial marketing management*, v. 37, n. 3, pp. 270-277.

<http://dx.doi.org/10.1016/j.indmarman.2007.07.005>

Deuze, Mark (2006). "Participation, remediation, bricolage: Considering principal components of a digital culture". *The information society*, v. 22, n. 2, pp. 63-75.

<http://jtc468.colostate.edu/docs/bricolage.pdf>

<http://dx.doi.org/10.1080/01972240600567170>

González-Lobo, María-Ángeles; Prieto-del-Pino, María-Dolores (2009). *Manual de publicidad*. Madrid: ESIC. ISBN: 978 84 7356 628 5

Google (2016). *Art, Copy & Code. A series of projects to re-imagine the future of advertising*.

<http://www.artcopycode.com/campaign/art-copy-code>

IABSpain (2015). *Inversión publicitaria en medios digitales. Resultados 2015*.

[http://www.iabspain.net/wp-content/uploads/downloads/2016/03/Estudio\\_Inversión\\_Publicitaria\\_Medios\\_Digitales\\_2015\\_IAB\\_Spain\\_vReducida.pdf](http://www.iabspain.net/wp-content/uploads/downloads/2016/03/Estudio_Inversión_Publicitaria_Medios_Digitales_2015_IAB_Spain_vReducida.pdf)

López, Belén (2007). *Publicidad emocional: estrategias creativas*. Madrid: ESIC. ISBN: 978 84 7356 488 5

Manovich, Lev (2008). *Software takes command*. New York: Georgetown University.

<http://faculty.georgetown.edu/irvinem/theory/Manovich-Software-Takes-Command-ebook-2008-excerpt.pdf>

Martí, José; Muñoz, Pablo (2008). *Engagement marketing. Una nueva publicidad para un marketing de compromiso*. Madrid: Pearson Educación. ISBN: 978 84 8322 449 6

**Nesamoney, Diaz** (2015). *Personalized digital advertising: How data and technology are transforming how we market*. Old Tappan: Pearson Education. ISBN: 978 0 13 403010 4

**Nichols, Wes** (2013). "Advertising analytics 2.0". *Harvard business review*, v. 91, n. 3, pp. 60-68.  
<https://hbr.org/2013/03/advertising-analytics-20>

**Nielsen** (2015). *Screen wars. The battle for eye space in a TV-everywhere world*.  
<http://www.nielsen.com/content/dam/corporate/us/en/reports-downloads/2015-reports/nielsen-global-digital-landscape-report-march-2015.pdf>

**Papí-Gálvez, Natalia** (2015). "Nuevos medios y empresas innovadoras. El caso de las agencias de medios". *El profesional de la información*, v. 24, n. 3, pp. 301-309.  
<http://dx.doi.org/10.3145/epi.2015.may.10>

**Peppers, Don; Rogers, Martha** (1996). *The one to one future. Building relationships one customer at a time*. New York: Currency. ISBN: 0 385 48566 2

**Percival, John** (2013). *HTML5 advertising*. New York: Springer Sciences+Bussines Media. ISBN: 978 1 4302 4602 2

**Rayport, Jeffrey F.** (2013). "Advertising's new medium: human experience". *Harvard business review*, v. 91, n. 3, pp. 76-82.  
<https://hbr.org/2013/03/advertisings-new-medium-human-experience>

**Roca, Meritxell** (2014). "La transformación de la industria publicitaria en la era digital. Percepciones de los profesionales en Estados Unidos: retos y oportunidades". *Telos: cu-*

*adernos de comunicación e innovación*, n. 99, pp. 64-74.  
[https://telos.fundaciontelefonica.com/seccion=1268&idioma=es\\_ES&id=2014102812250001&activo=6.do](https://telos.fundaciontelefonica.com/seccion=1268&idioma=es_ES&id=2014102812250001&activo=6.do)

**Sheehan, Kim-Bartel; Morrison, Deborah K.** (2009). "The creativity challenge: media confluence and its effects on the evolving advertising industry". *Journal of interactive advertising*, v. 9, n. 2, pp. 40-43.  
<http://dx.doi.org/10.1080/15252019.2009.10722154>

**Solana, Daniel** (2012). *Postpublicidad. Reflexiones sobre una nueva cultura publicitaria en la era digital*. Barcelona: Doubleyou. ISBN: 84 6140 584 6

**Tucker, Catherine E.** (2014). "Social networks, personalized advertising, and privacy controls". *Journal of marketing research*, v. 51, n. 5, pp. 546-562.  
[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1694319](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1694319)  
<http://dx.doi.org/10.1509/jmr.10.0355>

**Turow, Joseph** (2011). *The daily you. How the new advertising industry is defining your identity and your worth*. New Haven/London: Yale University Press. ISBN: 978 0 300 16501 2

**Van-Dijck, José; Poell, Thomas** (2013). "Understanding social media logic". *Media and communication*, v. 1, n. 1, pp. 2-14.  
<http://ssrn.com/abstract=2309065>

**Van-Doorn, Jenny; Hoekstra, Janny C.** (2013). "Customization of online advertising: The role of intrusiveness". *Marketing letters*, v. 24, n. 4, pp. 339-351.  
<http://dx.doi.org/10.1007/s11002-012-9222-1>

es flexible      es Diversidad      es Conocimiento

Es compromiso      es un Imán

Además ahora **SEDIC** es más por menos  
Consulta las nuevas tarifas de nuestros cursos en  
[www.sedic.es](http://www.sedic.es) y  
**¡RECÍCLATE!**

# ANÁLISIS DE LA RECOMENDACIÓN ENTRE IGUALES EN LA REPUTACIÓN ONLINE DE LAS ORGANIZACIONES

Analysis of peer recommendations in the online reputation of organizations

Lucía Benítez-Eyzaguirre



**Lucía Benítez-Eyzaguirre**, profesora de la *Universidad de Cádiz*, es doctora por la *Universidad de Sevilla* (con premio extraordinario) y premio RTVA a la mejor tesis doctoral de comunicación. *Master en tecnologías digitales y sociedad del conocimiento*. Directora de la revista científica *Redes.com*. Miembro del grupo de investigación *Compóliticas*, del campo de la comunicación para el desarrollo. Participa en proyectos de investigación y de enseñanza del periodismo y las nuevas tecnologías en Marruecos, Guinea Ecuatorial y Cabo Verde.  
<http://orcid.org/0000-0002-8879-7979>

*Universidad de Cádiz, Departamento de Marketing y Comunicación*  
Campus Universitario de Jerez, Avda. de la Universidad. 11405 Jerez de la Frontera (Cádiz), España  
[lucia.benitez@gm.uca.es](mailto:lucia.benitez@gm.uca.es)

## Resumen

La reputación corporativa se ve afectada por el creciente impacto de la reputación online, especialmente por el peso que adquiere la recomendación entre iguales en los medios sociales. El prosumidor (McLuhan; Nevitt, 1972; Toffler, 1981; Tapscott, 1997) tiene una autonomía y capacidad en la Red que le hace incorporarse a los agentes implicados como un emergente surgido del contexto digital, escasamente teorizado en relación con la reputación corporativa. En el contexto de la definición de estos conceptos, este artículo propone modelos de análisis de la reputación online en función de las acciones comunicativas de los prosumidores, desde un enfoque cualitativo para abordar el conocimiento de los públicos y de los intervinientes. Para ello, realiza una revisión teórica crítica fundamentada sobre fuentes secundarias.

## Palabras clave

Prosumidor; Reputación online; Comunicación organizacional; Recomendación entre iguales.

## Abstract

Corporate reputation is affected by the growing impact of online reputation, because of the weight of peer recommendations in social media. Digital prosumers' (McLuhan; Nevitt, 1972; Toffler, 1981; Tapscott, 1997) autonomy and ability on the internet have led them to integrate with stakeholders. Their role in relation to corporate reputation has barely been recognized. In the context of the definition of these concepts, this article proposes analysis models and tools for online reputation study, based on the communicative actions of prosumers. We take a qualitative methodology as the basic strategy to approach the knowledge of the public and the stakeholders. We carry out a critical theoretical review based on secondary sources.

## Keywords

Prosumer; Online reputation; Organizational communication; Peer recommendation.

**Benítez-Eyzaguirre, Lucía** (2016). "Análisis de la recomendación entre iguales en la reputación online de las organizaciones". *El profesional de la información*, v. 25, n. 4, pp. 652-660.

<http://dx.doi.org/10.3145/epi.2016.jul.15>

## 1. Introducción

Las transformaciones tecnológicas sitúan a las organizaciones y sus públicos en un plano de igualdad en sus intercambios comunicativos, en función del ecosistema horizontal propiciado por internet. Si antes las organizaciones podían

desempeñar el monopolio y la dirección de sus acciones de comunicación, en el actual contexto la estructura iguala la emisión y recepción de mensajes, que ya no son difusionistas sino conversacionales.

Los usuarios son prosumidores, es decir, productores y con-

sumidores de contenidos con capacidad y poder de alterar la reputación de las organizaciones. Sus acciones y conversaciones se desarrollan al margen de las empresas, en un sistema abierto y democratizado de difícil control y gestión para la reputación corporativa, por cuanto la interacción entre iguales (Tapscott; Williams, 2007) y la reputación colectiva operan desde abajo.

Se propone aquí una metodología de análisis de la conversación de los prosumidores como parte de la gestión de la reputación online.

Los usuarios son prosumidores, es decir, productores y consumidores de contenidos con capacidad y poder de alterar la reputación de las organizaciones

## 2. Metodología y objetivos

La reputación online forma parte esencial de la imagen de la empresa, aunque en su definición hay diferentes perspectivas. Justo Villafañe no diferencia el contexto digital en la construcción de la reputación, es decir, defiende el concepto como único por encima del canal en que se registre: “No existe una reputación online y una offline, la reputación es una”.

Sin embargo, Hernández (2011) entiende que es complementario, ya que el circuito de la comunicación digital se suma al que se registra offline: “La reputación online es el reconocimiento que los *stakeholders* (o personas implicadas o intervinientes) de una compañía manifiestan en internet sobre el comportamiento corporativo de la misma a partir del grado de cumplimiento de sus compromisos”. En la práctica, muchos profesionales entienden la reputación online a partir del análisis y la recopilación de datos sobre las opiniones en internet y que, por tanto, sólo es necesario atender a la monitorización. Este enfoque olvida la transformación fundamental del papel de los públicos a partir de la figura del prosumidor, que avanzó Tapscott (1997), y que describía su actividad y su perfil como una figura híbrida entre la producción y el consumo de contenidos, bajo la filosofía de la autonomía y la colaboración. En este sentido se puede entender a los prosumidores como *stakeholders* de las corporaciones y protagonistas activos de la comunicación online.

El concepto de reputación online, en función de los cambios registrados a partir de la web 2.0, precisa una reformulación porque los usuarios registran una conversación multidireccional y compleja que resulta de difícil análisis. Este comportamiento deja obsoleto el concepto de la reputación online basada en la actividad SEO (*search engine optimization*) e incluso el que entiende que la cuestión está relacionada con la monitorización de redes sociales o con la gestión de crisis. Se trata de mucho más, a la vista del papel protagonista que la interacción entre iguales, la valoración y opiniones en redes sociales, la autonomía y poder de los prosumidores en la red. Partiendo de la idea de que la reputación online forma parte de la comunicación corporativa, se plantea aquí la necesidad de mejorar la investigación y el análisis de los comportamientos y prácticas de los usuarios y prosumidores.

Este artículo parte de la hipótesis de que en el contexto de la conversación online la reputación se construye desde abajo con la participación de los prosumidores como usuarios muy activos que contribuyen a la reputación colectiva, a través de la recomendación entre iguales (Williamson; Stayner 1980), un modelo de producción de confianza a partir de la valoración de los públicos. El poder de los usuarios en la construcción del valor de las compañías se debe a su autonomía comunicativa y a su doble condición de productores y consumidores, que obliga a entender a los públicos como parte de los *stakeholders* en la estrategia de las compañías.

Con esta hipótesis, el objetivo de la presente investigación es establecer criterios de análisis sobre la recomendación entre iguales como parte de la comunicación corporativa, así como proponer metodologías cualitativas, más allá de la monitorización cuantitativa de las redes. La recomendación entre iguales (Williamson; Stayner 1980; Light, 2001) se entiende como un modelo de confianza generado por los clientes y usuarios de bienes y servicios a partir de la interacción online. Por tanto, este intercambio de opiniones tiene impacto sobre la reputación de las empresas.

A través del análisis fundamentado sobre fuentes secundarias, de un arco teórico y contextual fruto del análisis documental, se busca una redefinición de los públicos en función de su papel de prosumidores en la construcción de la reputación corporativa. Se describe la evolución del contexto digital para la nueva concepción de la actividad empresarial, de cara a las transformaciones económicas así como a su impacto sobre la gestión del *management*. Se profundiza en el objeto de estudio con la propuesta de modelos de estudio de la recomendación entre iguales, de la participación de los prosumidores en la conversación en red y de la influencia que ejercen sobre la reputación online.

En el contexto de la conversación online, la reputación se construye desde abajo con la participación de los prosumidores como usuarios activos que contribuyen a la reputación colectiva

## 3. La reputación online en las relaciones públicas

Con la web 2.0 y la aparición del prosumidor como figura híbrida en el contexto de la convergencia multicanal de internet, se registra una nueva fuente de creación de valor en la conversación global. McLuhan y Nevitt (1972) describieron por primera vez y con carácter prescriptor el comportamiento de este tipo de usuario. Toffler (1981) anticipó después el fin de los grandes medios masivos y de los modelos de control en la comunicación, como un cambio irreversible en el que los prosumidores tienen capacidad para incidir y transformar los mercados.

La acción conjunta de los públicos invierte el modelo de mando desarrollado en la gestión comunicativa, ya que se entiende que la reputación se impulsa también desde abajo por el efecto de la participación de los usuarios en la conversación online. En el campo de la comunicación,

las redes sociales han ocasionado la ruptura del tradicional paradigma de la comunicación de Lasswell para incorporar una estructura adecuada a la convergencia y la multidireccionalidad horizontal, con integración de los públicos en tanto que prosumidores y nuevos *stakeholders* en las corporaciones.

La reputación online forma parte en sí misma de la reputación corporativa, por lo que debe entenderse y valorarse según los mismos criterios. Dada la importancia creciente de la recomendación entre iguales en la reputación online y, por tanto, corporativa, se trata aquí de presentar criterios de evaluación y análisis de la actividad de los prosumidores a partir de las siguientes cuestiones esenciales:

3.1. La digitalización ha supuesto cambios económicos y sociales definitivos en la relación comunicativa con los *stakeholders* y los públicos, que debe ser contemplada de cara a la conceptualización de la reputación online.

3.2. El papel de la recomendación entre iguales dentro de la reputación online y corporativa, en el nuevo ecosistema comunicativo, tiene un peso decisivo por el efecto de las acciones y la experiencia de los prosumidores sobre la reputación de las compañías.

3.3. La evaluación de la reputación corporativa precisa de la inclusión de la conversación y la recomendación entre iguales por el valor que tienen para la estrategia corporativa, a través de la participación y la colaboración de los prosumidores.

3.4. Los modelos de análisis de la reputación online deben contemplar valores cualitativos para describir las acciones de los usuarios ya que son una fuente de conocimiento esencial para la gestión de la reputación social de las empresas.

3.5. La reputación corporativa suma los cauces online y offline porque se funden en una única experiencia, que se realimenta de las acciones y las prácticas de los públicos. De la misma forma que la convergencia propuesta en los procesos de digitalización, la integración de la comunicación digital y presencial dentro de una misma estrategia es la forma coherente de analizar estos procesos.

La recomendación entre iguales se entiende como un modelo de confianza generado por los clientes y usuarios de bienes y servicios a partir de la interacción online

Por tanto, a partir de estas cuestiones pasamos a profundizar en las transformaciones económicas y sociales con impacto en la gestión corporativa.

#### 4. Empresas y organizaciones en el contexto digital

La web 2.0 dibuja un escenario horizontal, participativo y más igualitario. Las transformaciones que describió **Himanen** (2002) sobre el espíritu de la sociedad de la información proporcionan las claves para el ejercicio del liderazgo y de la

gestión empresarial, ya que la lógica colaborativa y la participación en la toma de decisiones facilitan la competitividad colectiva.

De una forma más concreta, y aplicado al contexto empresarial, **Tapscott y Williams** (2007) habían descrito el impacto económico de la inteligencia colectiva en su libro *Wikinomics*, a partir de las experiencias de colaboración, producción, y la influencia en el valor de los productos y servicios. Los blogs, redes sociales y otras plataformas activan una conversación permanente que afecta a la reputación de las organizaciones. **Tapscott y Williams** (2007) también definieron la 'interacción entre iguales' desde la paridad en el acceso a la información y la autoorganización. El modelo ha conducido a la creación colectiva de soluciones y productos que van desde la elaboración de software libre hasta la *Wikipedia*, todo un mundo de opciones desde la acción colectiva.

La reputación corporativa suma los cauces online y offline porque se funden en una única experiencia, que se realimenta de las acciones y las prácticas de los públicos

La *wikinomía* describe la estructura social y productiva de intercambios y de colaboración sin mando, a partir de las conexiones entre el interior de las organizaciones y las redes externas, en una estructura que permite otros modos de interacción, de producción y de trabajo, según los principios de la economía digital (apertura de los sistemas, interacción entre iguales, compartir y colaborar, y capacidad de actuación global). Todas ellas muestran tendencias opuestas a las que se han venido defendiendo dentro de la economía física y mercantil durante las décadas anteriores a la irrupción de internet. En estos principios se basan los modelos de colaboración masiva como nuevos prototipos productivos pero también son modelos de influencia para la reputación corporativa.

El *Manifiesto Cluetrain* (1999) propuso cambios en las empresas y organizaciones para responder a las nuevas prácticas de comunicación que influyen en los mercados. También en el caso español, el *Manifiesto Crowd* (**Gutiérrez-Rubí; Freire**, 2013) concreta los elementos esenciales en la transformación de la cultura corporativa con el impulso de la comunicación, de otros modos de organización y del reconocimiento del talento o de la meritocracia. Estos elementos coinciden con modos de relación de un sistema abierto, con otra forma de implementar productos y servicios, y establecen dinámicas a considerar en la gestión de la comunicación y de la reputación corporativa.

Estamos por tanto ante una nueva cultura empresarial y económica, en la que la conectividad obliga a rediseñar las estrategias de la comunicación corporativa y a cuestionar las métricas dominantes del análisis del comportamiento de los públicos. Este escenario acaba con las jerarquías y con el control de la comunicación, así como con la estructura vertical desde la que a menudo se ha gestionado la reputación corporativa.

## 5. La web 2.0 rompe el equilibrio de la comunicación corporativa

O'Reilly (2005) definió la web 2.0 a partir de un software que ya no precisa de actualizaciones ni va unido a un dispositivo único porque es una plataforma en la que las experiencias de los usuarios perfeccionan los servicios ofertados. La inteligencia colectiva (Lévy, 2004) como concepto de la cultura digital hace referencia al concurso plural y la colaboración para la toma de decisiones, la explotación de la información de internet, la comparación y el estudio de los comportamientos colectivos, para la predicción y aprovechamiento de grandes datos en la toma de decisiones. Por tanto, la irrupción de la web 2.0 supuso cambios estructurales y tecnológicos como un ecosistema comunicativo y de acciones que ahora se caracteriza por la ausencia de mando. La estructura de la comunicación digital ha impuesto un proceso de desintermediación en las actividades empresariales, comerciales, sociales, económicas y políticas. Se rompe la cadena tradicional entre el emisor y el receptor, la relación entre los productores de contenidos y los receptores. Con el fin de la intermediación acaba el papel protagónico de los medios de comunicación y surge otro orden comunicativo de relaciones sociales y colectivas.

La conectividad obliga a rediseñar las estrategias de la comunicación corporativa y a cuestionar las métricas dominantes del análisis del comportamiento de los públicos

A partir de la web 2.0 también los buscadores pierden su papel protagonista en el acceso a contenidos para dar paso a las redes sociales. Si durante el primer desarrollo de la web el posicionamiento en buscadores o SEO (*search engine optimization*) ha sido esencial para mejorar la visibilidad y la notoriedad de las corporaciones dentro de internet, ahora la situación es otra, sobre todo si se trata de estudiar la reputación, que está más relacionada con el cumplimiento de acciones y compromisos.

El *Manifiesto Cluetrain* defiende la apertura y la horizontalidad de las empresas para adaptarse a la estructura que ha impuesto la red, en la que las organizaciones dependen de los vínculos para obtener como ventaja competitiva la inteligencia colectiva. El principio de apertura (Tapscott; Williams, 2007) exige cambios en el concepto de secreto empresarial, pues se entiende que el intercambio de información y la transparencia reducen los costes empresariales y, a la vez, proyectan una imagen de confianza. Por ello, el acceso a la información corporativa y el fin de los secretos forman parte de la filosofía de la cultura digital, y apuntan a otras tendencias en la recepción de la publicidad y en la gestión de las relaciones públicas, pues los usuarios con habilidades y competencias para obtener información directa exigen transparencia.

El modelo social y económico resultante lleva a considerar a los mercados como conversaciones, en las que las empresas se comunican de forma directa y abierta dentro de un sistema

que se enriquece y gana valor añadido a partir de la experiencia (Levine *et al.*, 1999). De cara a la comunicación, los autores recomiendan pautas diferenciales de comportamiento de las organizaciones, como sería la construcción de una masa crítica para recoger los beneficios que, a largo plazo, se producen. De hecho, el *Manifiesto Cluetrain* incorpora también la acción de los públicos como prosumidores, es decir, como parte del mercado y de la producción. En coherencia con la teoría de las relaciones públicas, se entiende que el fin último del proceso de comunicación y de alianzas está en establecer un diálogo entre organizaciones y públicos.

La web 2.0 dibuja un escenario horizontal, participativo y más igualitario

## 6. La comunicación horizontal y el prosumidor como mensaje en la reputación online

Las prácticas comunicativas de los prosumidores se desarrollan en un "espacio de encuentro y transformación en donde se articulan las percepciones plurales de la realidad" (Pérez; Massoni, 2009, p. 300). Esta visión compartida es definitiva en la comunicación y, aunque forma parte del capital reputacional de las empresas, se gestiona, cada vez más, al margen de estas instituciones y de las jerarquías (Tapscott; Williams, 2007, p. 80).

La interacción entre iguales configura comunidades colaborativas y autónomas –a menudo alrededor de gustos y preferencias– que vienen mostrando su eficacia en la gestión del conocimiento, pero también orientando consumos culturales como la lectura, la música o el cine. El prosumidor participa en los cambios culturales y transforma la relación entre producción y consumo, y los valores de marca (Toffler, 1981, p. 269), y con ello la reputación de las empresas y de sus productos, aun cuando no los consuma. En conjunto, las opiniones y valores de los prosumidores conducirían a la recomendación colectiva (*crowdcuration*), un concepto que surge del comercio electrónico como mecanismo de producción de confianza sobre las decisiones de compra en el que se apoyan muchos modelos de negocio como *eBay* o *TripAdvisor*.

La web social ha popularizado estos modelos de producción de confianza a partir de algoritmos que describen la información y las prácticas de los usuarios, que analizan sus preferencias y filtran la información, las opiniones y características de sus elecciones para orientar a los iguales. La monitorización del entorno y el valor de la inteligencia colectiva tienen capacidad para alterar y producir valor sobre los contenidos y la reputación online. Se entiende que la inteligencia colectiva puede superar el posible sesgo de los expertos a la hora de analizar la información de toda una organización, de cara a comprender comportamientos colectivos o a realizar predicciones. De hecho, la extracción de datos masivos o *big data* es una fuente de información relevante para las organizaciones.

Cabe aquí contemplar la capacidad de los usuarios para la transformación de los valores de marca de las empresas, que obliga a prestar atención a su impacto sobre la reputa-

ción online, como un desarrollo innovador de la digitalización. El modelo de la recomendación entre iguales resulta más solvente conforme los prescriptores sean conocidos o estén valorados en la red, o bien por el valor colectivo de la recomendación social. El concepto, en el ámbito de las relaciones públicas y la comunicación, procede del clásico boca-oído y en el mundo digital se conoce como eWoM (*electronic word of mouth*) (Westbrook, 1987), que contempla la comunicación informal entre consumidores sobre las empresas, sus servicios y productos. El valor de estas prescripciones puede ser superior al de los medios de comunicación ya que, además de estar libre de intereses comerciales, se autorregulan de forma abierta y transparente. En este sentido, forma parte de la tendencia de colaboración que mejora los servicios de las plataformas online (San-Millán-Fernández; Medrano-García; Cavarkapa, 2009, p. 4).

Con sus prácticas de comunicación, los usuarios terminan tejiendo una red, cuya acción colectiva y opiniones son la expresión de su poder en el mercado y, por tanto, de su impacto económico. Los prosumidores se erigen en jueces del desempeño organizacional, de la coherencia entre las acciones y la comunicación de las corporaciones, son testigos del compromiso empresarial con el entorno. Las experiencias y la valoración de los productos y bienes producen el entramado social de la reputación online que condiciona las percepciones de los públicos.

El peso del fenómeno, a la vista de las prácticas y de la acción colectiva, y del creciente número de plataformas que explotan la recomendación entre iguales, debe conducir a un cambio en la orientación de la estrategia de comunicación porque, por encima del efecto y la influencia que la gestión profesional pueda ejercer sobre los mensajes, la de la multitud inteligente logra mayor credibilidad y confianza en la reputación.

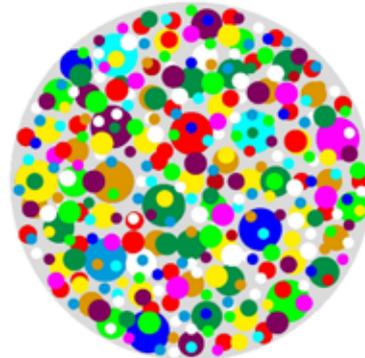
El intercambio de información y la transparencia reducen los costes empresariales y, a la vez, proyectan una imagen de confianza

## 7. Estrategia de una reputación online

En la web social, la interactividad propone un diálogo y un intercambio de opiniones en los que la escucha es también parte del proceso. Las organizaciones se deben insertar en el contexto social con apertura y flexibilidad para aprovechar las aportaciones de la inteligencia colectiva, de la innovación y creatividad. Así se entiende que, de cara a las relaciones públicas, más allá que publicidad y estrategia, el enfoque ha de ser el de la creación de una comunidad. En la construcción de la reputación online, son clave la participación en la conversación, los contenidos con acceso y enlaces, el análisis de los mensajes, la respuesta a los problemas y el estudio de la satisfacción. Los valores de la comunicación corporativa se transforman de acuerdo con el ecosistema digital en el que la cooperación es la base de la creación de alianzas.

# Manifiesto Crowd

La empresa y la inteligencia de las multitudes



#mc

Antoni Gutiérrez-Rubí  
Juan Freire

con la coordinación de Ricard Espelt  
y prólogo de Francis Pisani

<http://www.gutierrez-rubi.es/2013/03/18/manifiesto-crowd-libro-y-manifiesto>

El prosumidor avala la recomendación de productos y servicios con voz propia, autónoma e independiente. Se transforma en influenciador en la medida en que su criterio sea independiente, y fruto de su capacidad de obtener información en la Red, de su propio conocimiento o de la valía de sus opiniones. En el mundo 2.0, la influencia de los usuarios, la posibilidad de convertirse o ejercer de líder es muy superior a la del mundo físico.

En su origen, se detectó esta tendencia especialmente para el caso de los consumos culturales, como lecturas o cine, donde son decisivos factores estéticos, intelectuales o emocionales (Stebbins, 1996, p. 948). El mensaje emocional es muy influyente y debe ser considerado en la gestión de la comunicación online, más todavía si tiene clave empática y si se enfoca hacia la mejora de la reputación, tal y como propone Cleminson (2012, p. 119):

“La empatía y su compañía, la emoción, son reclamadas para mejorar nuestros sentimientos hacia instituciones y entidades impersonales que están dañadas y sufren como el mercado”.

Esta emoción social está en el origen de las visiones colectivas y de las relaciones. Por ello, el *Manifiesto Crowd* recomienda las relaciones emocionales con los *stakeholders* o participantes involucrados. En las percepciones se transforma el conocimiento en un sistema de necesidades que determina la acción, pero también por la relación emocional con el proceso y el objeto (Alloza, 2010, p. 134). La influencia social produce comportamientos imitativos, en los que se adoptan formas y expresiones similares a las que han conocido de otros iguales (Bonchi, 2011). De alguna forma,

las opiniones y valoración de los prosumidores dentro del mundo online conducen a experiencias emocionales offline, reactivando la necesaria relación entre las dos esferas.

La participación de los prosumidores en la difusión de contenidos, tanto en intensidad como en alcance –fruto de la viralidad–, logra una repercusión más allá de los círculos y las redes más próximas. El formato de esta comunicación es narrativo en la mayoría de los casos, dentro de la estrategia de *storytelling* (Salmon, 2008) con la que las organizaciones afrontan en un escenario líquido de comunicación relatos que fraguan la imagen corporativa. Al *storytelling* empresarial se suman las narrativas creativas y autónomas de los prosumidores, en un sistema abierto de democratización de la experiencia que salta las barreras entre medios hasta consolidar la estrategia *cross media*, donde el mensaje circula por diferentes medios y formatos (Lacalle; Sánchez-Navarro, 2012, p. 2). Cuando los prosumidores tienen un gran poder, cuestionan la autoridad y buscan nuevas formas de relación, las organizaciones se enfrentan a la dificultad de cumplir con una estrategia previa de comunicación y, a menudo, con la resistencia por parte de los profesionales del sector e incluso con la falta de formación en estrategias digitales para gestionar la reputación online.

Los usuarios terminan tejiendo una red, cuya acción colectiva y opiniones son la expresión de su poder en el mercado y, por tanto, de su impacto económico

La escucha se ha venido definiendo como un proceso fundamental de las relaciones públicas, y lo es más aún en el actual contexto. La autonomía e independencia de las recomendaciones obliga a las empresas a un ejercicio permanente de monitorización y escucha, a partir del cual se puede establecer un diálogo, dentro de la desintermediación digital. Como venimos defendiendo, los prosumidores forman parte de los *stakeholders* de las organizaciones y su intervención es crucial en la generación del valor añadido.

La estrategia de la comunicación online analizada desde la escucha se basa en tres grados de proximidad y de conversación con los públicos:

- el primero, que se podría calificar de pasivo, en el que la información canaliza los intercambios y que considera a los públicos sólo en sus prácticas de recopilación de la información;
- el segundo, desde una mayor implicación en la escucha planteada, ofreciendo respuestas de la organización a los comentarios y valoraciones de los públicos;
- en el último, además de la escucha y la respuesta, la organización promueve y dinamiza los comentarios y la comunicación con los públicos, estableciendo un diálogo como una posición de alto compromiso.

Por tanto, se trata del ámbito donde se construye la reputación ‘desde abajo’ en confluencia entre las prácticas online y offline, estableciendo vínculos de confianza.

## 8. Valores de la recomendación entre iguales

El modelo de creciente influencia social de la recomendación entre iguales presenta muchas dificultades para la gestión desde la estrategia corporativa. La toma de conciencia del poder de los prosumidores ha conducido en la mayoría de los casos a una gestión orientada a los riesgos para la reputación corporativa, pero a menudo se olvida el objetivo fundamental de consolidar las relaciones con los diferentes grupos de interés, que sería el objetivo de las relaciones públicas. Aceptar las prácticas sociales, el acceso a los contenidos a través de la web social, el diálogo horizontal e igualitario, así como la atención a las demandas de transparencia, al compromiso del equilibrio entre intereses económicos y sociales son aspectos esenciales de la reputación online.

El estudio de las prácticas de los usuarios es básico para comprender las claves de la comunicación distribuida en el contexto digital, pero también para valorar su impacto y participar en la conversación dentro de la lógica de la influencia de las relaciones públicas. Así, para el análisis del impacto de la recomendación entre iguales en la reputación online hay que estimar estos factores:

- La actividad y comunicación de los prosumidores debe ser el centro del análisis. Es más, en sí mismo, el prosumidor es una categoría de análisis por cuanto, como venimos defendiendo, es a la vez canal y mensaje de la comunicación corporativa en el modelo descentralizado de la red.
- El valor de la conversación, de la interactividad, de la autonomía y de la transparencia de las prácticas de los prosumidores, en un contexto social en el que las empresas deben formar parte de la conversación.
- La comparación de los diferentes prescriptores para determinar el papel de los influenciadores en la creación de esta reputación por el valor y el alcance de su recomendación.
- La importancia de la recomendación entre iguales en los procesos de toma de decisiones en contextos complejos, como el de la comunicación multilateral y no organizada de la red.
- El análisis de los contenidos, especialmente de los producidos por los públicos, en la toma de decisiones, especialmente para el caso en que estas propuestas de los usuarios muestren elementos negativos sobre la reputación corporativa.
- La comparación de los efectos en la reputación corporativa de la recomendación entre iguales con el análisis tradicional de los medios de comunicación. A pesar de su importancia, en la gestión de la comunicación se viene realizando un seguimiento exhaustivo de los medios en lugar del estudio de los comportamientos de los prosumidores.

En conjunto, las prácticas de los usuarios en la Red se pueden seguir a través de mecanismos de evaluación de los prescriptores, de su reputación y de la producción de confianza. En sí mismo, se convierte en un modelo de negocio sólido en internet, basado en las recomendaciones, porque

explota la suma de las prescripciones individuales y el valor de su confianza con sistemas de extracción de datos o *web mining* –incluso en tiempo real– para producir credibilidad y orientar decisiones de compra o elecciones de servicios.

Desde el *management* de la comunicación corporativa con frecuencia se analizan datos de corte cuantitativo a pesar de que no reflejan prácticas complejas de relaciones, de intereses y de valores que sintetizan la interactividad los usuarios. Aunque las aplicaciones de monitorización (que pueden ayudar al estudio de los comportamientos) más habituales muestran resultados inmediatos, no son suficientes para un estudio significativo; se hace imprescindible un giro en el estudio de la comunicación online que prime la información cualitativa, por cuanto ésta es mucho más explicativa sobre elementos esenciales de la actividad de los públicos, al contribuir a desvelar las emociones, motivaciones, expectativas y experiencias en su relación con las corporaciones.

La autonomía e independencia de las recomendaciones obliga a las empresas a un ejercicio permanente de monitorización y escucha, a partir del cual se puede establecer un diálogo

## 9. Análisis de la recomendación entre iguales para la reputación corporativa

A la hora de analizar la reputación online hay que diferenciar entre la monitorización, y la investigación y análisis de la información obtenida mediante *web mining*. En conjunto, el proceso de extracción de información se realiza a través de las palabras clave con las que se filtran los contenidos en todo tipo de formatos para recopilar los comentarios de los usuarios. El proceso inicial de monitorización detecta estas *keywords* con el filtrado de corte sintáctico, que no es suficiente para entender las opiniones y la valoración sobre los productos y servicios o los procesos de influencia que se registran sobre la imagen corporativa. Para el análisis y la investigación sería necesario que las referencias tuvieran valor semántico y, por tanto, permitieran el análisis cualitativo.

Cada una de estas estrategias responde a los desarrollos de la web y a su evolución. En primer lugar, la extracción de información a través de algoritmos de búsqueda, propia de la web 1.0, que condujo a estrategias SEO de posicionamiento que influyen en la notoriedad pero no en la reputación online. El uso social de internet también ha modificado las fórmulas de estos algoritmos, al incorporar a partir de 2009 las búsquedas sociales, es decir, las procedentes de redes sociales en función del perfil de usuario, las cuales, hoy por hoy, son uno de los factores de peso en el posicionamiento online.

En segundo lugar, el *web mining* o minería de datos, fruto de la web 2.0, de donde se obtiene información relevante para estudiar la recomendación, la agregación a grupos por gustos o aficiones, la clasificación y la puesta en valor de la información obtenida, es decir, permite obtener conocimiento colectivo.

El desarrollo de la web semántica o 3.0 se basa en estos resultados, en su capacidad descriptiva y también predictiva sobre los comportamientos individuales y colectivos. La combinación de técnicas inteligentes conduce a la minería de opiniones, al desarrollo de sistemas automáticos de procesamiento de información sobre las opiniones de usuarios expresadas en la Red.

Desde entonces, los principales buscadores como *Google* han modificado sus fórmulas con el objetivo de aproximarse a la búsqueda inteligente, a partir de la semántica y del procesamiento del lenguaje natural. Dada la mezcla y complejidad de la información que se puede recopilar, el reto se encuentra en lograr información significativa sobre las opiniones de los usuarios que resulte útil para el desarrollo de nuevos productos y servicios o bien para mejorar la reputación de las organizaciones. El resultado de esta evolución es un modelo más complejo de minería de opiniones de carácter cualitativo que permite conocer, además de los gustos de los consumidores, el impacto de una campaña o de la imagen de marca. Esta información es muy importante para que las organizaciones puedan reaccionar rápidamente y atajar las malas opiniones o los daños a la imagen corporativa.

La elección y la valoración colectiva y social de productos y servicios se sistematiza en algoritmos de recomendaciones, de aficiones comunes, para organizar la información o el conocimiento sobre determinadas materias, comprender la actividad y los comportamientos colectivos. Los algoritmos de recomendación se diseñan con diferentes criterios: en función del contenido, del contexto o de las prácticas sociales, a partir del análisis de la comunicación interactiva de los sujetos o bien de las etiquetas. Así se puede distinguir entre el tipo y la temática de los comentarios, el contexto de la comunicación como factor que influye en la recomendación y la influencia social en los procesos colaborativos.

De todas estas opciones cabe destacar los dos grandes modelos de análisis de los comportamientos:

- de una parte, el que se realiza sobre el contenido generado por los usuarios –comentarios, etiquetas, discusiones, interacciones sociales y recomendaciones–, y
- de otra, el estudio de las tareas colaborativas –producción de contenido, desarrollo de software libre o juegos–, que puede desembocar en otra modalidad de predicción de los comportamientos de los usuarios.

En el mar de información que proporciona internet, la metodología debe estar orientada a la búsqueda de significados, que interprete la información de corte cualitativo en función de las palabras clave o, mejor, de segmentos textuales de interés para la evaluación semántica. La información obtenida puede ser codificada según los criterios de la investigación para analizar elementos que afectan a la reputación corporativa, *insights* de interés para la gestión de la comunicación. El análisis complejo de esta información contempla que los comentarios y valoraciones no tienen necesariamente que ser coherentes ni responder a objetivos concretos, pero sí tener en consideración el aspecto emocional y clasificar los sentimientos, e incluso incorporar técnicas de aprendizaje al propio sistema.

La metodología debe incluir una sistematización progresiva.

- Primero, se debe definir el problema de análisis en función de los resultados previsibles y de distorsiones como el sentido del humor, la ambigüedad, las expresiones informales o la geolocalización.
- En segundo lugar, ha de plantear la extracción automática de información y de comentarios relacionados con la empresa o su actividad, así como sobre la competencia, a través de un proceso selectivo entre blog, foros, redes sociales o *Twitter*.
- En tercer lugar, debe realizar la clasificación de las opiniones según el análisis semántico (Balahur; Montoyo, 2009), de las emociones en las opiniones (Wiebe; Wilson; Cardie, 2005), la desambiguación (Gutiérrez et al., 2010), la informalidad de los textos (Mosquera; Moreda, 2011) y la geolocalización de las opiniones. Todo ello se presenta en un resumen eficaz para la reputación online, en función del comportamiento de los prosumidores y usuarios, y de su libre expresión de corte naturalista, que adquiere significado en su investigación.

## 10. Conclusiones

Si entendemos la figura del prosumidor como *stakeholder* de las organizaciones, el interés sobre sus acciones en la comunicación y su influencia en la reputación corporativa es definitivo, por cuanto se trata de un agente esencial en la propia organización. De ahí la necesidad de aplicar metodologías cualitativas al estudio de la conversación en las redes sociales, por la autonomía e independencia con la que los prosumidores afectan a la reputación corporativa. La perspectiva para este estudio debe ser un planteamiento integrador e interdisciplinar, incorporando a especialistas de diferentes perfiles para atender a la complejidad y el carácter holístico de la reputación corporativa. Nos referimos a:

- expertos en comunicación con formación tecnológica para estudio de la interactividad y de los comportamientos;
- tecnólogos que mejoren las posibilidades y las herramientas de análisis;
- perfiles analíticos y de investigación social que interpreten los grandes volúmenes de datos; y
- economistas que a partir del análisis multidisciplinar orienten a la obtención de valor y la cuantificación de resultados.

Es necesario desarrollar una nueva filosofía en la evaluación de la información obtenida y en su análisis dentro de la estrategia de la comunicación corporativa, para centrarse en el valor de las opiniones y prescripciones de los prosumidores que, de forma desinteresada, proponen en función de su experiencia para la orientación de otros consumidores.

El seguimiento y análisis de la recomendación entre iguales permite intervenir en la conversación y dar muestras del grado de apertura corporativa, de mayor adaptación al contexto digital: es una oportunidad para la gestión relacional entre la organización y los *stakeholders*. Además, hay que insistir en que la recomendación entre iguales aporta información esencial para lograr la mayor coherencia entre las acciones de las organizaciones y los compromisos públicos adquiridos.

## 11. Bibliografía

- Alloza, Ángel** (2010). "De Maquiavelo a la Neurociencia: las marcas que queremos". En: Villafañe, Justo (ed.). *La comunicación empresarial y la gestión de intangibles en España y Latinoamérica*. Madrid: Prentice Hall, pp. 131-171. ISBN: 978 8483227121
- Bonchi, Francesco** (2011). "Influence propagation in social networks: A data mining perspective". *IEEE intelligent informatics bulletin*, December, v. 12, n. 1, pp. 8-16. [http://www.comp.hkbu.edu.hk/~iib/2011/Dec/article1/iib\\_vol12no1\\_article1.pdf](http://www.comp.hkbu.edu.hk/~iib/2011/Dec/article1/iib_vol12no1_article1.pdf)  
<http://dx.doi.org/10.1109/WI-IAT.2011.286>
- Cleminson, Richard** (2012). "La emotividad del mercado". *Teknokultura*, v. 9, n. 1, pp. 117-121. <http://teknokultura.net/index.php/tk/article/view/32>
- Gutiérrez, Yoan; Fernández, Antonio; Montoyo, Andrés; Vázquez, Sonia** (2010). "UMCC-DLSI: Integrative resource for disambiguation task". En: *Procs of the 5th Intl workshop on semantic evaluation*, pp. 427-432. Uppsala, Sweden. ISBN: 978 1932432701  
<http://www.aclweb.org/anthology/S10-1095>
- Gutiérrez-Rubí, Antoni; Freire, Juan** (2013). *El manifiesto Crowd*. Laboratorio de tendencias. ISBN: 978 8493906672  
<http://laboratoriodetendencias.com>
- Hernández, Jerusalem** (2011). "Reputación online: Necesidad de un marco conceptual y modelo de gestión". En: Villafañe, Justo (ed.). *La comunicación empresarial y la gestión de intangibles en España y Latinoamérica*. Madrid: Prentice Hall, pp. 197-214. ISBN: 978 8483227121
- Himanan, Pekka** (2002). *La ética del hacker y el espíritu de la era de la información*. Barcelona: Destino. ISBN: 9788423336371  
<http://eprints.rclis.org/12851/1/pekka.pdf>
- Lacalle, Charo; Sánchez-Navarro, Jordi** (2012). "Audiovisual 2.0: narrativas, recepción y consumo en los nuevos hipertextos". *Anàlisi. Quaderns de comunicació i cultura*, n. extra 1, pp. 1-3. <https://goo.gl/qDEE2i>
- Levine, Rick; Locke, Christopher; Searls, Doc; Weinberger, David** (1999). *The Cluetrain manifesto. The end of business as usual*. New York: Cluetrain. ISBN: 978 0738204314  
<http://www.cluetrain.com>
- Lévy, Pierre** (2004). *Inteligencia colectiva*. Washington: Organización Panamericana de la Salud. ISBN: 2707126934  
<http://inteligenciacolectiva.bvsalud.org/public/documents/pdf/es/inteligenciaColectiva.pdf>
- Light, David A.** (2001). "Sure, you can trust us". *Sloan management review*, v. 43, n. 1. <http://sloanreview.mit.edu/article/ebusiness-sure-you-can-trust-us>
- McLuhan, Marshall; Nevitt, Barrington** (1972). *Take today: The executive as dropout*. New York: Harcourt Brace Jovanovich. ISBN: 978 0151878307
- Mosquera, Alfonso; Moreda, Paloma** (2011). "Caracterización de niveles de informalidad en textos de la web 2.0".

*Procesamiento del lenguaje natural*, n. 47, pp. 171-177. Alicante: Sociedad Española para el Procesamiento del Lenguaje Natural, Universidad de Alicante. [https://rua.ua.es/dspace/bitstream/10045/18525/1/PLN\\_47\\_18.pdf](https://rua.ua.es/dspace/bitstream/10045/18525/1/PLN_47_18.pdf)

**O'Reilly, Tim** (2005). *What is Web 2.0. Design patterns and business models for the next generation of software*. <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>

**Pérez, Rafael-Alberto; Massoni, Sandra** (2009). *Hacia una teoría general de la estrategia. El cambio de paradigma en el comportamiento humano, la sociedad y las instituciones*. Madrid: Ariel. ISBN: 9788434413108

**Salmon, Christian** (2008). *Storytelling. La máquina de fabricar historias y formatear las mentes*. Barcelona: Ediciones Península. ISBN: 9788483078358

**San-Millán-Fernández, Elvira; Medrano-García, María-Luisa; Cavarkapa, Branko** (2009). "Identidad y reputación digital: metodología en MK.20". En: *XXIII Congreso anual de la European Academy of Management and Business Economics*. Sevilla: Aedem. ISBN: 978 8473566094

**Stebbins, Robert** (1996). "Cultural tourism as serious leisure". *Annals of tourism research*, v. 23, n. 4, pp. 948-950. [http://dx.doi.org/10.1016/0160-7383\(96\)00028-X](http://dx.doi.org/10.1016/0160-7383(96)00028-X)

**Tapscott, Don** (1997). *La economía digital*. Colombia:

McGraw Hill; Interamericana de Colombia. ISBN: 978 9586005616

**Tapscott, Don; Williams, Anthony D.** (2007). *Wikinomics: La nueva economía de las multitudes inteligentes*. Barcelona: Paidós Ibérica. ISBN: 978 8449320149

**Toffler, Alvin** (1981). *La tercera ola*. Barcelona: Plaza & Janés. ISBN: 9788401374500 <https://cruceshernandezguerra.wikispaces.com/file/view/La+tercera+ola.pdf>

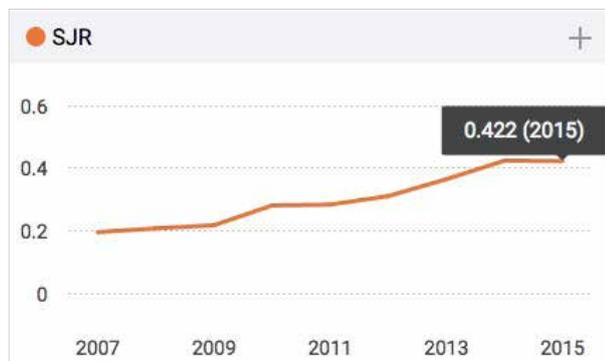
**Westbrook, Robert A.** (1987). "Product/consumption-based affective responses and postpurchase processes". *Journal of marketing research*, v. 24, n. 3, pp. 258-270. <http://dx.doi.org/10.2307/3151636> <http://www.jstor.org/stable/3151636>

**Wiebe, Janyce; Wilson, Theresa; Cardie, Claire** (2005). "Annotating expressions of opinions and emotions in language". *Language resources and evaluation*, v. 39, n. 2-3, pp. 165-210. <http://www.cs.cornell.edu/home/cardie/papers/lre05withappendix.pdf> <http://dx.doi.org/10.1007/s10579-005-7880-9>

**Williamson, Kirsty; Stayner, Richard** (1980). "Information and library needs of the aged". *The Australian library journal*, v. 29, n. 4, pp. 188-195. <http://www.tandfonline.com/doi/pdf/10.1080/00049670.1980.10755425> <http://dx.doi.org/10.1080/00049670.1980.10755425>

## Actualización del IF y del SJR

En junio de 2016 *SCImago* y *Thomson Reuters* han publicado sendas actualizaciones de los indicadores de impacto de las revistas que indexan en sus respectivas bases de datos.



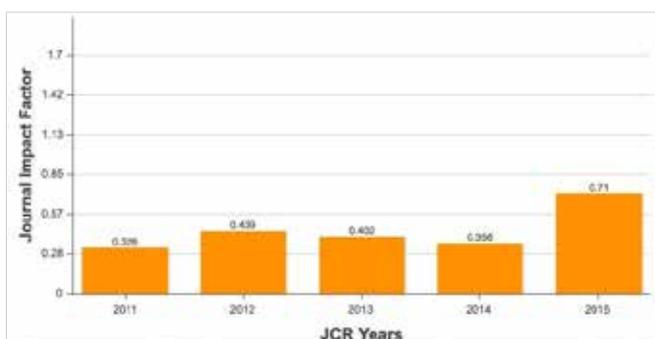
El grupo de investigación español *SCImago*, por encargo de *Elsevier*, elabora el indicador *SCImago Journal Rank (SJR)* basado en las citas que reciben las revistas indexadas en *Scopus*.

Esta revista *El profesional de la información* ha pasado de

SJR 2014 = 0,374 a **SJR 2015 = 0,422 (+12,8%)**

El *Impact Factor (IF)* que elabora *Thomson Reuters* sobre las bases de datos *Science Citation Index* y *Social Sciences Citation Index* (ambas integradas en la *Web of Science*) y que presenta en la base de datos *Journal Citation Reports (JCR)* también muestra un buen incremento de la popularidad de esta revista *El profesional de la información*: hemos pasado de

IF 2014 = 0,356 a **IF 2015 = 0,710 (+99,4%)**



**Damos las gracias a nuestros autores y a nuestros lectores**



# TECNOLOGÍAS PARA LA INCORPORACIÓN DE OBJETOS 3D EN LIBROS DE PAPEL Y LIBROS DIGITALES

Technologies for incorporating 3D objects in paper and digital books



**Carlos Carbonell-Carrera, José-Luis Saorín, Cecile Meier, Dámari Melián-Díaz y Jorge De-la-Torre-Cantero**



**Carlos Carbonell-Carrera** es profesor en la *Universidad de La Laguna*, e invited associate professor por la *University of Applied Life Sciences and Environment* de Iași, Rumanía. Es licenciado en documentación por la *Universidad Oberta de Cataluña* (2008), Master internacional en sistemas de información geográfica por la *Universitat de Girona* (2005) y doctor por la *Universidad de La Laguna* (2011).

<http://orcid.org/0000-0003-4733-1598>

[ccarbone@ull.es](mailto:ccarbone@ull.es)



**José-Luis Saorín** imparte docencia en expresión gráfica en ingeniería y en innovación docente en el master de profesorado en la *Universidad de La Laguna*. Es doctor por la *Universidad Politécnica de Valencia* (2006) y técnico en diseño de programas, entornos y materiales con apoyo de las TIC por la *Universidad Oberta de Cataluña*. Es coordinador del grupo de investigación *Dehaes* de la *Universidad de La Laguna*, entre cuyas líneas de investigación se encuentra el diseño, creación y evaluación de objetos tridimensionales para educación.

<http://orcid.org/0000-0003-3240-3317>

[jlsaorin@ull.es](mailto:jlsaorin@ull.es)



**Cecile Meier** realiza su doctorado en la *Facultad de Educación* de la *Universidad de la Laguna* bajo el título: *Incorporación del patrimonio escultórico en contextos educativos mediante el uso de impresoras 3D y tecnologías avanzadas de bajo coste*. Es licenciada en bellas artes, especialidad de diseño gráfico, con un máster en diseño gráfico por la *Universidad Politécnica de Valencia*. Realizó un master en formación del profesorado, especialidad dibujo diseño y artes plásticas, por la *Universidad de La Laguna* (2012/2013).

<http://orcid.org/0000-0001-7160-2154>

[alu0100305944@ull.edu.es](mailto:alu0100305944@ull.edu.es)



**Dámari Melián-Díaz** finalizó en 2011 sus estudios de arquitectura por la *Universidad de Las Palmas*, posteriormente realizó el máster en formación del profesorado en la especialidad de dibujo, diseño y artes gráficas. En la actualidad, cursa el tercer año de doctorado en educación por la *Universidad de La Laguna*, bajo el título *Aplicación de tecnologías avanzadas de bajo coste como estrategia para la incorporación de competencias artísticas y creativas en los procesos de enseñanza-aprendizaje del dibujo técnico*.

<http://orcid.org/0000-0002-7651-728X>

[damarimd@gmail.com](mailto:damarimd@gmail.com)



**Jorge De-la-Torre-Cantero** es profesor asociado del *Departamento de Técnicas y Proyectos en Ingeniería y Arquitectura* de la *Universidad de La Laguna* desde 2003. Doctor en 2013 por la *Universidad Politécnica de Valencia* e investigador del *Laboratorio de Fabricación Digital* de la *Universidad de La Laguna*. Sus líneas de investigación son: STEAM (*science, technology, engineering, arts, and mathematics*: influencia del arte y el diseño en las disciplinas científico-tecnológicas); tecnologías gráficas avanzadas, fabricación digital y tecnologías BIM (*building information modeling*).

<http://orcid.org/0000-0001-5516-0456>

[jcantero@ull.es](mailto:jcantero@ull.es)

*Universidad de La Laguna*

Avenida Ángel Guimerá Jorge, s/n. 38200 La Laguna (Tenerife), España

Artículo recibido el 21-01-2016  
Aceptación definitiva: 12-02-2016

## Resumen

El libro, como elemento de transmisión de conocimiento, muestra la información mediante texto, imágenes y dibujos. Cuando el texto contiene información de carácter tridimensional o geométrico, para complementar el libro, se utilizan recursos complementarios como objetos tangibles, desplegables o maquetas que facilitan y complementan la interpretación del texto. Con la aparición de los libros digitales es posible incorporar formatos multimedia como vídeos, gráficos interactivos, galerías de fotos u objetos 3D. Este artículo describe las posibilidades de incluir modelos tridimensionales en libros en formato tradicional de papel, así como en libros en versión electrónica tanto para PC como para dispositivos móviles. Para este trabajo se han realizado varios prototipos de libros en varios formatos y se presentan instrucciones para que el lector pueda visualizar los objetos 3D incluidos en los mismos.

## Palabras clave

Códigos QR; Libro digital; Modelos tridimensionales; 3D; *Pop-up*; *Paper toy*; Realidad aumentada.

## Abstract

Traditionally, the book uses text, pictures, and drawings to show information. To complement the information, various methods have been used, such as adding tangible objects or models. With the advent of digital books, it is possible to incorporate videos, interactive graphics, photo galleries, and 3D objects. This article describes the possibilities of including 3D models in books that are paper or electronic (read on a desktop or mobile device). For this article we have made several book prototypes in various formats, and instructions are presented so that the reader can visualize the 3D objects contained in them.

## Keywords

QR codes; Digital book; Three-dimensional models; 3D; *Pop-up*; *Paper toy*; Augmented reality.

**Carbonell-Carrera, Carlos; Saorín, José-Luis; Meier, Cecile; Melián-Díaz, Dámari; De-la-Torre-Cantero, Jorge (2016).** "Tecnologías para la incorporación de objetos 3D en libros de papel y libros digitales". *El profesional de la información*, v. 25, n. 3, pp. 661-670.

<http://dx.doi.org/10.3145/epi.2016.jul.16>

## 1. Introducción

Los libros, en sentido amplio, han sido el formato por excelencia para la difusión de contenidos. El acceso a información digital en línea, a través de foros, sitios webs, documentos auto-editados para su descarga, etc. supera al que se realiza mediante el formato clásico de libro para la lectura. La función del libro tradicional, tanto de texto como divulgativo, sigue teniendo vigencia en pleno siglo XXI para la difusión de ideas y conocimiento. En educación universitaria los libros siguen siendo el recurso más frecuentemente recomendado (Rodríguez-Bravo *et al.*, 2015).

Los libros, uno de los principales vehículos de transmisión de información, deberían poder incluir imágenes tridimensionales en su interior

Al tradicional formato de libro en papel se ha sumado el libro electrónico. Hay múltiples definiciones de libro electrónico: puede considerarse una versión digital de un libro, es decir, cualquier texto legible en pantalla o cualquier fichero en formato digital que puede descargarse en dispositivos electrónicos para su posterior visualización (Rao, 2005; Doctorow, 2004; Berube, 2005; Kang; Wang; Ling, 2009). El libro electrónico, sobre todo si es de texto, también puede ser entendido de forma más compleja, en el contexto de los objetos educativos digitales (Gértrudix-Barrio *et al.*, 2007), puesto que sus posibilidades de contenido, difusión, lectura, reutilización, interrelación con otros contenidos y muchos otros aspectos pueden ser repensados de forma innovadora en el espacio digital.

En ámbitos educativos, los objetos y modelos en tres dimensiones son utilizados de manera habitual en:

- ingeniería y arquitectura (maquetas);
- geografía (mapas con relieve);
- dibujo técnico y las vistas normalizadas (piezas técnicas);
- estudios artísticos (réplicas de obras escultóricas);
- para conocer órganos en clases de anatomía, elementos naturales en biología, etc.

En la figura 1 se muestra un ejemplo de maqueta y lámina de estudio utilizados en enseñanza superior. En diversos estudios (Andrade-Lotero *et al.*, 2012) se ha demostrado que un contenido se aprende más rápido mediante la manipulación de objetos tridimensionales.

Los libros, tanto electrónicos como de papel, al ser uno de los principales vehículos de transmisión de información, deberían disponer de la posibilidad de incluir objetos tridimensionales en su interior, ya que dichos objetos son diferentes a otros que pueden incluirse en un libro, como imágenes o videos. La interpretación detallada de la tridimensionalidad de un objeto a través de imágenes bidimensionales es un proceso que requiere, en algunos casos, un gran esfuerzo intelectual. Sin embargo, los modelos 3D, al permitir la observación desde distintos puntos de vista, facilita el aprendizaje de los conceptos espaciales asociados a los mismos.

En un escenario donde conviven los dos formatos: electrónico y papel, es interesante conocer de qué manera se puede conseguir que los objetos 3D se incluyan en los libros. En este artículo se describen y comparan nueve tecnologías para incorporar un objeto 3D con el que se pueda interactuar (mover y visualizar desde distintos ángulos) tanto en



Figura 1. Lámina y maqueta de estudio (modelo de vaca de la empresa *Medical Simulator*). Área de Zootecnia de la Universidad de La Laguna

libros de papel como en libros electrónicos. Aunque la inclusión de objetos tridimensionales se orienta más hacia el sector del libro de texto o divulgativo en los campos del arte, naturaleza, ciencia o tecnología, por citar los más extendidos, las tecnologías presentadas en este artículo son válidas para cualquier tipo de texto.

## 2. Incorporación de objetos 3D en libros

El intento de incorporar la tercera dimensión en los libros de texto o documentos informativos no es nuevo, tuvo su aparición hace siglos. Tanto con los libros *pop-up* como mediante el uso de imágenes en perspectiva (Damisch, 1997) se lograba el efecto de mostrar esa tercera dimensión. Sin embargo, las imágenes en perspectiva siguen siendo bidimensionales y no permiten visualizar el objeto desde un punto de vista distinto al representado. Los visualizadores 3D digitales se basan en las técnicas de representación clásicas, por lo que en realidad vemos una imagen 2D en la pantalla del ordenador o dispositivo de visualización. Sin embargo, dicha imagen es redibujada instantáneamente cuando el usuario cambia el punto de vista con un controlador (un ratón, el dedo...), por lo que se produce la sensación de manipular un objeto tridimensional.

La aparición del libro digital y tecnologías como la realidad aumentada permiten crear nuevas formas de visualización tridimensional. Con tecnologías digitales accesibles (muchas de ellas gratuitas), es posible actualizar las técnicas tradicionales para la incorporación de objetos 3D en formatos de papel como pueden ser los libros con *pop-ups* o los recortables. En este artículo se presentan nueve tecnologías para de incorpora-

ción de objetos 3D en libros (entendida como la posibilidad de que un objeto se pueda manipular o visualizar desde distintos ángulos), tanto en formato digital como en papel (figura 2).

Para la incorporación de un objeto 3D en un libro partiremos en todos los casos de su producción mediante tecnologías digitales, tanto si el destino es el soporte papel como si es digital. Para ilustrar cada una de las posibilidades descritas se ha seleccionado como ejemplo el modelo 3D digital de la escultura *Monumento al gato* del artista español Óscar Domínguez (figura 3), que formó parte de la *I Exposición internacional de esculturas en la calle* en Santa Cruz de Tenerife entre 1973 y 1974. El modelo 3D ha sido creado con la aplicación gratuita *SketchUp*.

### 2.1. Incorporación de objetos 3D en libros de papel

#### 2.1.1. Libros *pop-up* (a partir de software digital)

En libros con *pop-up*, al abrir sus páginas se despliega un

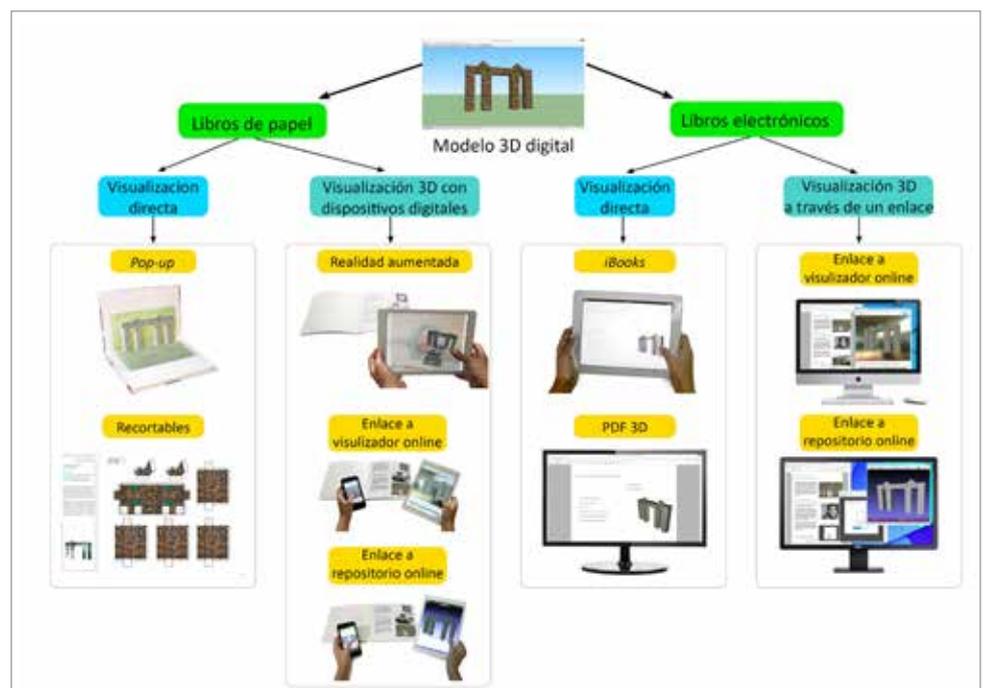


Figura 2. Tecnologías para la incorporación de objetos 3D en libros



Figura 3. Monumento al gato de Óscar Domínguez



Figura 4. Libro *pop-up*

modelo tridimensional de papel que permite una mejor comprensión de los contenidos volumétricos desarrollados en el texto. Existen referencias de estos libros desde 1240, cuando apareció el primero datado hasta la fecha (Connelly, 1999). Han sido utilizados en el campo de la medicina y en ámbitos educativos (Glassner, 2002), y se siguen empleando en la actualidad, sobre todo en el ámbito infantil y libro-juego, como método para incorporar objetos 3D en los libros de papel (figura 4).

El conjunto de técnicas para el diseño y creación de libros con objetos 3D desplegable (*pop-up*) de papel se denomina ingeniería de papel, término que hace referencia al conjunto de conocimientos para el diseño con papel, a través de sus posibilidades de plegado, recorte y ensamblado, dotándolos de tridimensionalidad.

Para el diseño de *pop-ups* es necesario entender el mecanismo de plegado y determinar las posiciones de los objetos emergentes para que no choquen entre sí o sobresalgan de la página (Iizuka et al., 2011). Con aplicaciones como *Pop-up Card Designer Pro* (disponible en versión gratuita) o *Pop-up Card Studio*, es posible crear *pop-ups* de manera digital. Estos programas permiten dibujar el contorno de un objeto y crear las líneas de recortado y doblado para construir la tercera dimensión. En la figura 5 se muestra el *pop-up* del objeto 3D utilizado como ejemplo en este artículo.

<https://goo.gl/LYOK9X>

La ventaja del libro *pop-up* radica en que, una vez terminado el proceso de creación, permite acceder a los modelos 3D tan sólo con abrir el libro, sin necesidad de ninguna tecnología ni procedimiento adicional; es por ello que estos libros son utilizados con frecuencia en educación infantil y primaria, sobre todo como material complementario para actividades, biblioteca o consumo particular. Por otro lado, un libro *pop-up* no es fácilmente modificable y requiere de un proceso de fabricación costoso en tiempo, material y mano de obra.

Se describen nueve tecnologías para incorporar un objeto 3D con el que se pueda interactuar tanto en libros de papel como en libros electrónicos

### 2.1.2. Libros con *papertoys* (recortables a partir de modelos digitales 3D)

En los libros con *papertoys*, el lector mediante recorte y plegado de papel puede obtener modelos tridimensionales a partir de patrones de piezas que se unen con pegamento. Los primeros recortables aparecen en Europa en el siglo XVIII, en el París de Luis XV, aunque es en 1840, con la aparición de la litografía en color, cuando experimentan

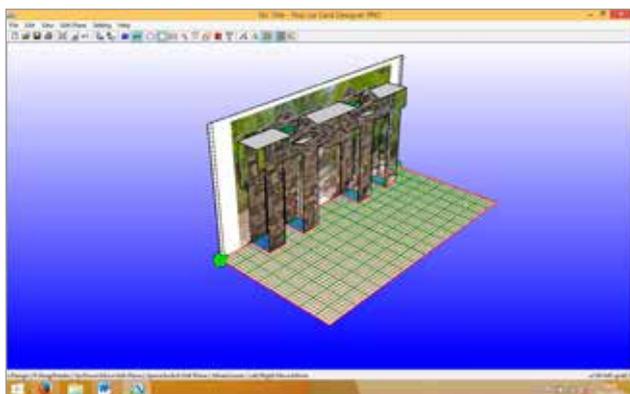


Figura 5. Ejemplo de un *pop-up* construido en *Pop-up Card Designer Pro*



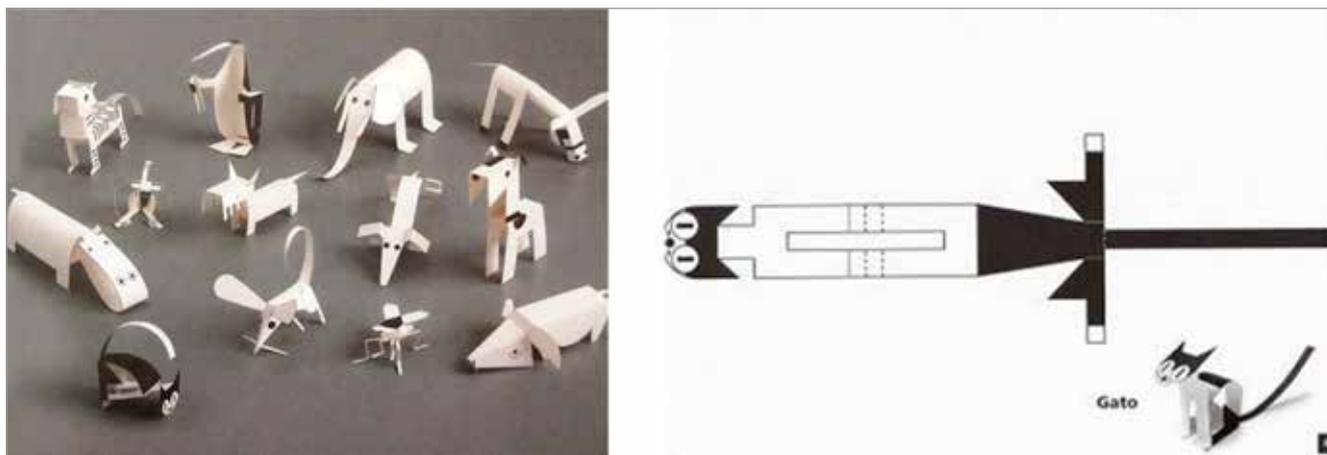


Figura 6. Ejemplo de libro de recortables

su mayor desarrollo y difusión (Haining, 1979; Reid-Walsh, 2012). Una muestra temprana de *papertoys* en España es la propuesta llevada a cabo por Antonio de Lara en los años 1933-1934 para la revista *Crónica*: se trata de una serie de recortables y plegables de animales coleccionables llamada *El arca de Noé* (IVAM, 2004) (figura 6).

El formato PDF 3D, una vez resueltos los problemas de visualización en dispositivos móviles, podría convertirse en el estándar para la inclusión de objetos 3D en libros digitales

El diseño de un papertoy con herramientas digitales constituye un proceso que engloba dos fases:

- la primera consiste en crear una plantilla recortable del modelo tridimensional, para lo cual se pueden utilizar aplicaciones como *123D Make*, *Pepakura Designer* o extensiones de *SketchUp* como *Unwrap and Flatten*, *Unfold* o *Flattery* (figura 7);
- en la segunda fase se lleva a cabo un proceso de posproducción para añadir detalles como solapas, pestañas y texturas con programas como *AutoCad*, *Illustrator*, *InDesign* o *Gimp*.

El diseño de un libro de recortables es una manera económica para la divulgación de objetos 3D en un aula sin acceso a internet ni dispositivos digitales. La creación de las planti-

llas recortables necesita varios programas, así como conocimientos informáticos y de retoque de imagen, pero el montaje final de los modelos 3D los realiza el lector con la única ayuda de unas tijeras y una barra de pegamento. Una vez realizados, son fácilmente reproducibles a través de fotocopias y posibles de incorporar en cualquier entorno escolar. El ejemplo del objeto 3D utilizado en este artículo (figura 8), además de otras 11 esculturas de Santa Cruz de Tenerife convertidas a formato *pop-up* se puede consultar en: <https://goo.gl/kgybMY>

### 2.1.3. Libros de papel con objetos 3D en realidad aumentada (*Magic book*)

La realidad aumentada (RA) permite al usuario ver el mundo real a través de un dispositivo digital con objetos virtuales superpuestos (Azuma, 1997). Los objetos virtuales pueden ser manipulados por el individuo, coordinando sus movimientos con las manos para obtener el punto de vista que desea. Se usa en distintos contextos: militar, medicina, diseño, ingeniería, robótica, enseñanza y aprendizaje, entretenimiento, tratamientos psicológicos, etc. (Azuma, 1997; Azuma et al., 2001).

La evolución de esta tecnología ha dado lugar a la aparición del libro aumentado o *magic book* (Billinghurst; Kato; Poupyrev, 2001; Juan; Beatrice; Cano, 2008; Ucelli et al., 2005). Este tipo de libro contiene unas marcas, las denominadas marcas de realidad aumentada, que permiten ver un objeto

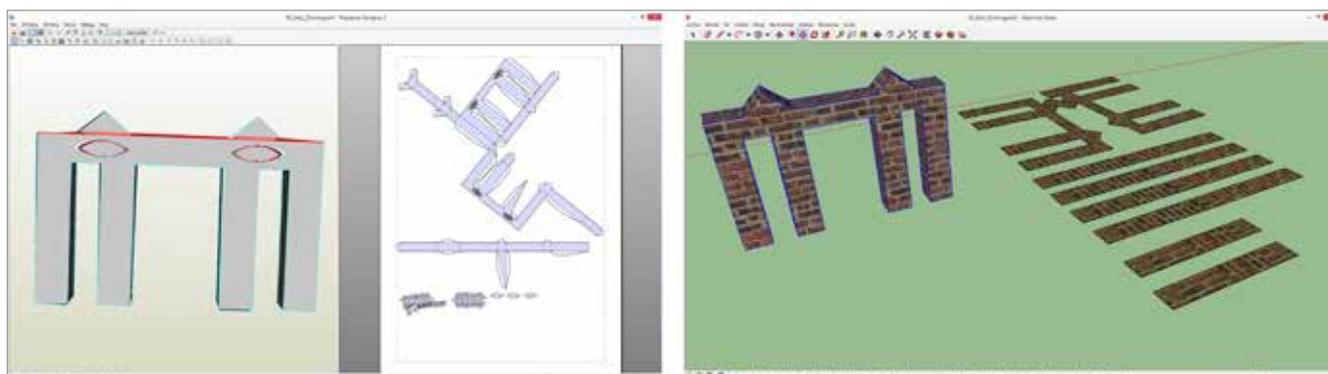
Figura 7. Ejemplos de creación de desplegado digital con *Pepakura Designer* (izq.) y *Unwrap and Flatten* de *Sketchup Make* (dcha.)



Figura 8. Libro de recortables con información (izq.) y el modelo montado en papel (dcha.)



Figura 9. Visualización del modelo 3D mediante realidad aumentada

3D a través de la cámara de un dispositivo móvil o la cámara de un ordenador (figura 9). Los libros aumentados ofrecen instrucciones del funcionamiento de la tecnología, donde se detalla la aplicación de visualización requerida para ver la realidad aumentada.

Un libro aumentado, es decir, un libro con marcas para visualizar realidad aumentada, permite convertir un modelo 3D en un archivo para visualizar a través de un dispositivo digital o una cámara de ordenador

Para crear un libro aumentado, es decir, un libro con dibujos o marcas para visualizar RA, existen programas, algunos con versiones gratuitas, como *Augment*, *Layar*, *Aurasma*, *Augmentaty*, *ARMedia*, etc. Estos programas permiten convertir un modelo 3D en un archivo para visualizar en RA a través de un dispositivo digital o una cámara de ordenador, enfocando a una marca (figura 9).

Para este artículo se han creado los ficheros de RA con *ARMedia* de *Inglobe Technologies*, un plugin (complemento que añade una funcionalidad adicional a un software) gratuito en su versión educativa para el programa de modelado 3D *SketchUp*. Permite transformar un modelo tridimensional en un fichero de realidad aumentada. En la tabla 1 se detallan los pasos para la visualización en RA del ejemplo de este artículo.

En este caso, a diferencia del anterior, aunque el libro es de papel es necesario usar un dispositivo digital con una aplicación específica y el modelo descargado para visualizarlo. Es importante señalar que para realizar la descarga del visualizador y del fichero de RA se precisa disponer de conexión a internet, aunque una vez realizada dicha descarga, el sistema funciona sin conexión a la Red.

La RA permite interactuar con el modelo 3D, movernos alrededor del mismo, hacer zoom, y girarlo con gestos de nuestras manos. Sin embargo, este proceso es sensible a las

condiciones externas como puede ser la luz, la posición de la marca, o el tipo de movimiento requerido, ya que si la cámara deja de ver la marca, se deja de visualizar el objeto 3D.

#### 2.1.4. Libro de papel con enlaces a visualizadores 3D online

Uno de los medios utilizados para conseguir que un libro de papel incluya información digital es mediante la referencia a una web, lo que obliga al lector a teclear la dirección web. Esta operación puede ser engorrosa ya que las direcciones suelen tener caracteres alfanuméricos complejos y de gran extensión. Existen estrategias para simplificar este proceso. Una de ellas es utilizar las direcciones url cortas y otra emplear aplicaciones que permiten, mediante un código, acceder a la dirección indicada, por ejemplo un código QR.

El código QR (*Quick Response*) fue creado en 1994 por la compañía japonesa *Denso Wave*. Los urls se convierten en una especie de código de barras que puede ser leído por las cámaras de los dispositivos móviles. Existen muchos generadores de código QR libres y gratuitos en internet (por ejemplo, *QR code creator*), que facilitan su generación y uso (Walsh, 2010).

<http://www.qrstuff.com>

Tabla 1. Instrucciones para la visualización en realidad aumentada (RA)

Instrucciones para la visualización en RA del modelo 3D de la escultura de Óscar Domínguez	
Instalar la aplicación <i>ARMedia Player</i> de <i>Inglobe Technologies</i> en el dispositivo móvil digital (versión <i>Android</i> o <i>iOS</i> ).	
Descargar el fichero de RA "18_Gato_Dominguez.armedia" en el dispositivo digital. <a href="https://goo.gl/bI23iG">https://goo.gl/bI23iG</a>	
Abrir con <i>ARMedia Player</i>	
Seleccionar el modelo "18_Gato_Dominguez.armedia" (para ello es posible que haya que actualizar la base de datos de la aplicación)	
Enfocar esta marca de RA con la cámara del dispositivo y visualizar el objeto en RA.	



Figura 10. Visualización de un objeto 3D en SketchFab. Acceso desde un código QR. Tras cargar el modelo 3D en internet, es posible interactuar con el mismo girándolo y haciendo zoom directamente con los dedos. Una vez se ha visualizado el enlace ya no es necesario que el dispositivo siga enfocando al código QR (tabla 2).



Figura 11. Modelo 3D en una aplicación de visualización de objetos 3D. Acceso desde un código QR. Al igual que con la RA, es posible interactuar con el modelo haciendo giros, cambios de escala y rotaciones directamente con movimientos de los dedos. Una vez está descargado el modelo 3D ya no es necesario que el dispositivo siga enfocando al código QR (tabla 3).

Para incluir un modelo 3D en un libro de papel se puede usar un enlace o código QR que conecte con un modelo tridimensional que se visualiza directamente en el navegador de internet (figura 10). Existen webs especializadas en la difusión y/o venta de modelos tridimensionales (*Thingiverse, Grabcad, 3D Warehouse, SketchFab*, etc.). También algunos museos e instituciones están poniendo disponibles para visualizar online sus modelos 3D, como por ejemplo la *Smithsonian Institution*, o el proyecto para divulgación de patrimonio en 3D del *Virtual World Heritage Laboratory*:  
<http://3d.si.edu>  
<http://vwhl.clas.virginia.edu>

El diseño de un libro de recortables es una manera económica de divulgar objetos 3D en un aula sin acceso a internet ni dispositivos digitales

Este sistema representa una alternativa más simple a la de visualización con realidad aumentada y además no requiere estar enfocando continuamente el código. Sin embargo sólo funciona mientras se disponga de conexión a internet. Además en muchos casos, debido a la configuración y políticas de los repositorios, no se permite descargar estos modelos en nuestros dispositivos. Esto no representa una gran limitación, siempre que nuestro objetivo sea la visualización del objeto 3D y no el disponer de él para por ejemplo editarlo o imprimirlo en 3D.

Tabla 2. Instrucciones para visualización online

Instrucciones para la visualización del ejemplo de este artículo desde un código QR enlazado a SketchFab	
Instalar una aplicación de lector de código QR (por ejemplo QR code reader) en el dispositivo móvil	
Abrir la aplicación y enfocar a la marca del código QR.	

### 2.1.5. Libro de papel con enlaces para descargar objetos tridimensionales

Otra opción de incorporar objetos 3D en un libro es enlazar con repositorios de objetos 3D. Desde la aparición de aplicaciones como *Dropbox, Drive, Mega* o *Box*, que permiten almacenar información en la nube (internet), cualquier usuario puede cargar modelos 3D en la nube y compartir estos ficheros a través de un enlace, como puede ser un código QR.

Descargar, abrir o trabajar con texto, imágenes, gráficos o vídeos está normalizado en todos los dispositivos, sin embargo la visualización e interacción con modelos 3D digitales es poco habitual. Tradicionalmente sólo se podían visualizar archivos tridimensionales en programas de modelado 3D. En la actualidad esta limitación se ha superado, pudiéndose visualizar e interactuar con modelos tridimensionales en cualquier dispositivo digital que disponga de una aplicación de visualización 3D, como *3D Viewer, HD Model Viewer, A360, i3D Viewer, CAD View 3D, MeshLab, Graphite*, etc. Una vez descargado y abierto el archivo, se queda en una galería dentro de la aplicación en nuestro dispositivo, permitiendo visualizarlo sin necesidad de conexión internet (figura 11).

Los libros de papel con enlaces a repositorios de objetos representan un proceso más complejo que el ejemplo ante-

Tabla 3. Instrucciones para visualizar objetos 3D en repositorios

Instrucciones para descargar y visualizar objetos 3D en repositorios	
Instalar una aplicación de lectura de código QR (por ejemplo QR code reader) en el dispositivo digital.	
Instalar el visualizador 3D, por ejemplo Meshlab (iOS) o Graphite (Android)	
Enfocar la marca del código QR con la aplicación descargada.	
Una vez en la carpeta del repositorio descargar el archivo con el objeto 3D seleccionado.	
Utilizar el visualizador 3D para ver el objeto descargado.	



Figura 12. Visualización de un objeto 3D en un *iBook*

rior de enlazar a un visualizador 3D online, ya que implica (al igual que con la realidad aumentada) la instalación de una aplicación y la descarga de un fichero. Sin embargo, tiene la ventaja de que no requiere estar enfocando continuamente el código QR y que una vez descargados los objetos se pueden visualizar sin necesidad de conexión a internet ya que se quedan almacenadas en nuestro dispositivo.

## 2.2. Incorporación de objetos tridimensionales en libros digitales

Los libros digitales se pueden crear en múltiples formatos, no todos ellos con el mismo grado de estandarización (Browne, 2012). Prácticamente, cualquier formato de libro digital permite incluir con facilidad enlaces, imágenes, vídeo y sonido, una vez superados los problemas aún existentes de normalización en visores y tabletas. Sin embargo la mayoría de estos formatos no contempla incluir objetos tridimensionales, con excepción del formato PDF 3D o el *iBook*. A continuación se detallan cuatro opciones para incluir objetos 3D en libros digitales.

### 2.2.1. Libro digital en formato *iBook*

La aplicación *iBooks Author* presentada por *Apple* en 2012 permite incluir contenidos multimedia (imágenes, vídeos, cuestionarios, objetos 3D, animaciones, gráficos interactivos). En concreto, los modelos 3D hay que importarlos en formato *Collada* y una vez insertados pasan a formar parte integral del libro. Esta herramienta permite crear libros mul-

timedia para dispositivos *Apple*, a través de un entorno gratuito de desarrollo muy intuitivo, pensado para profesores y divulgadores (Kwok, 2012).

La ventaja de un *iBook* es que una vez descargado el archivo, se puede visualizar el contenido 3D interactivo sin necesidad de conexión a internet, con lo cual reproduce a la perfección las características de movilidad de un libro de papel. Para este artículo, se ha realizado un *iBook* (figura 12) con modelos 3D que se pueden descargar y visualizar en cualquier dispositivo *Apple* (*iPhone*, *iPad* y *Mac*).

<https://goo.gl/sYdA95>

### 2.2.2. Libros digitales en formato *Adobe PDF 3D*

El formato PDF ha sido durante años el estándar para documentos digitales imprimibles y, desde 2008, es la norma *ISO 32000*. El formato especial 3D-PDF permite incorporar contenido multimedia y modelos 3D (Ruthensteiner; Heß, 2008).

Para crear un PDF 3D se puede usar el programa *Adobe PDF Pro*, que permite diseñar un libro e incorporar contenido 3D. Sin embargo no es un programa diseñado para usuarios sin experiencia y además sólo está disponible de manera gratuita en versión de prueba de 30 días. El PDF-3D se puede abrir con cualquier lector de ficheros PDF, sin embargo para la correcta visualización de los modelos 3D es necesario utilizar *Adobe Reader*. Este programa visualiza el fichero en cualquier dispositivo, aunque todavía presenta problemas en smartphones y tabletas digitales. Debido a ello se han creado otras aplicaciones, como *3D PDF Reader*, que intentan mejorar la visualización en dispositivos móviles. Para este artículo se ha creado un ejemplo (figura 13) de PDF-3D que se puede descargar en:

<https://goo.gl/Kxq5Co>

El formato PDF es un formato ya estandarizado para texto e imágenes. Si se resuelven los problemas de visualización en dispositivos móviles, el PDF 3D, podría convertirse en el formato estándar para incluir objetos tridimensionales en libros digitales.

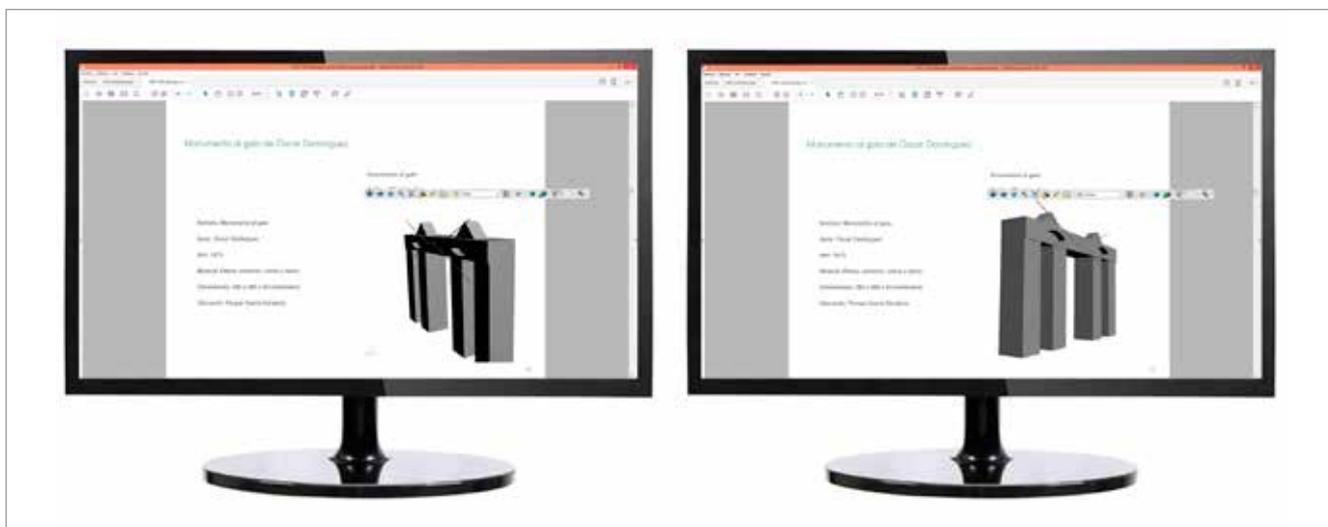


Figura 13. Visualización de un objeto 3D en PDF (*Adobe Acrobat*)

### 2.2.3. Libro digital con enlaces a visualizadores 3D online y repositorios de objetos 3D

Incluir enlaces en un libro digital es una opción en todos los formatos existentes (PDF, epub, *Word*...). Los enlaces pueden ser a visualizadores 3D online o a repositorios de objetos 3D. Estas dos opciones son iguales que las detalladas en el caso de los libros de papel en los apartados 2.1.4. y 2.1.5. En este caso se ha realizado un ejemplo en PDF con enlace a la escultura del *Monumento al gato* de Óscar Domínguez que se puede descargar en:

<https://goo.gl/BymCxz>

Dicho ejemplo incluye dos enlaces, uno a un visualizador online y otro a un repositorio que permite la descarga del objeto 3D.

De esta manera cualquier persona puede crearse un documento digital que incluya objetos 3D en su contenido. Las limitaciones de cada uno de estos dos sistemas (descarga de aplicaciones y ficheros, conexión a internet, etc.) serán las mismas que las vistas en incluir objetos 3D a través de enlaces en los libros de papel.

### 3. Conclusiones

En este artículo se han descrito 9 propuestas para la inclusión de modelos tridimensionales en libros, tanto de papel como digitales. No todas las propuestas presentan las mismas características, por lo que se podrá optar por una u otra alternativa dependiendo de:

- tipo de libro que se quiera crear;
- recursos disponibles;
- conocimientos tecnológicos de los autores;
- destinatarios.

Por ejemplo, si los destinatarios son estudiantes de educación infantil o primaria, lo normal, ya que no requiere ninguna formación, será utilizar libros de papel tipo *pop-up* o visualización directa en libros electrónicos como *iBooks* o PDF 3D. Los estudiantes en educación infantil o primaria no siempre tienen dispositivos móviles propios por lo que la utilización de libros electrónicos con contenido 3D directo, tendría que contar con materiales del centro escolar o de la propia familia del estudiante.

Para educación secundaria y superior, los libros (tanto de papel como digitales) con enlaces a visualizadores directos en internet (tipo *SketchFab*), pueden ser muy interesantes al permitir que los alumnos utilicen sus propios dispositivos móviles para ver los modelos 3D. Por otro lado, los libros con *papertoys* pueden ser una buena herramienta de divulgación mediante actividades de construcción de un objeto tridimensional en niveles educativos preuniversitarios.

Para autores que quieran incluir contenido 3D en sus libros, sería necesario un mínimo de formación tecnológica para poder crear o al menos manejar modelos 3D digitales. La manera más simple de incluir estos contenidos tridimensionales sería a través de enlaces (códigos QR o hipervínculos) a modelos ya creados en internet (ej. *SketchFab*). Autores con conocimientos avanzados pueden crearse sus propios *papertoys*, libros *pop-up* o repositorios de modelos 3D o realidad aumentada de una manera gratuita.

En ámbitos educativos, donde las aulas virtuales tienen una gran implantación, los libros digitales con contenidos tridimensionales son adecuados, por su formato, en este tipo de entornos de enseñanza-aprendizaje. Por otro lado, los contenedores online de objetos 3D, entendidos como pequeñas unidades de contenido, cada vez serán más habituales, pudiendo estar enlazados o no a libros digitales.

### 4. Bibliografía

**Andrade-Lotero, Luis-Alejandro; Espitia-Gómez, Carolina; Huertas-Franco, Elquin-Antonio; Aldana-Ahumada, Derly-Rocío; Bacca-Pachón, Paola-Andrea** (2012). "Tocar o mirar: comparación de procesos cognitivos en el aprendizaje con o sin manipulación física". *Psicología educativa*, v. 18, n. 1, pp. 29-40.

<http://dx.doi.org/10.5093/ed2012a3>

**Avella, Natalie** (2003). *Paper engineering: 3D design techniques for a 2D material*. Rotovision. ISBN: 978 2880467111

**Azuma, Ronald** (1997). "A survey of augmented reality". *Presence*, v. 6, n. 4, pp. 355-385.

<http://www.cs.unc.edu/~azuma/ARpresence.pdf>

<http://dx.doi.org/10.1162/pres.1997.6.4.355>

**Azuma, Ronald; Baillet, Yohan; Behringer, Reinhold; Feiner, Steven** (2001). "Recent advances in augmented reality". *IEEE Computer graphics and applications*, v. 21, pp. 34-37.

<http://dx.doi.org/10.1109/38.963459>

**Berube, Linda** (2005). "E-books in public libraries: a terminal or termination technology?". *Interlending and document supply*, v. 33, n. 1, pp. 14-18.

<http://dx.doi.org/10.1108/02641610510582090>

**Billinghurst, Mark; Kato, Hirokazu; Poupyrev, Ivan** (2001). "The magic book—moving seamlessly between reality and virtuality". *IEEE computer graphics and applications*, v. 21, n. 3, pp. 6-8

<http://dx.doi.org/10.1109/38.920621>

**Browne, Glenda** (2012). "Indexes in tomorrow's world". En: *Australian Society for Technical Communicators (NSW) conference*, Sydney.

<http://www.webindexing.com.au/indexes-in-tomorrows-world-astc-nsw-conference-presentation-november-2012>

**Connolly, Daniel K.** (1999). "Imagined pilgrimage in the itinerary maps of Matthew Paris". *The art bulletin*, v. 81, n. 4, pp. 598-622.

[http://www.jstor.org/stable/3051336?seq=1#page\\_scan\\_tab\\_contents](http://www.jstor.org/stable/3051336?seq=1#page_scan_tab_contents)

**Damisch, Hubert** (1997). *El origen de la perspectiva*. España: Alianza Editorial. ISBN: 84 206 7143 6

**Doctorow, Cory** (2004). *Ebooks: Neither E, nor books*. FQ Books.

<http://craphound.com/ebooksneitherenorbooks.txt>

<http://jamillan.com/doctorow.htm>

**Gértrudix-Barrio, Manuel; Álvarez-García, Sergio; Galisto-del-Valle, Antonio; Gálvez-de-la-Cuesta, María-del-Carmen; Gértrudix-Barrio, Felipe** (2007). "Acciones de diseño y desarrollo de objetos educativos digitales: programas ins-

titucionales". *Revista de universidad y sociedad del conocimiento (RUSC)*, v. 4, n. 1.

[http://www.uoc.edu/rusc/4/1/dt/esp/gertrudix\\_alvarez\\_galisteo\\_galvez.pdf](http://www.uoc.edu/rusc/4/1/dt/esp/gertrudix_alvarez_galisteo_galvez.pdf)

**Glassner, Andrew** (2002). "Interactive pop-up card design". *Computer graphics and applications*, v. 22, n. 1, pp. 79-86. <http://dx.doi.org/10.1109/38.988749>

**Haining, Peter** (1979). *Movable books: An illustrated history*. London: New English Library. ISBN: 978 0450039492

**Iizuka, Satoshi; Endo, Yuki; Mitani, Jun; Kanamori, Yoshihiro; Fukui, Yukio** (2011). "An interactive design system for pop-up cards with a physical simulation". *The visual computer*, v. 27, n. 6-8, pp. 605-612.

<http://www.npal.cs.tsukuba.ac.jp/~iizuka/projects/popup/data/fulltext.pdf>

<http://dx.doi.org/10.1007/s00371-011-0564-0>

**IVAM** (2004). *Los talleres didácticos del IVAM 1998-2005*. Valencia: Instituto Valenciano de Arte Moderno. ISBN: 978 8448238605

**Juan, Carmen; Beatrice, Francesca; Cano, Juan** (2008). "An augmented reality system for learning the interior of the human body". En: *Procs of the 8<sup>th</sup> IEEE intl conf on advanced learning technologies*, pp. 186-188.

<http://users.dsic.upv.es/~mcarmen/docs/HumanBody.pdf>

<http://dx.doi.org/10.1109/ICALT.2008.121>

**Kang, Yen-Yu; Wang, Mao-Jiun; Ling, Rungtai** (2009). "Usability evaluation of e-books". *Displays*, v. 30, pp. 49-52.

<http://dx.doi.org/10.1016/j.displa.2008.12.002>

**Kwok, Roberta** (2012). "Going digital". *Nature*, v. 485, pp. 405-407.

<http://dx.doi.org/10.1038/nj73>

**Rao, Siriginidi-Subba** (2005). "Electronic books: their inte-

gration into library and information center". *The electronic library*, v. 23, n. 1, pp. 116-140.

<http://dx.doi.org/10.1108/02640470510582790>

**Reid-Walsh, Jacqueline** (2012). "Books or toys?: A traveller's tale: Researching early movable books for and by children in material and virtual collections". *Explorations into children's literature*, v. 22, n. 1, pp. 156-169.

**Rodríguez-Bravo, Blanca; Pacios, Ana-Reyes; Vianello-Osti, Marina; Moro-Cabero, Manuela; De-la-Mano-González, Marta** (2015). "Digital transition of teaching learning resources at Spanish universities". *El profesional de la Información*, v. 24, n. 6, pp. 737-748.

<http://dx.doi.org/10.3145/epi.2015.nov.05>

**Ruthensteiner, Bernhard; Heß, Martin** (2008). "Embedding 3D models of biological specimens in PDF publications". *Microscopy research and technique*, v. 71, n. 11, pp. 778-786.

<http://dx.doi.org/10.1002/jemt.20618>

**Smithsonian Institution** (2015). *Smithsonian x3D*.

<http://3d.si.edu/browser>

**Ucelli, Giuliana; Conti, Giuseppe; De-Amicis, Raffaele; Servidio, Rocco** (2005). "Learning using augmented reality technology: Multiple means of interaction for teaching children the theory of colours". En: *Procs of the 1<sup>st</sup> intl conf on intelligent technologies for interactive entertainment*, pp. 193-202.

[http://dx.doi.org/10.1007/11590323\\_20](http://dx.doi.org/10.1007/11590323_20)

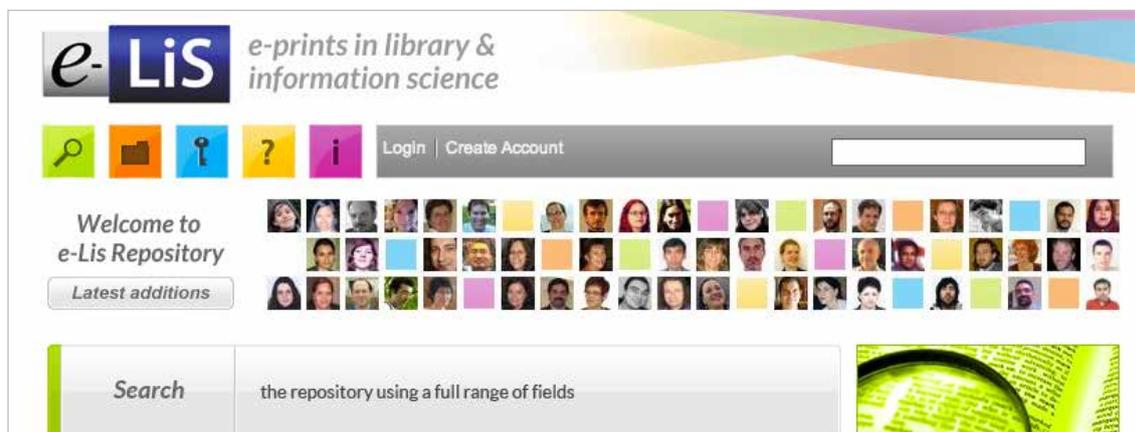
**Virtual World Heritage Laboratory** (2015). *The digital sculpture project*.

<http://www.digitalsculpture.org/laocoon/index.html>

**Walsh, Andrew** (2010). "QR codes – Using mobile phones to deliver library instruction and help at the point of need". *Journal of information literacy*, v. 4 n. 1, pp. 55-65.

<http://dx.doi.org/10.11645/4.1.1458>

## Da visibilidad a tu trabajo depositándolo en e-LIS, el mayor repositorio internacional sobre biblioteconomía, documentación y comunicación



<http://eprints.rclis.org>

# SOFTWARE

## DISEÑO DE UN AGREGADOR PARA LA GESTIÓN DE LOS *BIG DATA* INFORMATIVOS

Design of an aggregator for managing informative big data

**Manuel Blázquez-Ochando**



**Manuel Blázquez-Ochando** es profesor del *Departamento de Biblioteconomía y Documentación* de la *Universidad Complutense de Madrid (UCM)*. Su investigación se centra en el desarrollo de software documental, agregadores de contenidos, algoritmos de recuperación, programas *web crawler* y técnicas de *scraping*. Es autor del buscador *WauSearch* y de la distribución de utilidades para bibliotecas, archivos y museos *AMPdoc*. Su actividad docente aborda la automatización de unidades de información, arquitectura de la información y metadatos, diseño de bases de datos y aplicaciones web documentales.

<http://www.mblazquez.es>

<http://orcid.org/0000-0002-4108-7531>

*Universidad Complutense de Madrid, Facultad de Ciencias de la Documentación  
Departamento de Biblioteconomía y Documentación  
Santísima Trinidad, 37. 28010 Madrid, España  
manublaz@ucm.es*

### Resumen

El artículo explica el diseño y características de un sistema de agregación de contenidos de código abierto. Entre las especificaciones del programa destaca un motor de procesamiento colaborativo, capacidades de monitorización de la recuperación de información en tiempo real, configuración del comportamiento del agregador, clasificación automática, alertas de noticias filtradas y nuevos tipos de representación de la información a partir de mapas relacionales interactivos. Por otra parte el programa de agregación está diseñado para gestionar miles de canales de sindicación en formato RSS. Además proporciona estadísticas que pueden servir para estudiar la producción informativa de cualquier sujeto productor y el impacto de la información publicada en otras fuentes. El resultado obtenido se traduce en módulos capaces de comparar las relaciones entre distintas informaciones o fuentes, determinar su grado de influencia, repercusión, impacto y cuantificación.

### Palabras clave

Agregadores; *Big data*; Producción informativa; Canales de sindicación RSS; Recuperación de información.

### Abstract

The design and characteristics of a new open source content aggregation program, XYZ, are described. Several features of the program stand out, including the processing engine of syndication channels, monitoring capability of information recovery in real time, possibility of configuration of the aggregator behavior, automatic content classification, and new models for representing information from relational interactive maps. On the other hand, the aggregation program is designed to manage thousands of syndication channels in the RSS format. It also provides statistics that can be used to study the production of any information producer and the impact of the information published in other sources. The XYZ modules are capable of comparing the relationship between news or information from different sources and the degree of influence which is detected by patterns.

### Keywords

Aggregators; Big data; Information production; RSS feeds; Information retrieval.

**Blázquez-Ochando, Manuel** (2016). "Diseño de un agregador para la gestión de los *big data* informativos". *El profesional de la información*, v. 25, n. 4, pp. 671-683.

<http://dx.doi.org/10.3145/epi.2016.jul.17>

## 1. Introducción

La creciente producción de información y su disponibilidad en la Red son el origen del concepto *big data*. **Mayer-Schönberger** (2013) lo define centrándose en dos aspectos: el almacenamiento de datos a gran escala y los métodos de análisis mediante patrones, con los que se advierten repeticiones y tendencias que sirven para crear modelos predictivos en campos tan diversos como la medicina, meteorología, criminología, finanzas, consumo o genómica, entre otros.

También se incide en el estudio de las correlaciones entre variables, para descubrir relaciones ocultas entre la información y datos masivos previamente almacenados, que pasarían inadvertidas. Con ello se consigue un mayor conocimiento de las causas y efectos que afectan a un objeto de estudio dado, todo lo cual revierte en ventajas competitivas. Tomando como punto de partida esta definición, se plantea la creación de un programa de agregación de contenidos, capaz de trabajar con textos informativos procedentes de miles de medios de comunicación, para su posterior filtrado y análisis correlacional.

Para abordar el objeto de estudio, se analiza la bibliografía científica sobre *big data* en relación con las aplicaciones de agregación en medios de información y comunicación y su tecnología. Una vez revisadas las últimas investigaciones, se propone un pliego de especificaciones con las características deseables para el diseño del nuevo programa de agregación, sus módulos, sets de datos y desarrollo. Finalmente se aborda el funcionamiento y sus características principales.

La búsqueda de artículos para la revisión del estado del arte, se ha ceñido a las bases de datos y buscadores: *Google Scholar*, *Web of Science* y *Scopus*.

La fórmula de consulta utilizada para la recuperación fue la siguiente:

```
(intitle:"aggregator" AND intext:"RSS") OR (intitle:"RSS reader" AND intext:"aggregator") OR (intitle:"feed reader" AND intext:"aggregator") OR (intitle:"feed aggregator" AND intext:"RSS") OR (intitle:"RSS crawler" AND intext:"feed") OR (intitle:"syndication" AND intext:"feed").
```

## 2. Big data y agregación en periodismo y comunicación

Una cuestión recurrente cuando se aborda el tema de *big data* corresponde a los métodos de gestión, procesamiento y extracción de los datos. Según **Bansal y Kagemann** (2015) aún deben mejorarse los procedimientos de almacenamiento para cada tipo de datos. Esto se debe a su heterogeneidad temática y tipológica. Para dificultar aún más la tarea, el método de extracción y su procesamiento varía en función de su contexto de aplicación, género, formato e incluso fuente de procedencia. Por ejemplo la metodología para el tratamiento de textos planos, estructurados, imágenes y otros documentos multimedia, requiere patrones de clasificación específicos que permitan reconocer expresiones, vocabulario técnico y marcas audiovisuales basadas en polígonos, o bien espectrogramas de audio, entre otros. Otros autores

también coinciden en la necesidad de dominar la extracción y el almacenamiento de datos, como paso previo a estudios analíticos (**Chen; Zhang**, 2014), que aún se encuentran en una etapa inicial. De hecho todavía resulta difícil combinar los *big data* procedentes de fuentes y temáticas dispares, si no existe una guía de relaciones hipotéticas, construida a priori por el investigador. Sin embargo el periodismo puede aprovecharse de infraestructuras ya conocidas, tales como los programas de agregación de contenidos (**Bazargani; Brinkley; Tabrizi**, 2013) para estudiar las relaciones entre las noticias y contenidos publicados en la Web por los medios de comunicación.

Este enfoque permite hablar del concepto *big data* informativo, por aludir al tratamiento masivo de los canales de comunicación de prensa, radio y televisión digitales. Sin embargo existen dificultades inherentes al producto informativo a la hora de diferenciar la originalidad, calidad, valor e interés de la información (**Katakis et al.**, 2009). Ello requiere aún del análisis documental y periodístico del profesional, que deberá dominar las técnicas de minería de datos para realizar periodismo de investigación, para ayudar a desvelar las correlaciones entre eventos y hechos a priori inconexos (**Colle**, 2013). Esto facilita las actividades de curación de contenidos y repercute en la sociedad a través de medios y plataformas digitales (**Guallar; Leiva-Aguilera**, 2013). Todo ello justifica la necesidad de construir sistemas de agregación de contenidos orientados a los *big data*, a la clasificación de contenidos, sistemas de alerta y medición de impacto.

“ Los *big data* generados por los medios puede ser tratado por agregadores. La dificultad estriba en diferenciar el valor de los contenidos y su impacto ”

También se plantean otros retos derivados como la representación de noticias en tiempo real, para crear una fotografía instantánea del interés social y mediático. De hecho las últimas investigaciones en el área centran su atención en la revisión de sucesos y eventos, así como sus repercusiones en los flujos informativos internacionales (**Severo; Beauguitte; Pecout**, 2015). Previamente otros investigadores también atendieron al análisis de las noticias de propagación global con el propósito de identificar las fuentes de información más trascendentales e influyentes (**Gallé; Renders; Kars- tens**, 2013). Incluso se han llevado a cabo métodos para conocer la relevancia de los contenidos con mayor interés en la opinión pública (**Thelwall; Prabowo; Fairclough**, 2006). Esto fue posible usando canales de sindicación y tecnologías de agregación, que han servido como pilares básicos en la elaboración de dichos trabajos. Quizá el ejemplo que mejor encarna los *big data* informativos es el trabajo de **Travers et al.** (2014) en el que se revisaron 10 millones de noticias obtenidas durante 8 meses de agregación. Su análisis desveló que la *Ley de Pareto* se cumple al identificar que una minoría de canales de sindicación producía la mayor parte de los contenidos recopilados. Su trabajo desveló cuáles eran los medios digitales más productivos e influyentes y con qué publicaciones se alcanzó más éxito. También ayudó

a proporcionar datos métricos y de caracterización de las fuentes de información, la extensión media de los artículos, los fenómenos de duplicación y replicación, estrategias de posicionamiento y la capacidad de propagación informativa.

A pesar del valor que los canales de sindicación pueden tener en investigaciones de *big data*, uno de los obstáculos que los investigadores pueden encontrar es su localización y extracción. Según **Reichert et al.** (2011) esto es un problema de minería de datos que puede resolverse combinando estrategias de consulta en buscadores y programas de tipo *webcrawler*. Sin embargo la creación de una colección de canales de sindicación para su estudio y análisis, lejos de ser una recopilación masiva, también debe basarse en factores de calidad y pertinencia. Esto es, que además de obtener datos masivos las fuentes de información sean relevantes y suficientemente reconocidas en un dominio temático concreto, que facilite la interpretación de los contenidos que sean recuperados.

Los agregadores deben incorporar métodos que ayuden a clasificar mejor la información y aumentar su escalabilidad

La realidad de las tecnologías de la información en torno a los *big data* y la agregación de contenidos han servido para crear nuevas empresas periodísticas, basadas en la reutilización de información que han cambiado el paradigma del sector (**Carlson; Usher**, 2015). Es posible competir con grandes grupos de comunicación, cuya adaptación al cambio es más lenta, situándolos en una posición de desventaja, con un adecuado empleo de los agregadores y la curación de contenidos. Un ejemplo de las consecuencias fue el caso *Google News España* (**Rodríguez-Cano**, 2015) cuyo servicio fue cerrado tras la aprobación de la reforma de *Ley 21/2014 de 4 de noviembre de propiedad intelectual (España, 2014)*. La imposición del canon AEDE (*Asociación de Editores de Diarios Españoles*) a todos los portales de agregación de noticias con o sin ánimo de lucro, obligó a numerosas empresas a cesar sus actividades y a limitar las posibilidades de innovación en el sector. Todo ello fue debido en gran medida a la crisis de las empresas periodísticas que se vieron incapaces de competir ante los servicios de comparación de noticias y *newsclipping*, que de forma gratuita se proporcionaban (**Guallar**, 2015). A pesar de todo la *Ley* no impide que los agregadores puedan ser usados o implementados en servidores locales (**López-Maza**, 2015). En torno al ámbito de las libertades cabe destacar que la transparencia se logra asegurando el libre acceso a la información y uno de los mejores medios que tiene la sociedad siguen siendo los agregadores de contenidos y la tecnología de sindicación (**Marty et al.**, 2010). Para alcanzar ese fin se tiene que asegurar un flujo constante de información, procedente de los medios, gobiernos y administraciones públicas. La sociedad tiene derecho a cotejar y contrastar la información de forma automática con miles de fuentes de información, para intentar conformar una opinión crítica, exhaustiva, fiel a la realidad y debidamente documentada (**Leaver; Willson; Balnaves**, 2012).

### 3. Tecnologías de agregación

El desarrollo de agregadores surge de la necesidad de controlar y procesar la información que generan millones de autores y editores en la Web. Si se asume que cada sitio web basado en CMS genera al menos un canal de sindicación, la cifra estimada superaría los 300 millones (*BuiltWith*, 2016). A esta cantidad habría que añadir los canales generados por blogs y redes sociales, obteniendo un volumen difícilmente abarcable. Una forma de afrontar el problema son los métodos de análisis multidimensional que tratan de procesar los *big data* en sets de datos no necesariamente caracterizados, para su posterior análisis (**Cuzzocrea**, 2015).

Dos de los programas más reconocidos en la materia son *Apache Hadoop* y *Apache Hive*. Su objeto es la agregación y análisis de cualquier tipo de contenido y su tratamiento a gran escala, de forma distribuida a través de una red de servidores. No podría considerarse un sistema específico para el tratamiento de canales de sindicación, aunque sí para el desarrollo de buscadores y programas *webcrawler*.

Otro enfoque para resolver el problema, corresponde a los métodos de agrupación de contenidos mediante algoritmos *K-mean* (**Li et al.**, 2007). Sin embargo su aplicación es compleja ya que los resultados que se obtienen pueden ser igualmente inabarcables. Incluso con un número relativamente bajo de canales de sindicación se pueden producir miles de grupos de noticias por efecto de una excesiva granularidad. Una solución a éste problema es el método de agrupación multimodal de las noticias (**Messina; Montagnuolo**, 2009). Esto implica usar el corpus textual para la agrupación de los contenidos, y con ello obtener un modelo estadístico-probabilístico que permita clasificar sucesivas noticias. Tal como se indica, resulta de gran utilidad para su aplicación en colecciones de textos con información heterogénea o rasgos semánticos bien caracterizados, que facilitan los procesos de clasificación.

Otro método de clasificación es el referido al empleo de inteligencia artificial. Un agregador podría ser capaz de distinguir los artículos relevantes para el lector, basándose en su comportamiento (**Chen; Bøen**, 2008). Esto ayuda a eliminar los contenidos accesorios con un margen de error cercano al 20%. Otros investigadores también observaron el inconveniente de encontrar canales de sindicación no conocidos. Ello requería usar programas *web crawler* adaptados al caso, abriendo la investigación al concepto de agregadores híbridos, con funciones de lectura, clasificación y descubrimiento (**Lee et al.**, 2008).

Otro enfoque es concebir al agregador como un intermediario entre las fuentes de información y el suscriptor que recibe la información en un formato enriquecido, diferente a RSS (**O’Riordan; O’Mahoney**, 2011). Si bien el concepto representa ventajas a la hora de ampliar los metadatos originales, representa una redundancia en la multiplicación de los procesos de agregación. Esto se debe a que el suscriptor necesitaría otro programa de lectura diferente al intermediario. También se han creado proyectos de reutilización de código abierto, para poner en marcha agregadores en la plataforma *WordPress* (**Isah**, 2012). Ello hace posible la creación de portales autoalimentados con noticias de terceros medios, con un bajo coste de mantenimiento.

Otro asunto abordado en la literatura científica es el rendimiento de los agregadores. Si se desea utilizar esta técnica en la investigación de los *big data*, se necesita mejorar sus capacidades y escalabilidad. Algunos autores afirman que la actualización periódica de los contenidos puede ser una solución para controlar la producción de información publicada (Sia; Cho; Cho, 2007). Sin embargo cuando las investigaciones requieren el análisis de varios miles de canales de sindicación, éste método puede provocar pérdidas de datos, al no poder abarcar a tiempo todo el volumen de noticias que se genera (Horincar; Amann; Artières, 2010).

En esta línea, se viene desarrollando el proyecto *RoSeS* (Creus et al., 2011) que apuesta por un esquema de recuperación a gran escala, que radica en determinar tiempos de refresco para la revisión de los canales de sindicación, basándose en los factores del límite del ancho de banda, la cantidad de noticias publicadas y la frecuencia de actualización del canal. Ello determina un tiempo de revisión específico para cada canal de sindicación, ordenando su prioridad de procesamiento dentro de la lista que los agrupa. Esto se enmarca en un intento por predecir y caracterizar el comportamiento de la publicación de los contenidos sindicados (Hmedeh et al., 2011) para crear mejores sistemas de agregación.

Otro caso es el proyecto *NectarRSS* que determina prioridades de procesamiento en los canales de sindicación, según las preferencias del usuario conforme a los contenidos consultados, su frecuencia y las palabras clave de sus búsquedas (Samper et al., 2008). Ello favorece la concentración del esfuerzo de análisis en las fuentes que satisfacen sus necesidades de información por encima de las demás.

Todas las investigaciones parecen coincidir en la necesidad de mejorar los agregadores en torno a una serie de aspectos muy concretos:

- capacidad para filtrar la información, evitando el ruido y noticias irrelevantes para el lector: se recomienda el empleo de métodos de clasificación automática, autoaprendizaje y análisis estadístico del comportamiento del usuario en referencia a sus necesidades informativas;
- aumentar la escalabilidad para procesar más canales de sindicación: se proponen frecuencias de actualización periódica de las fuentes de información, reducir el régimen de actualización según estadísticas de producción, diseñar reglas predictivas según el comportamiento del usuario o bien establecer prioridades de procesamiento.

#### 4. Metodología y desarrollo

Una vez revisada la bibliografía y el estado del arte en materia de *big data* informativos y agregación, se han planteado las siguientes fases metodológicas:

- 1) Definición de las especificaciones para diseñar el nuevo programa de *big data* informativos, teniendo en cuenta los aspectos sin resolver, advertidos en las investigaciones analizadas.
- 2) Diseño de un esquema modular que organice las funciones del programa conforme a cada especificación declarada.
- 3) Diseño de una base de datos que permita el almacena-

miento eficiente de la información procedente de los medios de comunicación.

- 4) Elección del entorno de programación y construcción del agregador.

Una solución que mejora notablemente el rendimiento del agregador es la introducción de programas *parser* colaborativos. Esto ayuda a repartir la carga de trabajo y priorizar las fuentes más relevantes

#### 4.1. Especificaciones

Los puntos clave del programa se han basado en las conclusiones y experiencias publicadas por los investigadores en materia de *big data* informativos y diseño de programas de agregación. Se ha procurado el cumplimiento de las siguientes:

- diseñar un método que aumente la capacidad para procesar *big data* procedentes de canales RSS 2.0, clasificar o filtrar los contenidos relevantes y eliminar aquellos que superen un tiempo de archivo establecido;
- desarrollar un sistema de clasificación automática que permita equiparar los filtros definidos por el administrador y los contenidos recopilados en los canales de sindicación;
- introducir un buscador a texto completo en lenguaje natural con operadores de consulta exacta;
- elaborar métodos estadísticos que permitan conocer el número de canales de sindicación, fuentes de información, noticias activas y archivadas, cronología de publicaciones y contenidos clasificados;
- crear utilidades para el análisis de impacto y correlación de noticias con el fin de evaluar eventos y contenidos de interés social o mediático;
- hacer posible el uso del agregador fuera del dominio público de internet, para evitar la aplicación de la *Ley de propiedad intelectual*. Esto es lograr su funcionamiento autónomo con servidor y base de datos local en el equipo del usuario, sin que suponga dificultades de instalación y configuración;
- crear métodos alternativos de representación de la información que permitan observar la publicación de noticias en tiempo real o comprobar sus relaciones;
- concebir un método de monitorización activa de los procesos de análisis y recopilación de datos, para observar el desempeño del programa.

#### 4.2. Diseño de módulos

En base a las especificaciones planteadas, el agregador *XYZ* dispone de:

- área administrativa: formada por los módulos de configuración, mantenimiento, estadísticas, importación y exportación de datos, edición de canales de sindicación, filtros para la clasificación de contenidos y procesamiento o ejecución del programa;

Tabla 1. Módulos y funciones del agregador XYZ

Módulo	Funciones	Módulo	Funciones
Configuración	Gestión de procesos	Filtros	Edición de filtros con operadores
	Opciones de archivo	Aplicación de análisis de impacto	Cálculo automático del impacto de una noticia
	Opciones de filtrado		Cálculo de correlación entre dos grupos de noticias
	Opciones de representación	Portada de noticias	Representación de noticias según relevancia de la fuente
	Opciones de alerta		Filtro de noticias por categorías y períodos temporales
Mantenimiento / Estado	Comprobación de errores	Tiempo real	Información recopilada en el momento en que se procesa
	Estado de la base de datos		Filtro de noticias por categorías y períodos temporales
	Reiniciar agregador	Noticias filtradas	Selección de últimas noticias filtradas
Estadísticas	Estadísticas generales		Impresión de dossiers según filtros
	Estadísticas personalizadas	Buscador	Consulta en lenguaje natural y frase exacta
Importación	Importar listas de canales de sindicación		Filtrar resultados según fuentes e intervalos temporales
	Importación OPML		
	Exportación OPML		
Edición	Edición masiva de canales de sindicación	Mapa de noticias	Mapa de noticias relacionadas según categorías
	Edición en detalle		
Procesamiento	Recopilación de noticias	Bloc de noticias	Guardar noticias seleccionadas por el lector
	Monitorización de procesos		Guardar noticias filtradas automáticamente
	5 hilos de ejecución <i>parser</i>		Buscar entre las noticias guardadas
			Filtrar noticias según fuentes e intervalos temporales

- área de representación de contenidos: presenta los módulos de estudio de impacto, mapa de noticias, noticias en tiempo real, portada de noticias según categorías temáticas, contenidos filtrados y bloc de noticias guardadas (tabla 1).

Este diseño permite controlar todas las fases de tratamiento de la información. El primer paso consiste en la configuración del programa para controlar la velocidad de recopilación de datos y su comportamiento para guardar o eliminar todas las noticias o solamente aquellas que fueron filtradas por el programa.

Una vez parametrizado, el módulo de importación permite cargar listas de enlaces correspondientes a canales de sindicación RSS, indicando su valor, clasificación temática e importancia. En el proceso se discriminan aquellos que no sean validados debido a errores de codificación, estructura o integridad, obteniendo al mismo tiempo los datos meta-descriptivos de los que sí sean aceptados.

Una vez importados, el módulo de edición favorece la modificación puntual de los canales de sindicación para su preparación final.

Una vez dispuestos los canales, el módulo de filtros permite elaborar estrategias de filtrado con todos los descriptores que el documentalista requiera. Ello permite al programa clasificar las noticias antes de ser registradas en base de datos y facilita su posterior tratamiento estadístico.

Creados los filtros, el módulo de procesamiento pone en marcha el programa aplicando todos los parámetros previamente configurados.

Mientras XYZ funciona, los módulos de representación de noticias, monitorización en tiempo real, mapa de noticias, buscador y bloc permanecen activos para proporcionar ins-

trumentos que permitan realizar un seguimiento activo de los contenidos y publicaciones recuperadas.

### 4.3. Tablas y sets de datos

El programa XYZ opera una base de datos *MySQL* con tablas y sets de datos diseñados a medida para registrar:

- canales de sindicación activos y temporales;
- noticias activas, archivadas, pendientes de eliminar y filtradas;
- categorías temáticas;
- filtros que se utilizarán para clasificar los contenidos;
- datos de control de ejecución.

A continuación se exponen las tablas que utiliza y una muestra del set de datos que ilustra el método de almacenamiento utilizado.

#### Tabla *categories*

Registra las categorías clasificatorias para los canales de sindicación de contenidos, de acuerdo con un máximo de tres bloques o facetas. Por ejemplo en el set de datos de la tabla 2 se identifica que la categoría "Prensa digital" está disponible en el bloque 1 y su número de orden es 100.

Tabla 2. Ejemplo de set de datos de la tabla *categories*

```
INSERT INTO 'categories' SET id='1', type='1', category='Prensa digital', value='100';
```

#### Tabla *control*

Registra el último canal de sindicación analizado, el número parcial de ítems analizados, el tiempo global de funcionamiento en formato *UNIX* y el identificador del próximo canal de sindicación. En el ejemplo de la tabla 3 se muestra que el

*parser* con identificador 1 acaba de analizar el canal de sindicación 4.309, recopilando un total de 39.440 noticias en 214 ciclos de análisis en un tiempo de 3 horas y 36 minutos. También se indica que el próximo canal de sindicación que se analizará tiene el identificador 1.075

Tabla 3. Ejemplo de set de datos de la tabla *control*

```
INSERT INTO 'control' SET id='1', lastFeed='4309', parItems='39440', numLaps='214', timer='12960.5947565', start='1075';
```

### Tabla *filters*

Contiene los datos y parámetros de los filtros creados por el usuario para clasificar las noticias automáticamente. Concretamente almacena el título del filtro, descripción, palabras y frases clave para ser usadas con operadores booleanos. La tabla 4 presenta el filtro 1 relativo a movilidad sostenible cuyas palabras clave se guardan con el formato original introducido por el usuario y en un formato depurado, separando los descriptores por plecas (barras verticales). Además se añade un campo destinado a indexación que recoge todos los términos clave del filtro para facilitar su recuperación.

Tabla 4. Ejemplo de set de datos de la tabla *filters*

```
INSERT INTO 'filters' SET id='34', title='Term. Transporte subterráneo', description='Noticias especializadas en metro', filter1and='Metro', filter2and='metro', filter1or='suburbano, transporte, subterráneo, túnel, estación', filter2or='suburbano|transporte|subterráneo|túnel|estacion', filter1not='', filter2not='', filter1noise='Tren ligero', filter2noise='tren ligero', indexer='transporte subterráneo metro suburbano transporte subterráneo tunel estacion';
```

### Tablas *feeds* y *feedstemp*

La tabla *feeds* registra los canales de sindicación activos, que van a ser usados. Contiene campos para la metadescripción del canal, la asignación de categorías, así como el número de identificación del programa *parser* que se encargará de procesar el canal. Otros datos de interés corresponden al valor del canal, fuente o dominio de procedencia y campo de indexación con los términos clave ya depurados y normalizados.

La tabla *feedstemp* es una copia de la tabla *feeds* y sirve para almacenar temporalmente los canales de sindicación durante el proceso de importación.

Tabla 5. Ejemplo de set de datos de la tabla *feeds*

```
INSERT INTO 'feeds' SET id='4449', regdated='2016-06-16 16:24:28', title='Tendencias 21. Ciencia, tecnología, sociedad y cultura', link='http://www.tendencias21.net/xml/syndication.rss?r=190619', description='Revista electrónica de ciencia, tecnología, sociedad y cultura. ISSN 2174-6850', language='es', copyright='', managingEditor='', webMaster='', pubDate='Thu, 16 Jun 2016 16:00:01 +0200', lastBuildDate='Thu, 16 Jun 2016 16:00:01 +0200', category='Science', generator='', ttl='60', image='http://www.tendencias21.net/var/style/logo.jpg', source='http://tendencias21.net', cat1='1', cat2='3', cat3='5', priority='1', core='5', value='90', indexer='tendencias ciencia tecnologia sociedad cultura revista electronica ciencia tecnologia sociedad cultura issn 2174 6850 science';
```

### Tablas *items* e *items365*

La tabla principal *items* almacena las noticias y contenidos de cada canal de sindicación. Recoge el identificador del filtro con el que se clasifica la noticia y el identificador del canal de sindicación en el que fue publicada. Otros datos fundamentales son el código *hash* de la noticia, su fecha de registro, título, enlace, descripción o contenido, autor, categoría temática, comentarios, archivos adjuntos o encapsulados, enlace permanente, fecha de publicación, dominio de la fuente, código fuente original, notas y campo de indexación.

La tabla *items365* es una copia de la tabla *items*, cuya función es almacenar las noticias que no han sido filtradas y que son susceptibles de eliminación automática, al cabo de un período de tiempo previamente configurado por el usuario. De esta forma el programa restringe el almacenamiento a los contenidos deseados, filtrados o seleccionados para su conservación.

Tabla 6. Ejemplo de set de datos de la tabla *items*

```
INSERT INTO 'items' SET id='61901', idfilter='34', idfeed='4449', hash='52a838848f5140224013d0a427376cfe0efc0f4a', regdated='2016-06-27 14:30:04', title='Los nuevos sistemas de metro ligero son una solución sostenible y eficiente', link='http%3A%2F%2Fwww.tendencias21.net%2FLos-nuevos-sistemas-de-metro-ligero-son-una-solucion-sostenible-y-eficiente_a11267.html', description='Las unidades de metro ligero de última generación presentan interesantes ventajas como solución para los sistemas de transporte público...;', author='Pablo Javier Piacente', category='Transporte sostenible', comments='', enclosure='', guid='www.tendencias21.net%2C2016%3Arss-4110225', pubDate='Fri, 20 Apr 2012 09:45:00 +0200', source='http://tendencias21.net', htmlcontent='', notes='', indexer='nuevos sistemas metro ligero solucion sostenible eficiente unidades metro ligero ultima generacion presentan interesantes ventajas solucion sistemas transporte publico...;';
```

## 4.4. Programación del agregador

El entorno de desarrollo elegido para el agregador fue *Apache*, PHP y *MySQL* (AMP) por ser uno de los más extendidos y estandarizados. El método de programación empleado es de tipo estructurado y orientado a objetos. Esto significa que cada módulo dispone de un código de programación específico y unas rutinas que delimitan las opciones y funciones que se han predefinido en el apartado de especificaciones. Un ejemplo relevante de la metodología empleada, puede encontrarse en el algoritmo de procesamiento de noticias, encargado de revisar los canales de sindicación de los medios de comunicación, también conocido como *parser*. Su código estructurado permite inferir los siguientes pasos:

- 1) Cargar función de indexación, eliminación de etiquetas html y de iteración de tiempos para *benchmark* en tiempo real.
- 2) Selección de la base de datos.
- 3) Recuperación del identificador del último canal de sindicación analizado.
- 4) Liberación de memoria, antes de iniciar operaciones de recuperación.

- 5) Recuperación de los datos de la tabla control.
- 6) Calcular la fecha y hora de inicio y fin de ejecución.
- 7) Recuperación en tabla *feeds* del enlace al canal de sindicación que será consultado.
- 8) Envío de petición al servidor mediante funciones *cURL*.
- 9) Obtención del código fuente del canal de sindicación en formato XML.
- 10) Creación de un objeto DOM con el código fuente obtenido.
- 11) Selección de etiquetas contenedoras de las noticias mediante *XPath*.
- 12) Generación de *hash* identificativo de la noticia.
- 13) Comprobación de duplicados en base de datos usando el código *hash*.
- 14) Notación de tiempos de ejecución y número de noticias recopiladas para su supervisión.
- 15) Indexación del texto de las noticias.
- 16) Aplicación de filtros diseñados por el usuario.
- 17) Inserción de datos en la tabla *items*.
- 18) Eliminación de variables y liberación de memoria.
- 19) Iteración del proceso hasta la finalización de todos los canales de sindicación asignados al programa *parser*.

Un caso de programación orientada a objetos, corresponde al algoritmo de clasificación y filtrado. Su ejecución se produce únicamente cuando se declara su nombre clave y se transmiten las variables requeridas, devolviendo como resultado la clasificación positiva o negativa de las noticias. Su código a su vez es estructurado y se define con los siguientes pasos:

- 1) Recepción del texto depurado de la noticia.
- 2) Recuperación de los filtros definidos por el usuario y previamente almacenados en listas de tipo *array*.
- 3) Búsqueda de coincidencias parciales o totales de los términos del filtro conforme a la lógica booleana.
- 4) Cálculo de similaridad booleana en función de la frecuencia de aparición de los términos.
- 5) Almacenamiento del coeficiente de similaridad obtenido y cálculo del resto de filtros.
- 6) Selección del filtro con mejor coeficiente de similaridad, remitiendo su número de identificación.
- 7) En caso de que la noticia no sea clasificada, se devuelve el valor 0, que indica al programa que la noticia puede ser archivada.

La posibilidad de calcular el impacto y la correlación entre grupos de noticias puede facilitar la investigación de tendencias periodísticas y sociales en los medios de comunicación

El resultado del trabajo se traduce en más de 20.000 líneas de código, distribuidas en 53 archivos originales. También se utilizaron 7 librerías externas para desempeñar las funciones de alerta por correo electrónico, generación de gráficos estadísticos, representaciones gráficas y rutinas de procesamiento de textos de tipo *wysiwyg* que se encuentran debidamente citadas en el código fuente del programa.

Terminada la fase de programación, el agregador fue instalado en la plataforma portable *Wamp Server2go* (Haberker, 2007). Esto ha permitido que pueda ser utilizado

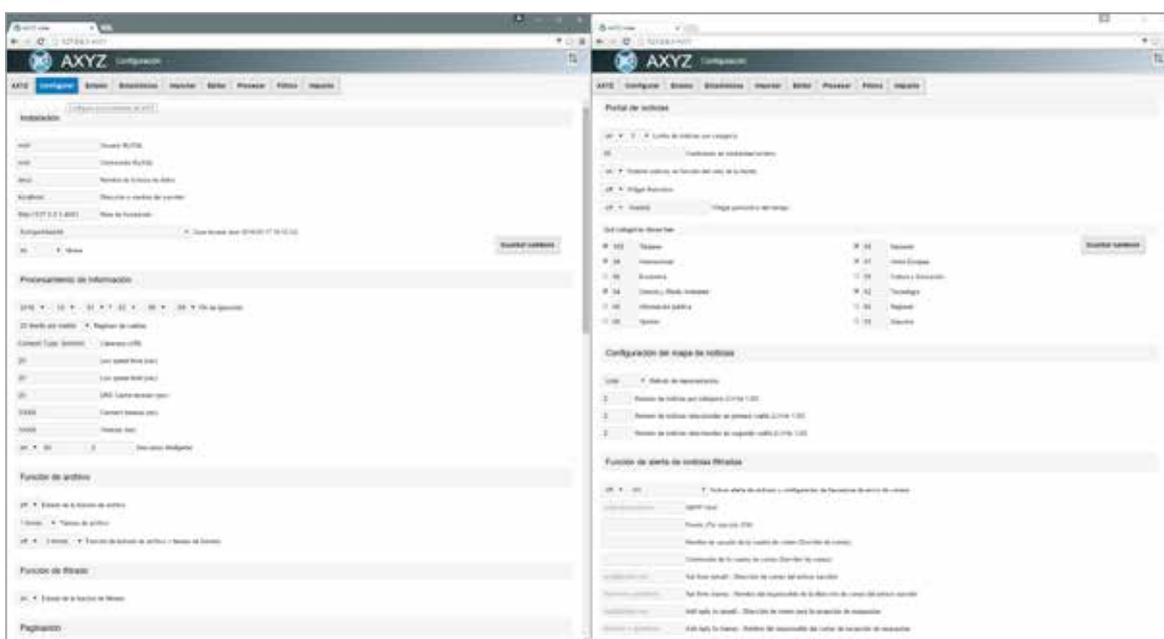


Figura 1. Módulo de configuración

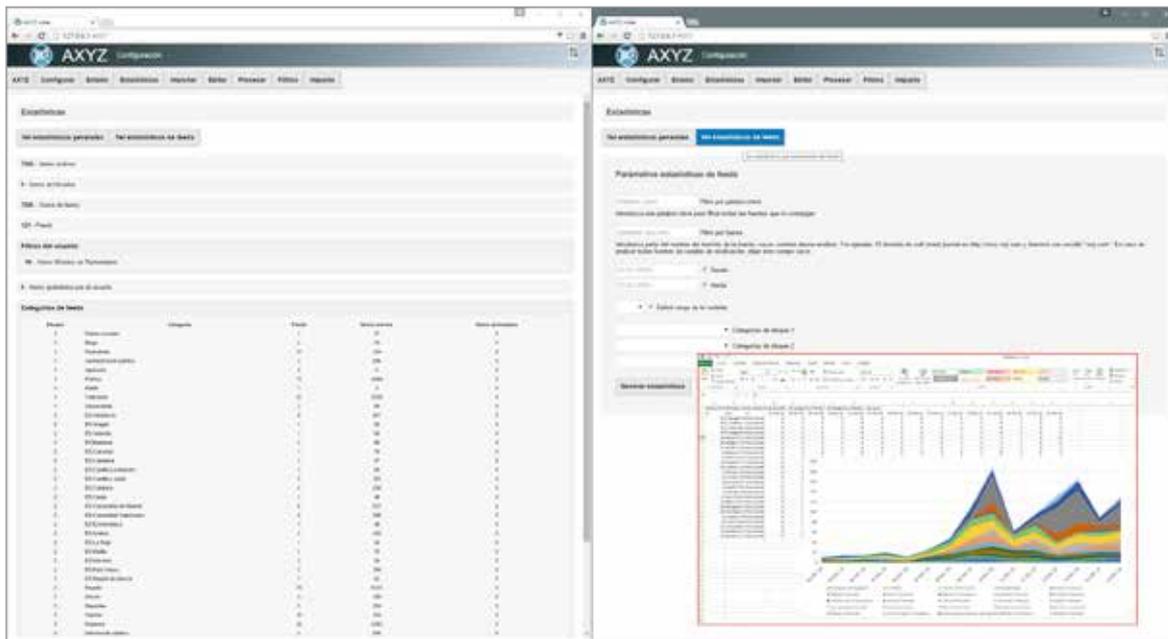


Figura 2. Opciones estadísticas del agregador

de forma local y autónoma sin necesidad de instalación o configuración. De hecho cualquier persona puede descargar el agregador en su ordenador y trabajar libremente como mejor considere, en el marco del derecho de acceso a la información pública, sin interferir en los dominios legales establecidos. El agregador, denominado XYZ, se encuentra disponible para su experimentación, explotación científica y académica en el repositorio de software libre *SourceForge.net* en la dirección:

<http://sourceforge.net/projects/axyznews>

### 5. Características destacables del agregador XYZ

El programa ha sido pensado para integrar todas las opciones y parámetros que controlan el funcionamiento de los módulos desde un módulo de configuración centralizado. Por ejemplo:

- temporizador de tareas
- progreso de la velocidad de recopilación de datos
- definición del método de archivo y expurgo
- filtrado de contenidos
- particularidades de la portada de noticias
- alertas informativas.

El módulo Estado supervisa las tablas de la base de datos, habilitando el mantenimiento, optimización o borrado de todos los contenidos recopilados, en caso de querer reiniciar la investigación en curso. En relación con las estadísticas, se proporcionan datos exactos sobre el número de contenidos activos, archivados, clasificados y distribuidos según categorías temáticas definidas por el usuario. También suministra datos sobre el número de canales de sindicación importados y su producción informativa según cronología, rangos temporales, fuente o palabras clave. La información puede ser exportada en hojas de cálculo para su posterior tabulación.

El módulo Importación se dedica a la inserción de listas de canales de sindicación, que pueden ser pre-clasificadas en torno a tres bloques de categorías temáticas, definidas por el administrador. También incorpora el concepto de valoración o importancia de las fuentes y la prioridad de procesamiento de las mismas. Por ejemplo un nivel de prioridad 1 forzaría al agregador a revisar con más asiduidad los canales de sindicación. Si además el administrador valora la fuente, el programa ponderará positivamente las noticias que ésta produzca, ayudando a organizar mejor los resultados. También incorpora las opciones de importación y exportación en formato OPML (*outline processor markup language*), permitiendo archivar listas de canales de sindicación.

El agregador XYZ puede ser descargado libremente para usos académicos y científicos en: <http://sourceforge.net/projects/axyznews>

Cuando se importan canales de sindicación en el agregador, éstos son validados y puestos a su disposición en el módulo de edición. Resulta destacable la capacidad para editar bloques de 100 canales de sindicación, facilitando la modificación masiva de sus títulos, descripciones, fuentes, dominios, prioridad, importancia, fechas y temática. Otra ventaja se basa en la aparición de señales gráficas que avisan del estado de la clasificación de los registros, de forma que puedan ser localizados inmediatamente para su subsanación.

Resuelta la edición de los canales, el agregador puede ponerse en marcha usando la función de Procesamiento. Esta función abre una ventana de monitorización estratigráfica de las tareas de recopilación de noticias en los canales de sindicación previamente registrados. Sin embargo presenta innovaciones que han multiplicado la capacidad del agrega-

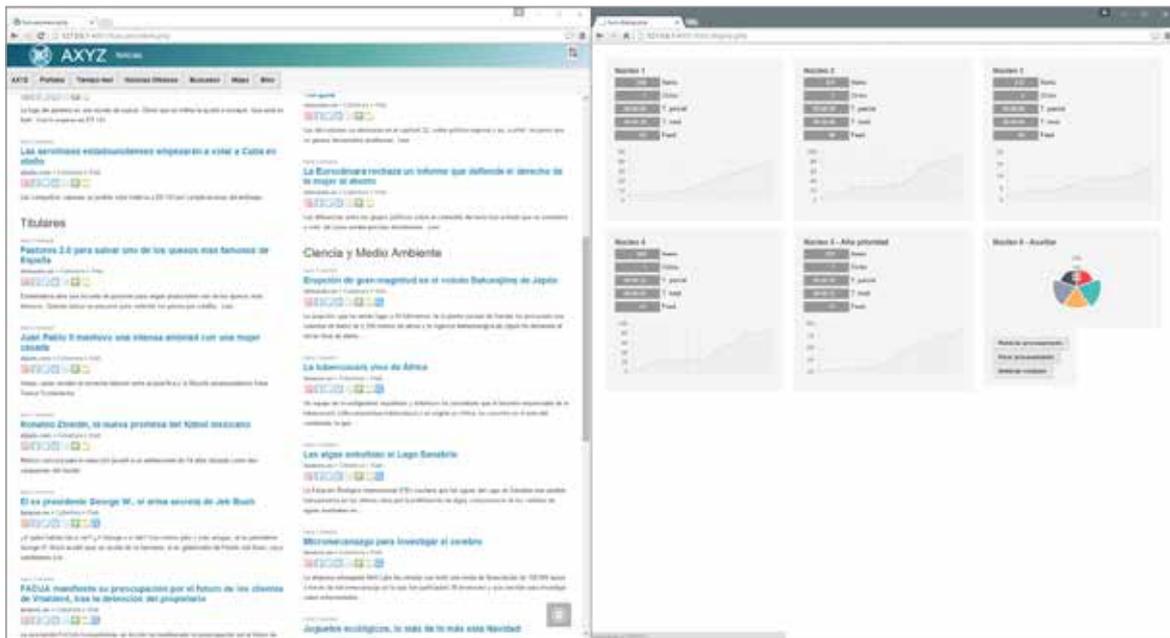


Figura 3. Portada de noticias y monitorización de la actividad del agregador

dor. Se trata de la introducción de 5 programas *parser* cuya ejecución independiente favorece la recuperación simultánea de las noticias actualizadas en los canales de sindicación. Esto se consigue asignando una lista de canales que será explorada por cada *parser* sin interferir en las operaciones de los demás. El trabajo colaborativo de los programas *parser* hace deseable la monitorización estadística de la actividad en tiempo real. Esto es, proporcionar al investigador información constante sobre:

- número de noticias recopiladas en cada *parser*;
- tiempo de ejecución parcial y total del proceso;
- número de ciclos o revisiones de los canales;
- número de identificación del canal que está siendo analizado.

Se ha prestado especial atención a completar todas las funciones que cabría esperar en un agregador de contenidos, incluyendo buscadores, grafos de noticias, portada personalizable, edición de filtros y migración de datos

Todo ello favorece que el investigador pueda controlar mejor el programa, detectar errores, fallos y realizar pruebas de rendimiento, que permitan verificar el funcionamiento del agregador y su comportamiento.

El módulo Filtros ha sido diseñado para recoger todos los criterios de clasificación que serán utilizados para discriminar la información. El método de recuperación aplicado es el booleano, con los operadores de intersección, unión, negación y ruido. Cada filtro es registrado en base de datos y en un archivo independiente con el que se genera un patrón de consulta con expresiones regulares. Este detalle aumenta significativamente la eficiencia y capacidad de filtrado

del agregador, que evita tener que emitir más consultas a la base de datos, obteniendo vía *NoSQL* los patrones que necesita para calcular la similitud de las noticias en relación con los filtros.

Tabla 7. Ejemplo de *array NoSQL* para el filtrado de noticias

```
$arrayFilters[] = array(id => "34", filterAND => "metro", filterOR => "suburbano|transporte|subterráneo|tunel|estacion", filterNOT => "", filterNOISE => "tren ligero");
```

Uno de los módulos de mayor utilidad para la investigación es Impacto. El programa comprende una interfaz para buscar las noticias recopiladas por el agregador y opciones para calcular el factor de impacto relativo y absoluto, según se explica en la tabla 8. Las fórmulas empleadas han sido diseñadas específicamente para poner en relación las noticias seleccionadas por el investigador con otras similares o bien con respecto a la totalidad. Esto permite conocer qué relevancia ha tenido un evento en los medios y qué cobertura absoluta ha representado en la producción informativa total. En cuanto al método para obtener noticias cuyo tema y contenido sea similar, se emplean consultas *SQL* a texto completo con el cuerpo de la noticia, usando su título y su clasificación o filtro si lo tuviera. El programa contabiliza en tal caso el número total de resultados cuyo coeficiente de similitud sea superior a 20. Esta medida obliga a que los contenidos recuperados tengan un alto nivel de semejanza con respecto a la noticia seleccionada, asegurando un cálculo más preciso. Sin embargo, el investigador puede modificar estos valores para perfeccionar su funcionamiento o satisfacer necesidades de cálculo específicas.

Cabe indicar que el modelo matemático está inspirado en el cálculo del factor de impacto aplicable a las revistas científicas (Garfield, 2006). Sin embargo, a diferencia de éste, no incluye el factor temporal en el denominador de las fór-

Tabla 8. Factores de impacto y correlación calculados en el agregador XYZ

$Q = Qt_1 + Qt_2 + Qt_n = Nr \& Sr$
<p>La consulta Q con unos términos <math>Qt_1, Qt_2</math> y <math>Qt_n</math> proporciona una serie de noticias relacionadas <math>Nr</math> y canales de sindicación relacionados <math>Sr</math></p>
$N_{dr \rightarrow nu} = \sum Nr \rightarrow (Qtnu_1 \cup Qtnu_2 \cup Qtnu_n) \geq sim = 20$
<p>El número de noticias directamente relacionadas con la seleccionada por el usuario <math>N_{dr \rightarrow nu}</math> es igual a la suma de noticias resultantes <math>\sum Nr</math>, de la consulta de los términos de dicha noticia (<math>Qtnu_1 \cup Qtnu_2 \cup Qtnu_n</math>), cuya similaridad sea igual o mayor que 20.</p>
$IFA_{nu} = \frac{N_{dr \rightarrow nu}}{\sum N}$
<p>El factor de impacto absoluto de la noticia seleccionada por el usuario <math>IFA_{nu}</math> es igual al número de noticias directamente relacionadas <math>N_{dr \rightarrow nu}</math> fraccionado por el número total de noticias recopiladas en el agregador <math>\sum N</math>.</p>
$IFR_{nu} = \frac{N_{dr \rightarrow nu}}{Nr}$
<p>El factor de impacto relativo de la noticia seleccionada por el usuario <math>IFR_{nu}</math> es igual al número de noticias directamente relacionadas <math>N_{dr \rightarrow nu}</math> fraccionado por el número total de noticias recuperadas en la consulta de origen <math>Nr</math>.</p>
$IFA_s = \frac{S_{dr \rightarrow nu}}{\sum S}$
<p>El factor de impacto absoluto entre fuentes <math>IFA_s</math> es igual al número de canales de sindicación directamente relacionados con la consulta del usuario <math>S_{dr \rightarrow nu}</math> fraccionado por el número total de canales de sindicación registrados en el agregador <math>\sum S</math>.</p>
$IFR_s = \frac{S_{dr \rightarrow nu}}{Sr}$
<p>El factor de impacto relativo entre fuentes <math>IFR_s</math> es igual al número de canales de sindicación directamente relacionados con la consulta del usuario <math>S_{dr \rightarrow nu}</math> fraccionado por el número de canales de sindicación <math>Sr</math> relacionados con la consulta original Q.</p>

mulas, ya que la frecuencia de publicación en los medios de comunicación es muy elevada. Por ello su aplicación se aconseja en rangos cronológicos breves de horas, días, semanas o pocos meses.

También se ha programado un método para calcular la correlación aproximada entre dos grupos de noticias filtradas. El algoritmo tiene en cuenta los siguientes pasos:

- 1) El investigador selecciona una noticia A, que será comparada y correlacionada con otra denominada B.
- 2) Extracción de las noticias similares en función de su coeficiente de similaridad mínimo.
- 3) Creación de listas de noticias similares para A y B.
- 4) Cálculo del coeficiente de similaridad de cada noticia de la lista A para cada noticia de la lista B.
- 5) Obtención de listas de coeficientes de similaridad entre el grupo de noticias de A con respecto a B y viceversa.

6) Aplicación de la fórmula de correlación de Pearson para calcular la correlación entre dos variables que corresponden a los coeficientes de similaridad obtenidos por el grupo de noticias de A y B. El resultado obtenido se puede exportar en formato CSV para su tratamiento estadístico y con ello obtener un diagrama de dispersión que permita determinar el grado de correlación obtenido y su desviación típica. Esta información ayudaría a los investigadores a trabajar con grandes cantidades de datos y demostrar que se suceden correlaciones entre diversos eventos y contenidos publicados.

Otros módulos de interés son los vinculados con las funciones de representación y consulta como la Portada de noticias y el Buscador. En ambos casos se proporcionan métodos de filtrado según categoría temática, intervalos temporales y dominio. En este marco también se encuentra el módulo de noticias en tiempo real, que muestra las últimas 15 informaciones recopiladas con un intervalo de refresco de 30 segundos.

También es destacable el módulo Mapa de noticias, ya que hace posible la creación de un grafo arborescente o circular, con las noticias relacionadas de cada categoría temática. En su desarrollo se ha utilizado la librería *Cytoscape.js* (Franz et al., 2016) diseñada para la representación de redes. En este caso *Cytoscape* ha sido adaptado a la representación de las noticias, mostrando el título y el enlace de las mismas, así como su relación por afinidad. El criterio de selección usado por el programa es el grado de actualidad de las noticias, eligiendo siempre las más recientes y el coeficiente de similaridad basado en la materia, textos y titulares recuperados. Finalmente el agregador tiene la capacidad de generar dossiers de noticias filtradas y almacenar las noticias favoritas del lector, incluyendo también en este caso un buscador específico.

## 6. Conclusiones

El agregador XYZ cumple los objetivos de la investigación aportando soluciones para calcular la repercusión e impacto mediático de la información, añadiendo funciones estadísticas personalizadas que cuantifican la producción informativa en las fuentes y canales de sindicación registrados. Por otra parte también es posible obtener el coeficiente de correlación entre dos eventos y sus noticias próximas.

En cuanto a rendimiento, el agregador proporciona un sistema de 5 programas *parser* colaborativos, que ayudan a redistribuir el trabajo de recopilación de contenidos y con ello aumentar la velocidad y capacidad de procesamiento. Además habilita la supervisión de su funcionamiento, a través de una ventana de monitorización que suministra datos y gráficas estadísticas en tiempo real. Ello contribuye a detectar errores si los hubiere y también a medir su rendimiento. Otras funciones adaptadas a los *big data* son los módulos de edición y de clasificación automática, diseñados para tratar bloques de registros y reducir la concurrencia de consultas en la base de datos.

La clasificación automática de contenidos se aplica al mismo tiempo que son recopilados en sus respectivas fuentes. El agregador aplica todos los filtros definidos por el administrador a

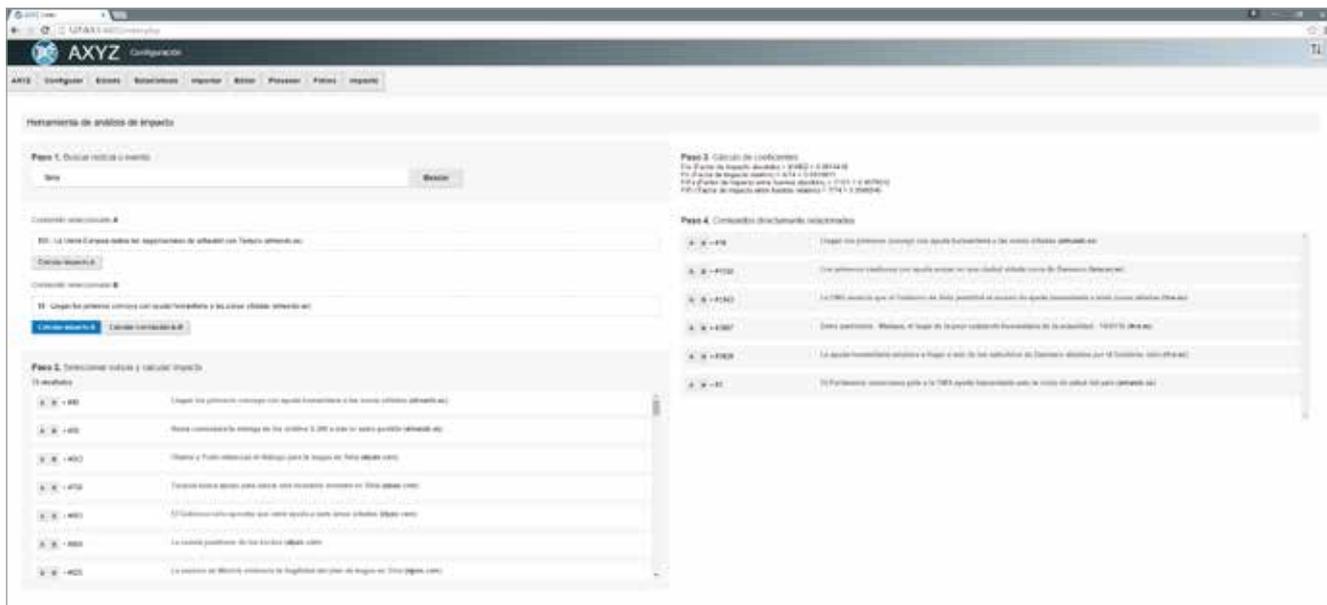


Figura 4. Módulo para el cálculo de impacto y correlación entre noticias

través de expresiones regulares, que combinan los términos y palabras clave con el método booleano de recuperación. A diferencia de otros programas, XYZ genera un archivo *NoSQL* con las expresiones de consulta, al que se acude para clasificar la información. Esto evita tener que recurrir a la base de datos por cada noticia que se recuperada, lo que supone un ahorro del 50% en el número de peticiones al servidor.

El programa ha sido creado con todos los módulos que cabría esperar en un agregador de contenidos. Además incorpora otros menos comunes como:

- grafo de noticias relacionadas
- vista de noticias en tiempo real

- generador de cobertura informativa
- métodos de consulta predefinida en buscadores.

Las limitaciones que plantea la *Ley 21/2014 de 4 de noviembre de propiedad intelectual (España, 2014)* constituyen una amenaza para el desarrollo de tecnologías basadas en sindicación de contenidos. Afortunadamente el agregador XYZ ha sido diseñado para funcionar en entorno local, lo que garantiza el derecho de acceso a la información publicada en medios de comunicación. Ello es debido a que la ley mantiene la excepción del derecho de copia privada permitiendo el uso de agregadores en el entorno particular, científico y académico. Cualquier investigador puede descargar el agrega-

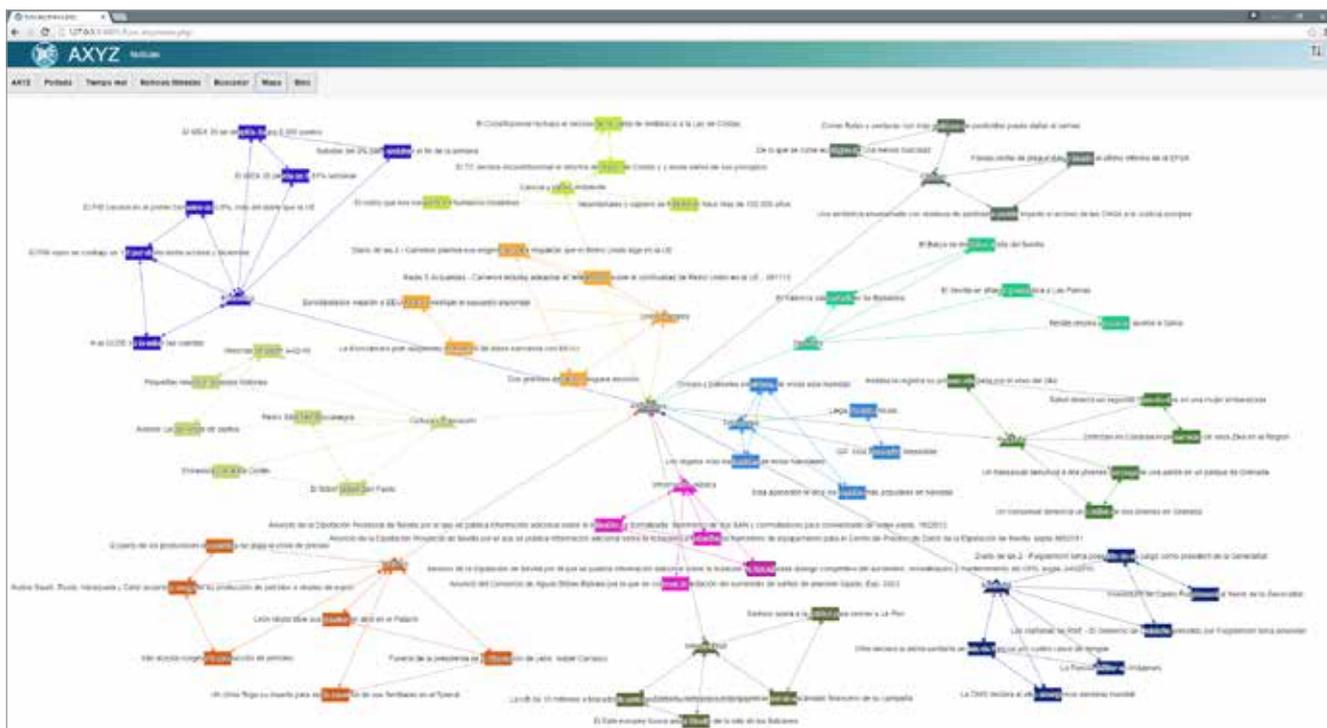


Figura 5. Mapa interactivo y relacional de noticias

dor y obtener un sistema de información equivalente a los extintos *Google Reader* y *Google News* (en España). Por otra parte puede conformar sus propias colecciones de fuentes, canales y noticias para iniciar proyectos de investigación informativa y documental de tipo métrico y cualitativo.

El agregador puede ser mejorado en sucesivas versiones atendiendo a la capacidad de inteligencia artificial para discernir las tendencias de consulta, corregir dinámicamente la relevancia de los contenidos basada en la experiencia del lector, ampliar las opciones de generación de grafos o enriquecer las noticias recopiladas con datos obtenidos a través de técnicas de rastreo web. También pueden mencionarse nuevas líneas de trabajo como el benchmarking comparativo entre agregadores, poner en práctica portales de canales de sindicación, análisis léxico, aplicaciones en otras áreas de conocimiento y adaptación del agregador para el intercambio de información bibliográfica o documental.

## Bibliografía

- Bansal, Srividya K.; Kagemann, Sebastian** (2015). "Integrating big data: A semantic extract-transform-load framework". *Computer*, v. 48, n. 3, pp. 42-50.  
<http://doi.ieeecomputersociety.org/10.1109/MC.2015.76>
- Bazargani, Sahar; Brinkley, Julian; Tabrizi, Nassehzadeh** (2013). "Implementing conceptual search capability in a cloud-based feed aggregator". En: *3<sup>rd</sup> Intl conf on innovative computing technology (Intech)*, 29-31 Aug., pp. 138-143.  
<http://dx.doi.org/10.1109/INTECH.2013.6653631>
- BuiltWith* (2016). *CMS usage statistics. Statistics for web-sites using CMS technologies*.  
<http://trends.builtwith.com/cms>
- Carlson, Matt; Usher, Nikki** (2015). "News startups as agents of innovation: For-profit digital news startup manifestos as metajournalistic discourse". *Digital journalism*, v. 4, n. 5, pp. 1-19.  
<http://dx.doi.org/10.1080/21670811.2015.1076344>
- Chen, Philip; Zhang, Chun-Yang** (2014). "Data-intensive applications, challenges, techniques and technologies: A survey on big data". *Information sciences*, v. 275, pp. 314-347.  
<https://goo.gl/kQJM5H>  
<http://dx.doi.org/10.1016/j.ins.2014.01.015>
- Chen, Weiqin; Bøen, Torbjørn** (2008). "A personalized RSS news filtering agent". En: Ellis, Richard; Allen, Tony; Petridis, Miltos. *Applications and innovations in intelligent systems XV*. Londres: Springer, pp. 321-326. ISBN: 978 1 84800 086 5  
[http://dx.doi.org/10.1007/978-1-84800-086-5\\_25](http://dx.doi.org/10.1007/978-1-84800-086-5_25)
- Colle, Raymond** (2013). "Prensa y big data: el desafío de la acumulación y análisis de datos". *Mediterranean journal of communication*, v. 4, n. 1, pp. 275-282.  
<http://dx.doi.org/10.14198/MEDCOM2013.4.1.13>
- Creus, Jordi; Amann, Bernd; Travers, Nicolas; Vodislav, Dan** (2011). "RoSeS: A continuous content-based query engine for RSS feeds". En: *Procs of the 20<sup>th</sup> ACM Intl conf on information and knowledge management*, 2011, pp. 2549-2552.  
[http://cedric.cnam.fr/fichiers/art\\_2086.pdf](http://cedric.cnam.fr/fichiers/art_2086.pdf)  
<http://dx.doi.org/10.1145/2063576.2064016>
- Cuzzocrea, Alfredo** (2015). "Aggregation and multidimensional analysis of big data for large-scale scientific applications: models, issues, analytics, and beyond". En: *Procs of the 27<sup>th</sup> Intl conf on scientific and statistical database management*, pp. 23.  
<http://dx.doi.org/10.1145/2791347.2791377>
- España (2014). "Ley 21/2014, de 4 de noviembre, por la que se modifica el texto refundido de la Ley de propiedad intelectual, aprobado por Real decreto legislativo 1/1996, de 12 de abril, y la Ley 1/2000, de 7 de enero, de enjuiciamiento civil". *BOE*, n. 268, 5 de noviembre, pp. 90404-90439.  
[https://www.boe.es/diario\\_boe/txt.php?id=BOE-A-2014-11404](https://www.boe.es/diario_boe/txt.php?id=BOE-A-2014-11404)
- Franz, Max; Lopes, Christian T.; Huck, Gerardo; Dong, Yue; Sumer, Onur; Bader, Gary D.** (2016). "Cytoscape.js: a graph theory library for visualisation and analysis". *Bioinformatics*, v. 32, n. 2, pp. 309-311.  
<http://dx.doi.org/10.1093/bioinformatics/btv557>
- Gallé, Matthias; Renders, Jean-Michel; Karstens, Eric** (2013). "Who broke the news?: an analysis on first reports of news events". En: *Procs of the 22<sup>nd</sup> Intl conf on World Wide Web companion*, pp. 855-862.  
<http://www2013.org/companion/p855.pdf>  
<http://dx.doi.org/10.1145/2487788.2488066>
- Garfield, Eugene** (2006). "The history and meaning of the journal impact factor". *Jama*, v. 295, n. 1, pp. 90-93.  
<http://garfield.library.upenn.edu/papers/jamajif2006.pdf>  
<http://dx.doi.org/10.1001/jama.295.1.90>
- Guallar, Javier** (2015). "Prensa digital en 2013-2014". *Anuario ThinkEPI*, v. 9, pp. 153-160.  
<http://dx.doi.org/10.3145/thinkepi.2015.37>
- Guallar, Javier; Leiva-Aguilera, Javier** (2013). *El content curator. Guía básica para el nuevo profesional de internet*. Barcelona: UOC. Colección El profesional de la información, n. 24. ISBN: 978 84 9064 018 0
- Haber Kern, Timo** (2007). *Server2Go*.  
[http://www.server2go-download.de/download/server2go\\_a22\\_psm.zip](http://www.server2go-download.de/download/server2go_a22_psm.zip)
- Hmedeh, Zeinab; Vouzoukidou, Nelly; Travers, Nicolas; Christophides, Vassilis; Du-Mouza, Cedric; Scholl, Michel** (2011). "Characterizing web syndication behavior and content". En: Bouguettaya, Athman; Hauswirth, Manfred; Liu, Ling. *Web information system engineering (2011)*. Sydney: Springer, pp. 29-42. ISBN: 978 3 642 24434 6  
[http://cedric.cnam.fr/fichiers/art\\_2162.pdf](http://cedric.cnam.fr/fichiers/art_2162.pdf)  
[http://dx.doi.org/10.1007/978-3-642-24434-6\\_3](http://dx.doi.org/10.1007/978-3-642-24434-6_3)
- Horincar, Roxana; Amann, Bernd; Artières, Thierry** (2010). "Best-effort refresh strategies for content-based RSS feed aggregation". En: Chen, Lei; Triantafyllou, Peter; Suel, Torsten. *Web information systems engineering (2010)*. Hong Kong: Springer, pp. 262-270. ISBN: 978 3 642 17616 6  
[http://dx.doi.org/10.1007/978-3-642-17616-6\\_24](http://dx.doi.org/10.1007/978-3-642-17616-6_24)
- Isah, Haruna** (2012). "Full data controlled web-based feed aggregator". *International journal of computer science & information technology*, v. 4, n. 3, pp. 71-84.  
<http://dx.doi.org/10.5121/ijcsit.2012.4307>

- Katakis, Ioannis; Tsoumakas, Grigorios; Banos, Evangelos; Bassiliades, Nick; Vlahavas, Ioannis** (2009). "An adaptive personalized news dissemination system". *Journal of intelligent information systems*, v. 32, n. 2, pp. 191-212. <http://dx.doi.org/10.1007/s10844-008-0053-8>
- Leaver, Trama; Willson, Michele; Balnaves, Mark** (2012). "Transparency and the ubiquity of information filtration?". *Ctrl-Z: New media philosophy*, v. 1, n. 2. <http://www.ctrl-z.net.au/articles/leaver-willson-balnaves-transparency-and-the-ubiquity-of-information-filtration>
- Lee, Bum-Suk; Im, Jin-Woo; Hwang, Byung-Yeon; Zhang, Du** (2008). "Design of an RSS crawler with adaptive revisit manager". En: *SEKE*, 2008, pp. 219-222. <http://dblp.uni-trier.de/db/conf/seke/seke2008.html>
- Li, Xin; Yan, Jun; Deng, Zhihong; Ji, Lei; Fan, Weiguo; Zhang, Benyu; Chen, Zheng** (2007). "A novel clustering-based RSS aggregator". En: *Procs of the 16<sup>th</sup> Intl conf on World Wide Web*, 2007, pp. 1309-1310. ISBN: 978 1 59593 654 7 <http://www2007.org/posters/poster931.pdf> <http://dx.doi.org/10.1145/1242572.1242824>
- López-Maza, Sebastián** (2015). "El límite sobre agregadores y buscadores". En: Rodríguez-Cano, Rodrigo B. *La reforma de la Ley de propiedad intelectual*. Valencia: Tirant lo Blanch, pp. 89-111. ISBN: 978 84 9086 664 1
- Marty, Emmanuel; Rebillard, Franck; Smyrniaios, Nikos; Touboul, Annelise** (2010). "Variété et distribution des sujets d'actualité sur Internet. Une analyse quantitative de l'information en ligne". *Mots. Les langages du politique*, n. 93, pp. 107-126. <http://dx.doi.org/10.4000/mots.19832>
- Mayer-Schönberger, Viktor; Cukier, Kenneth** (2013). *Big data: la revolución de los datos masivos*. Madrid: Turner. ISBN: 978 84 15427 81 0
- Messina, Alberto; Montagnuolo, Maurizio** (2009). "A generalised cross-modal clustering method applied to multimedia news semantic indexing and retrieval". En: *Procs of the 18<sup>th</sup> Intl conference on World Wide Web*, 2009, pp. 321-330. ISBN: 978 1 60558 487 4 <http://ra.ethz.ch/CDstore/www2009/proc/docs/p321.pdf> <http://dx.doi.org/10.1145/1526709.1526753>
- O'Riordan, Adrian P.; O'Mahoney, Oliver** (2011). "Engineering an open web syndication interchange with discovery and recommender capabilities". *Journal of digital information*, v. 12, n. 1. <https://cora.ucc.ie/handle/10468/980>
- Reichert, Sandro; Urbansky, David; Muthmann, Klemens; Katz, Philipp; Wauer, Matthias; Schill, Alexander** (2011). "Feeding the world: a comprehensive dataset and analysis of a real world snapshot of web feeds". En: *Procs of the 13<sup>th</sup> Intl conf on information integration and web-based applications and services*, 2011, pp. 44-51. ISBN: 978 1 4503 0784 0 <http://dx.doi.org/10.1145/2095536.2095546>
- Rodríguez-Cano, Rodrigo B.** (2015). "Tasa Google o canon AEDE: una reforma desacertada". *Aranzadi civil-mercantil. Revista doctrinal*, v. 1, n. 11, pp. 53-94.
- Samper, Juan J.; Castillo, Pedro A.; Araujo, Lourdes; Merello, Juan J.; Cordon, Oscar; Tricas, Fernando** (2008). "NectarRSS, an intelligent RSS feed reader". *Journal of network and computer applications*, v. 31, n. 4, pp. 793-806. <http://dx.doi.org/10.1016/j.jnca.2007.09.001>
- Severo, Marta; Beauguitte, Laurent; Pecout, Hugues** (2015). "Archiving news on the Web through RSS flows. A new tool for studying international events". En: *Resaw. Web archives as scholarly sources: Issues, practices and perspectives*. <https://halshs.archives-ouvertes.fr/halshs-01187828>
- Sia, Ka-Cheung; Cho, Junghoo; Cho, Hyun-Kyu** (2007). "Efficient monitoring algorithm for fast news alerts". *Knowledge and data engineering*, v. 19, n. 7, pp. 950-961. <http://dx.doi.org/10.1109/TKDE.2007.1041>
- Thelwall, Mike; Prabowo, Rudy; Fairclough, Ruth** (2006). "Are raw RSS feeds suitable for broad issue scanning? A science concern case study". *Journal of the American Society for Information Science and Technology*, v. 57, n. 12, pp. 1644-1654. <http://dx.doi.org/10.1002/asi.20334>
- Travers, Nicolas; Hmedeh, Zeinab; Vouzoukidou, Nelly; Du-Mouza, Cedric; Christophides, Vassilis; Scholl, Michel** (2014). "RSS feeds behavior analysis, structure and vocabulary". *International journal of web information systems*, v. 10, n. 3, pp. 291-320. <http://dx.doi.org/10.1108/IJWIS-06-2014-0023>

## El profesional de la información

<http://www.elprofesionaldeinformacion.com/autores.html>

### PRÓXIMOS TEMAS

Número	Mes año	Tema	Envío textos
25, 5	Sept 2016	Evaluación de la ciencia	20 mayo 2016
25, 6	Nov 2016	TIC para información y comunicación	10 julio 2016
26, 1	Ene 2017	Públicos vulnerables y empoderamiento digital	10 sept 2016
26, 2	Mar 2017	Ética, investigación y comunicación	10 nov 2016
26, 3	May 2017	Información pública	10 enero 2017
26, 4	Jul 2017	Comunicación política	10 marzo 2017

# INDICADORES



## PROPUESTA DE RANKINGS DE UNIVERSIDADES ESPAÑOLAS EN REDES SOCIALES

A proposal of rankings of Spanish universities in social networking sites



**Carmen Zarco, Salvador Del-Barrio-García y Óscar Cordon**



**Carmen Zarco** es doctora en documentación científica y master en marketing y comportamiento del consumidor por la *Universidad de Granada*. Su vida profesional se ha desarrollado durante más de quince años en varias empresas como responsable de gestión del conocimiento. A lo largo de su carrera ha publicado una decena de artículos científicos, algunos de ellos en revistas *JCR*, y ha participado en varios proyectos y contratos de investigación. Es profesora asociada en la *Universidad Internacional de La Rioja*, donde imparte clase en el grado de ADE y en el MBA.  
<http://orcid.org/0000-0002-2530-249X>

*Universidad Internacional de la Rioja, Departamento de Comercialización e Investigación de Mercados  
Gran Vía Rey Juan Carlos I, 41. 26002 Logroño, España  
carmen.zarco@unir.net*



**Salvador Del-Barrio-García** es profesor titular de comercialización e investigación de mercados en la *Universidad de Granada*. Sus áreas de especialización son la comunicación integrada de marketing, el comportamiento del consumidor online, marketing cross-cultural y marketing turístico. Ha dirigido 3 tesis doctorales y ha publicado diversos libros y capítulos de libros en editoriales internacionales de prestigio como *Springer*, *Wiley* e *IGI Global*. Ha publicado más de 30 artículos en revistas científicas de prestigio, entre ellas *International journal of advertising*, *Journal of advertising research*, *European journal of marketing*, *Crosscultural management: An international journal*, *Computers in human behaviour*, *Business services*, *Cornell hospitality quarterly*, *Cities*, y *Public relations review*.  
<http://orcid.org/0000-0002-6144-0240>

*Universidad de Granada, Departamento de Comercialización e Investigación de Mercados  
Campus Universitario de Cartuja, s/n. 18071, Granada, España  
dbarrio@ugr.es*



**Óscar Cordon**, catedrático de la *Universidad de Granada*, es director del *Laboratorio de Soft Computing Applications for Complex EnviRonments (Soccer)* y delegado de la Rectora (Vicerrector) para la universidad digital. Ha dirigido 16 tesis doctorales y publicado más de 100 artículos en revistas internacionales, 88 de ellos en revistas *JCR*. Ha coordinado 26 proyectos y contratos de investigación. Es *Premio nacional de informática Aritmel 2014*, *Premio IEEE computational intelligence society outstanding early career award 2011* y *Premio a la trayectoria de jóvenes investigadores de la UGR 2004*.  
<http://orcid.org/0000-0001-5112-5629>

*Universidad de Granada, Departamento de Ciencias de la Computación e Inteligencia Artificial  
Periodista Daniel Saucedo Aranda, s/n. 18071, Granada, España  
ocordon@decsai.ugr.es*

### Resumen

La elaboración de rankings es una aproximación cada vez más popular para la evaluación de resultados de las universidades. Las propuestas existentes de rankings de universidades españolas en redes sociales presentan limitaciones como su falta de cobertura, su especificidad y especialmente su falta de transparencia en la metodología de definición. El objetivo de este trabajo es proponer un marco general de diseño de clasificaciones coherentes para analizar el impacto de todas las universidades españolas, tanto públicas como privadas, en varias redes sociales (*Twitter*, *Facebook*, *LinkedIn* y *YouTube*). El

Artículo recibido el 25-10-2015  
Aceptación definitiva: 04-04-2016

marco global propuesto está basado en la definición de una taxonomía de criterios, la selección de los disponibles en las cuatro redes sociales mencionadas y la construcción de distintos tipos de indicadores a partir de ellos. Para ilustrarlo, presentamos dos casos de estudio relacionados con la presencia y la influencia de las universidades españolas en dos de esas redes, *Twitter* y *LinkedIn*.

## Palabras clave

Rankings de universidades; Redes sociales; *Facebook*; *Twitter*; *YouTube*; *LinkedIn*.

## Abstract

University rankings have become popular tools to evaluate the results of academic institutions. The existing proposals of rankings of universities in Spain, based on social networking sites, show some limitations related to their lack of coverage, their specificity, and their lack of transparency in the ranking definition. The goal of this work is to introduce a general design framework for coherent classifications to analyze the impact of all Spanish universities, both public and private, in four social networking sites: *Twitter*, *Facebook*, *LinkedIn*, and *YouTube*. The proposed framework is based on the definition of a criteria taxonomy, the selection of those criteria available in the four social networking sites, and the proposal of various kinds of indicators using them. To illustrate our framework, two case studies are developed evaluating the presence and influence of Spanish universities in *Twitter* and *LinkedIn*.

## Keywords

University rankings; Social networking sites; *Facebook*; *Twitter*; *YouTube*; *LinkedIn*.

Zarco, Carmen; Del-Barrio-García, Salvador; Cordón, Óscar (2016). "Propuesta de rankings de universidades españolas en redes sociales". *El profesional de la información*, v. 25, n. 4, pp. 684-698.

<http://dx.doi.org/10.3145/epi.2016.jul.18>

## 1. Introducción

### 1.1. redes sociales y universidades

Los progresos tecnológicos que se han ido produciendo en el campo de la comunicación siempre han sido objeto de investigación dentro de las ciencias sociales al generar nuevas formas de relación social que provocan una transformación en los hábitos y costumbres de la sociedad. Las redes sociales virtuales (en adelante RS) nos permiten interactuar con otras personas aunque no las conozcamos. Estas redes cuentan con una serie de herramientas tecnológicas muy sencillas de utilizar y posibilitan la creación de comunidades de personas en las que se establece un intercambio dinámico y comunicativo (O'Keeffe; Clarke-Pearson, 2011).

Las RS son especialmente adecuadas para alcanzar a uno de los públicos objetivos de mayor interés para las instituciones educativas, los estudiantes, que son los usuarios que más rápidamente las han adoptado

En este contexto, la incursión de las RS supone un salto cualitativo que potencia significativamente los modos de comunicación y generación del conocimiento de los escenarios involucrados, concretamente en la enseñanza (Gonzalo-Brito; Laaser; Toloza, 2012). Las RS son especialmente adecuadas para alcanzar a uno de los públicos objetivos de mayor interés para las instituciones educativas, los estudiantes, que son los usuarios que más rápidamente las han adoptado. Tienen todo el potencial para convertirse en un recurso valioso para conectar e implicar a este colectivo y para ser utilizadas in-

cluso como herramientas docentes (Del-Barrio-García; Arquero; Romero-Frías, 2015; Roblyer *et al.*, 2010). Por esta razón, las universidades han sido pioneras en incorporar su uso como herramienta de comunicación y han comenzado a otorgarle más importancia dentro de su estrategia de comunicación institucional (Valerio, 2015). El impacto de las RS en España es muy destacado, siendo uno de los países con mayor tasa de penetración. Concretamente, según datos de la tercera oleada del *Estudio general de medios* (octubre-noviembre) (EGM 2015), el 64,2% de la población frecuenta alguna red social (AIMC, 2015). De acuerdo con el informe internacional de *GlobalWebIndex*, las RS ocupan ya un tercio del tiempo online de los usuarios españoles, los cuales son activos en 3,5 RS de media (GWI, 2016). De todas ellas, las más utilizadas son (IAB, 2015)<sup>1</sup>:

- *Facebook* (96%)
- *YouTube* (66%)
- *Twitter* (56%)
- *Google+* (34%)
- *LinkedIn* (31%)

Como consecuencia de este crecimiento en el uso de las RS, especialmente entre los jóvenes, la mayoría de las universidades españolas se han sumado a este movimiento y han abierto perfiles institucionales en los últimos años (Paniagua-Rojano; Gómez-Calderón, 2012; Gómez-Calderón; Paniagua-Rojano, 2014; Simón-Onieva, 2014; González-Díaz; Iglesias-García; Codina, 2015).

### 1.2. Rankings de universidades

Las clasificaciones académicas de universidades o rankings<sup>2</sup> de universidades han experimentado un gran avance en los últimos años. Estos rankings tienen por objeto realizar jerarquizaciones de las instituciones de educación superior

teniendo en cuenta diversos indicadores que buscan medir aspectos relacionados con la calidad de la investigación, la docencia u otros aspectos de la actividad académica para informar y orientar a los estudiantes, investigadores y público en general (Liu, 2015; Tomàs-Folch et al., 2015).

Cada vez se tienen más en cuenta como medidas para evaluar la calidad o el prestigio de una universidad. Su estrecha incidencia sobre la opinión pública y su canalización a través de los medios de comunicación permitiría caracterizarlos como una herramienta de marketing que promociona las universidades mejor calificadas (Ordorika; Rodríguez-Gómez, 2010).

Existe un gran número de rankings universitarios de distinta naturaleza a nivel nacional e internacional, muchos de ellos basados en criterios relacionados con la calidad y la cantidad de la producción investigadora. Entre los rankings de universidades del mundo destaca el *Academic Ranking of World Universities (ARWU)*, más conocido como *Ranking de Shanghai*, considerado actualmente como el referente para conocer la posición de las universidades a nivel mundial (Docampo, 2012; Docampo et al., 2012).  
<http://www.shanghairanking.com/es>



El suplemento educativo de *Times, Higher*, publica también un ranking mundial anual de universidades desde 2004: *Times Higher Education (THE) World University Rankings*. A partir de 2010 comenzó a colaborar con *Thomson Reuters* utilizando una nueva metodología (Ordorika; Rodríguez-Gómez, 2010).  
<https://www.timeshighereducation.com/world-university-rankings>

Otros rankings internacionales de referencia son:

- *QS-Ranking*  
<http://www.topuniversities.com/university-rankings>
- *SCImago Institutions Rankings*  
<http://www.scimagoir.com>
- *Leiden Ranking*  
<http://www.leidenranking.com>



- *U-Multirank*  
<http://www.umultirank.org>



- *NTU Ranking*  
<http://nturanking.lis.ntu.edu.tw>
- *URAP*  
<http://www.urapcenter.org/2015>

Todos estos rankings se basan principalmente en indicadores de producción científica.

Existen también rankings internacionales que se han centrado en clasificar las universidades en función de su presencia web. En este sentido cabe destacar el *Ranking Mundial de Universidades en la Web (Webometrics)*, una iniciativa del *Laboratorio de Cibermetría* del CSIC. La intención de este ranking es destacar la importancia de las publicaciones en web como forma de transmisión de conocimiento académico además de medir el rendimiento y el impacto de la actividad científica.  
<http://www.webometrics.info>

En lo referente a rankings de universidades españolas, encontramos:

- *Ranking Buela-Casal (Buela-Casal et al., 2009)*: cuyo objetivo es analizar la productividad científica de las universidades;
- *Rankings I-UGR*: una serie de rankings basados en la investigación publicada en las revistas internacionales de mayor impacto (Torres-Salinas et al., 2011).  
<http://www.rankinguniversidades.es>

Otra iniciativa que ha destacado recientemente es el proyecto *U-Ranking* de la *Fundación BBVA* y el *Instituto Valenciano de Investigaciones Económicas*, que presenta importantes singularidades como el hecho de permitir estudiar tres dimensiones: docencia, investigación e innovación, y desarrollo tecnológico, así como realizar dos tipos de formatos.  
<http://www.u-ranking.es/index2.php>

Finalmente, es preciso mencionar las diversas clasificaciones de las universidades españolas que realiza el *Observatorio IUNE* en función a una serie de indicadores de volumen de profesorado, sexenios y premios nacionales obtenidos, actividad científica, innovación, competitividad y capacidad formativa.  
<http://www.iune.es>



### 1.3. Rankings de universidades en redes sociales

Hasta donde alcanza nuestro conocimiento, son prácticamente inexistentes los trabajos académicos que han tratado de examinar de una manera más directa o indirecta la presencia de las universidades en las redes sociales de carácter horizontal o genéricas.

Holmberg (2015) analizó la presencia de las universidades finlandesas en varias redes sociales (*Twitter, LinkedIn, YouTube, Facebook*) teniendo en cuenta únicamente el número de seguidores o suscriptores.

**Otto y Williams** (2014) hicieron lo propio para el caso de varias universidades africanas.

Por su parte, **Permatasari et al.** (2013) analizaron la penetración de las redes sociales horizontales en el sistema universitario de Indonesia, encontrando una gran variación entre instituciones de educación superior.

**Campos-Freire, Rivera-Rogel y Rodríguez-Hidalgo** (2014) estudiaron la presencia e impacto de las universidades andinas en *Facebook*, *Twitter*, *YouTube*, *Google+* y *ResearchGate*, analizando únicamente el número de seguidores en cada una de ellas.

**Paniagua-Rojano y Gómez-Calderón** (2012) realizaron un ranking de las 15 universidades españolas con más seguidores en *Facebook*, *Twitter*, *YouTube*, *LinkedIn*, *Tuenti* y *Flickr*.

Al margen de estos intentos más académicos de realizar una clasificación de las universidades en RS, dada la proliferación de rankings de universidades a nivel nacional e internacional y la repercusión que tienen las RS a la hora de considerar el prestigio de las instituciones de educación superior, diversas organizaciones se han lanzado también a diseñar rankings de universidades en RS con un objetivo más centrado en obtener tráfico, notoriedad y reputación online.

En la esfera internacional podemos encontrar algunos rankings de universidades en RS, si bien sólo están centrados en universidades anglosajonas de EUA o Reino Unido. Uno de ellos es el realizado por *Colleagueatlas.org*, que analiza la presencia e influencia social de cerca de 400 instituciones de EUA en *Facebook*, *Twitter*, *Google+*, *LinkedIn* e *Instagram* con más de 30 parámetros de esas redes. Con esos datos, aplicando determinados algoritmos que no quedan nada claros en cuanto a la metodología seguida, obtienen diversos rankings: Top 20 y Top 30 de influencia y presencia social en dichas redes.

<http://www.colleagueatlas.org/top-colleges-in-social-media-main.html>

Otro intento reciente es el realizado por *The World University Rankings*, que centra la atención en la presencia de universidades americanas e inglesas en *Twitter*. El ranking se construye de manera simple considerando el número de seguidores, *Siguiendo*, la ratio *Seguidores/Siguiendo*, la ratio tweets/día y una medida de influencia social de la cuenta obtenida con una aplicación externa de *Twitter*. <https://www.timeshighereducation.com/news/top-100-most-influential-uk-and-us-universities-on-twitter/2013373.article>

Asimismo, nos encontramos con los rankings realizados por *Theunipod.com* para las universidades británicas en *Facebook*, *Twitter* y *YouTube*, si bien sólo consideran como criterio de clasificación el número de seguidores en cada una de estas redes. Además, la información es de 2013 y no ha sido actualizada.

<http://www.theunipod.com/making-your-choice/university-rankings/social-media-rankings>

Al igual que ocurre en el terreno internacional, los intentos llevados a cabo para elaborar rankings de universidades españolas en RS generales son escasos. A continuación se comentan brevemente.

*Alianzo*

<http://blogs.alianzo.com/socialnetworks>



Esta institución creó inicialmente sus propios rankings, que servían para medir la popularidad de blogs por categorías basándose en los datos públicos que ofrecían dichos blogs. Conforme la empresa fue creciendo, fue haciendo rankings más relevantes incluyendo RS como *Twitter* y *Facebook*. Recientemente realizó un cambio radical en su web. En la actualidad considera una gran cantidad de RS y ofrece la posibilidad de conocer la influencia combinada de una marca en el mundo 2.0. No obstante, no se especifican los criterios de valoración y ponderación que se aplican para calcular los rankings.

*Avanzaentucarrera.com*

<http://www.avanzaentucarrera.com/llegaraser/especiales/especial-carreras/las-universidades-en-las-redes-sociales>



El canal de formación *Infoempleo* ha realizado durante dos años seguidos un ranking de universidades españolas, tanto públicas como privadas, centrándose únicamente en dos redes sociales: *Facebook* y *Twitter*. El primero de estos rankings se publicó en un blog paralelo que mantiene esta web y apenas contó con repercusión mediática. El segundo, en cambio, sí que ha producido un fuerte impacto, publicándose en múltiples medios de comunicación, la mayoría de ellos online. Este ranking se basa únicamente en el número de seguidores (presencia) que cada universidad tiene en esas RS.

*Blog Juan Carlos Martínez*

<http://es.slideshare.net/juancarlosmtnez/informe-del-estado-de-las-redes-sociales-en-las-universidades-espaolas>

Es un ranking bastante completo en el que se consideran seis RS, si bien no existe transparencia en cuanto a la metodología utilizada por el autor. No se explica el procedimiento seguido ni se justifica por qué usa esos y no otros indicadores. Directamente define las fuentes de información y realiza distintos rankings. Los datos son de 2013.

*IY Magazine.es*

<http://www.iymagazine.es/noticia/1231/universidad/top-20-del-ranking-de-universidades-espanolas-en-redes-sociales.html>

Ranking realizado por la web *IYMagazine.es* en el que el único indicador considerado es el número de fans o seguidores en las redes sociales *Facebook*, *Twitter* y *Tuenti* (dimensión de presencia), al igual que el ranking de *Avanzaentucarrera.com*.

De la revisión realizada podemos concluir que al contrario de lo que ocurre con los rankings de universidades basados en indicadores de producción científica, los intentos por jerarquizar las universidades a través del uso de indicadores más sociales son mucho menos numerosos y en la mayoría de los casos surgen de iniciativas propias de webs o consultoras y no desde la propia academia. Esto lleva a que la metodología que se utiliza para la definición de los rankings no siempre sea del todo transparente y fiable. Además, se trata de rankings muy sesgados hacia determinados ámbitos geográficos y que se centran en muy pocas RS, en algunos casos en una sola.

### 1.4. Objetivos

El objetivo de este trabajo es proponer un marco general de definición de rankings que permita analizar el impacto de todas las universidades españolas, tanto públicas como privadas, en las principales RS genéricas (en concreto, *Twitter*, *Facebook*, *LinkedIn* y *YouTube*) según un amplio número de criterios. Dicho marco se basa en:

- definición de una taxonomía de criterios que los clasifica en varias dimensiones;
- selección de los disponibles en estas RS para cada categoría;
- construcción de distintos tipos de indicadores a partir de ellos, permitiendo seleccionar la dimensión de evaluación deseada por el usuario en cada caso.

Para ello, llevaremos a cabo una recogida de datos y un establecimiento de variables lo más exhaustivo y riguroso posible.

Para ilustrar el marco de trabajo introducido, presentaremos dos casos de estudio en los que aplicaremos cuatro rankings específicos relacionados con la presencia y la influencia de las universidades españolas en *Twitter* y *LinkedIn* empleando datos recogidos en los últimos diez días de agosto de 2015. Las limitaciones de espacio no nos permiten la inclusión de resultados del resto de rankings propuestos.

## 2. Metodología

Según la *Sexta oleada del Observatorio de redes sociales (The Cocktail Analysis, 2015)*, las tres RS horizontales más extendidas son *Facebook*, *Twitter* y *Google+*, por ese orden. En nuestro marco de trabajo hemos decidido seleccionar las dos primeras, dada la poca penetración de las universidades españolas en la última. Además, hemos considerado también dos RS especializadas de distinto tipo: *LinkedIn*

como red profesional y *YouTube* como repositorio audiovisual. Ambas son muy empleadas por las universidades españolas como canales de comunicación activos (**Paniagua-Rojano; Gómez-Calderón, 2012**).

Cada una de estas RS posee características distintivas y, por tanto, no es conveniente emplear los mismos indicadores en todas ellas sino hacer un diseño de rankings personalizado en cada caso. A continuación, describiremos los indicadores que se han valorado:

### Twitter

- *Seguidores*: usuarios con un perfil personal o profesional que agregan perfiles de otros usuarios. En este caso, consideramos el número absoluto de seguidores de cada universidad.
- *Ratio Siguiendo/Seguidores*: proporción entre los usuarios a los que sigue una institución y los que la siguen a ella. Se trata de ver si existe un equilibrio entre ambos datos, ya que si una institución es muy seguida y por el contrario no sigue prácticamente a nadie, da lugar a una sensación de lejanía y de falta de diálogo. Cuanto más cerca de 1 esté este valor, más equilibrada será dicha proporción.
- *Tweets*: número de tweets generados por la cuenta de una universidad desde su creación. Mide la actividad de la universidad en *Twitter*. Cuanto más elevado sea, mayor será la intención de convicción por parte de la institución.
- *Alcance estimado (estimated reach)*: alcance global estimado de los tweets de una universidad en un período de tiempo concreto, es decir, el número de cuentas únicas de *Twitter* que reciben los tweets.
- *Exposición (exposure)*: número de impresiones generadas, es decir, de tweets entregados en *timelines*, incluidas repeticiones en un período de tiempo concreto.

### Facebook

- *Likes*: número de usuarios que al pulsar el botón “Me gusta” se convierten en observadores activos de la web de la universidad a la que han decidido seguir.
- *Internacionalización*: refleja la proyección internacional que tiene cada institución. Se mide como el porcentaje de seguidores (*likes*) no españoles con respecto al total.

### LinkedIn

- *Seguidores*: número de usuarios de la red social que deciden conocer cada uno de los movimientos que hace la institución dentro de la red social.

Tabla 1. Taxonomía de indicadores

Indicadores específicos en cada red social				
Tipo de indicador	Twitter	Facebook	LinkedIn	YouTube
Presencia en la red	Seguidores	Likes	Seguidores	Suscriptores
Modelo de comunicación relacional	Siguiendo/ Seguidores	--	--	--
Modelo de comunicación tradicional	Tweets	--	--	Videos
Influencia	Exposición y Alcance estimado	--	Interesados en asistir y Egresados	Reproducciones
Internacionalización	--	Internacionalización	--	--

- Empleados: número de empleados que siguen el perfil de la institución desde su perfil profesional de *LinkedIn*. El dato se recoge de los propios usuarios cuando especifican su puesto de trabajo y coincide con una de las instituciones presentes en la red social.
- *Interesados en asistir*: número de personas que han añadido la universidad a su tablón de decisiones.
- *Egresados*: número de antiguos alumnos que siguen a la universidad.

### YouTube

- *Suscriptores*: número de usuarios que deciden seguir al canal de la institución para recibir notificaciones de novedades.
- *Vídeos*: cantidad de material multimedia que la institución ha publicado en su canal dentro de la red social.
- *Reproducciones*: número de veces que se ha visualizado el contenido del canal.

De este modo, podemos clasificar los indicadores seleccionados en función de su tipo, generalizando los criterios a considerar en la batería de rankings a definir, tal y como se muestra en la tabla 1.

Dado que las características de cada red social son muy distintas, no parece razonable agrupar todas las variables consideradas en un solo ranking ya que se combinarían criterios heterogéneos sin nada en común. De hecho, estos agrupamientos son complicados incluso dentro de la misma red como es el caso de *Twitter* o *Facebook* dada la diversidad de los indicadores existentes. Por esta razón, hemos optado finalmente por desvincular las distintas RS y considerarlas por separado. Al disponer de una tipología de criterios de distinta naturaleza para cada una de ellas, nuestro marco general de trabajo permite diseñar rankings específicos para medir la importancia de cada universidad en la dimensión concreta que queramos evaluar, tomando variables individuales o combinando indicadores del mismo tipo para obtener un ranking más robusto.

En 2015 las redes más extendidas por número de perfiles de universidades españolas eran *Facebook*, *Twitter* y *LinkedIn*

A continuación, presentamos varias propuestas de rankings para cada una de las RS consideradas en una o varias de las cinco dimensiones definidas.

### Twitter

- *Presencia en la red*: sólo se considera la variable "*Seguidores*". Se trata de realizar un ranking muy básico y simultáneamente significativo pues esta variable muestra la popularidad que posee una institución en términos absolutos.
- *Modelo de comunicación relacional* (o de diálogo): basado en el marketing relacional donde la prioridad es la comunicación y el diálogo, es decir, que exista una complicidad y acercamiento entre emisores y receptores (Duncan; Moriarty, 1998; Bortree; Seltzer, 2009; Rybalko; Seltzer, 2010). La variable "*Ratio Siguiendo/Seguidores*" es la uti-

lizada en este caso para conocer el grado de diálogo y conexión que existe entre los *Seguidores* de la universidad y la propia institución. Para definir un indicador que mida de forma adecuada la dispersión del valor obtenido en la variable con respecto a su valor óptimo de 1, proponemos emplear la siguiente expresión matemática:

$$MCR = 1 - \left| 1 - \frac{\text{Siguiendo}}{\text{Seguidores}} \right|$$

El indicador está basado en una medida no lineal de distancia y se define en (0,1], siendo 1 su valor óptimo. Aquellas universidades con una ratio adecuada entre *Siguiendo* y *Seguidores* toman valor 1 en el indicador mientras que las que lo superen por arriba o no lo alcancen por debajo reciben un valor que tiende a 0 cuanto más desbalanceada sea la ratio.

- *Modelo de comunicación tradicional* (o de persuasión): basado en el marketing tradicional en el que la emisión masiva de mensajes unidireccionales para persuadir a los grupos de interés es el único motor de comunicación (Petty; Cacioppo, 1986; 2011). Para realizar el ranking en este modelo se ha utilizado la variable "*Tweets*" ya que proporciona una medida directa de insistencia por parte de las instituciones.
- *Indicador global de influencia en la red*: pretende medir el alcance global de la influencia que tiene el perfil de la institución dentro de la red social, propuesto en el trabajo actual. Se propone combinar dos variables de la misma dimensión para construir el indicador correspondiente: "*Exposición*" y "*Alcance estimado*". Para ello se usa un operador de agregación basado en la raíz cuadrada del producto de ambas (una vez normalizadas):

$$IGI - \text{Twitter} = \sqrt{\text{Exposición} \cdot \text{Alcance estimado}}$$

Dentro de la gran variedad de operadores de agregación existentes, se ha escogido el actual porque proporciona la misma importancia a los dos criterios considerados, como pretendíamos, además de ser más restrictivo que la media al exigir valores altos en ambos para obtener un valor alto en el indicador.

### Facebook

- *Presencia en la red*: ranking realizado con la variable "*Likes*", basado en datos absolutos. Es equivalente al del mismo nombre en *Twitter* al presentar ambos una intención de seguimiento de las instituciones en la red social.
- *Internacionalización*: se trata de evaluar la proyección más internacional de las universidades. Para este ranking sólo se ha considerado una variable, "*Internacionalización*".

### LinkedIn

- *Presencia en la red*: al igual que en el ranking de *Twitter*, se considera directamente la variable "*Seguidores*".
- *Indicador global de influencia en la red*: ranking que pretende medir el alcance global de la influencia que tiene

el perfil de la institución dentro de la red social, también propuesto por los autores del presente trabajo. Se propone combinar dos variables para construir el indicador correspondiente: “Interesados en asistir” y “Egresados”. En este caso hemos decidido considerar una combinación lineal como operador de agregación ya que pensamos que las dos métricas no deberían tener el mismo grado de importancia. En nuestra opinión, los *Interesados* tie-

nen un valor añadido al ser los que toman la iniciativa de buscar a la institución y deciden agregarla a su tablón de decisiones de *LinkedIn*. Por tanto, nuestro criterio fue darle un peso de 0,75 a la variable “Interesados” y de 0,25 a “Egresados” (después de normalizar ambas). El indicador toma la expresión:

$$IGI - LinkedIn = 0,75 \cdot Interesados + 0,25 \cdot Egresados$$

Tabla 2. Cuentas oficiales de universidades españolas en *Twitter* consideradas en el presente estudio:

	Universidad	Username
<b>IEU</b>	IE Universidad	@leuniversity
<b>UMU</b>	Mondragón Unibersitatea	@Munibersitatea
<b>UDIMA</b>	Universidad a Distancia de Madrid	@UDIMA
<b>UAX</b>	Universidad Alfonso X El Sabio	@universidad_uax
<b>UAN</b>	Universidad Antonio de Nebrija	@Nebrija
<b>UAM</b>	Universidad Autónoma de Madrid	@UAM_Madrid
<b>UCJC</b>	Universidad Camilo José Cela	@universidadcjc
<b>UCH-CEU</b>	Universidad Cardenal Herrera-CEU	@uchceu
<b>UC3M</b>	Universidad Carlos III de Madrid	@uc3m
<b>UCV</b>	Universidad Católica de Valencia San Vicente Mártir	@UCV_es
<b>UCAM</b>	Universidad Católica San Antonio	@UCAM
<b>UCST</b>	Universidad Católica Santa Teresa de Jesús de Avila	@UCAVILA
<b>UCM</b>	Universidad Complutense de Madrid	@unicomplutense
<b>UAH</b>	Universidad de Alcalá de Henares	@UAHes
<b>UA</b>	Universidad de Alicante	@UA_Universidad
<b>UAL</b>	Universidad de Almería	@ualmeria
<b>UBU</b>	Universidad de Burgos	@UBUEstudiantes
<b>UCA</b>	Universidad de Cádiz	@univcadiz
<b>UNICAN</b>	Universidad de Cantabria	@unican
<b>UCLM</b>	Universidad de Castilla-La Mancha	@uclm_es
<b>UCO</b>	Universidad de Córdoba	@Univcordoba
<b>UDE</b>	Universidad de Deusto	@deusto
<b>UEX</b>	Universidad de Extremadura	@infoeux
<b>UGR</b>	Universidad de Granada	@CanalUGR
<b>UHU</b>	Universidad de Huelva	@CanalUHU
<b>UJA</b>	Universidad de Jaén	@ujaen
<b>ULL</b>	Universidad de La Laguna	@CanalULL
<b>UR</b>	Universidad de La Rioja	@Unirioja
<b>ULPGC</b>	Universidad de Las Palmas de Gran Canaria	@ULPGC
<b>ULE</b>	Universidad de León	@unileon
<b>UMA</b>	Universidad de Málaga	@InfoUMA
<b>UM</b>	Universidad de Murcia	@umnoticias
<b>UNA</b>	Universidad de Navarra	@unav
<b>Uniovi</b>	Universidad de Oviedo	@uniovi_info
<b>USAL</b>	Universidad de Salamanca	@usal
<b>USC</b>	Universidad de Santiago de Compostela	@UniversidadeUSC
<b>USE</b>	Universidad de Sevilla	@unisevilla
<b>UVA</b>	Universidad de Valladolid	@UVa_es
<b>UVIGO</b>	Universidad de Vigo	@uvigo

<b>UZ</b>	Universidad de Zaragoza	@unizar
<b>UPV-EHU</b>	Universidad del País Vasco/Euskal Herriko Unibersitatea	@upvehu
<b>UEM</b>	Universidad Europea de Madrid	@UEuropea
<b>UEMC</b>	Universidad Europea Miguel de Cervantes	@UEMC
<b>UFV</b>	Universidad Francisco de Vitoria	@ufvmadrid
<b>UNIA</b>	Universidad Internacional de Andalucía	@UNIAuniversidad
<b>UIC</b>	Universidad Internacional de Catalunya	@UICbarcelona
<b>UNIR</b>	Universidad Internacional de La Rioja	@UNIRuniversidad
<b>UIIC</b>	Universidad Internacional Isabel I de Castilla	@ui1Universidad
<b>UIMP</b>	Universidad Internacional Menéndez Pelayo	@UIMP
<b>UIV</b>	Universidad Internacional Valenciana	@UniversidadVIU
<b>UJI</b>	Universidad Jaume I de Castellón	@UJI_noticies
<b>ULA</b>	Universidad Loyola Andalucía	@LoyolaAnd
<b>UMH</b>	Universidad Miguel Hernández de Elche	@UniversidadMH
<b>UNED</b>	Universidad Nacional de Educación a Distancia	@UNED
<b>UPO</b>	Universidad Pablo de Olavide	@pablodeolavide
<b>UPCA</b>	Universidad Politécnica de Cartagena	@UPCTnoticias
<b>UPM</b>	Universidad Politécnica de Madrid	@La_UPM
<b>UPCO</b>	Universidad Pontificia de Comillas	@UCOMILLAS
<b>UPSAL</b>	Universidad Pontificia de Salamanca	@upsa_salamanca
<b>UPN</b>	Universidad Pública de Navarra	@Unavarra
<b>URJC</b>	Universidad Rey Juan Carlos	@urjc
<b>USJ</b>	Universidad San Jorge	@_usj_
<b>USP-CEU</b>	Universidad San Pablo CEU	@USPCEU
<b>UDC</b>	Universidade da Coruña	@UDC_gal
<b>UAO-CEU</b>	Universitat Abat Oliba CEU	@UAOCEU
<b>UAB</b>	Universitat Autònoma de Barcelona	@UAB_info
<b>UB</b>	Universitat de Barcelona	@UB_endirecte
<b>UGI</b>	Universitat de Girona	@univgirona
<b>UIB</b>	Universitat de les Illes Balears	@UIBuniversitat
<b>UL</b>	Universitat de Lleida	@trampoli_udl
<b>UV</b>	Universitat de València-Estudi General	@UV_EG
<b>UVIC</b>	Universitat de Vic	@uvic_news
<b>UOC</b>	Universitat Oberta de Catalunya	@UOCuniversitat
<b>UPC</b>	Universitat Politècnica de Catalunya	@BarcelonaTech
<b>UPV</b>	Universitat Politècnica de València	@upv
<b>UPF</b>	Universitat Pompeu Fabra	@univpompeufabra
<b>URL</b>	Universitat Ramon Llull	@uramonllull
<b>URV</b>	Universitat Rovira i Virgili	@universitatURV

## YouTube

- *Presencia en la red*: se considera únicamente la variable “Suscriptores”, que representa el mismo concepto que las consideradas en las RS anteriores para esta dimensión.
- *Modelo de comunicación tradicional*: se utiliza la variable “Videos” como medida directa de insistencia por parte de las instituciones.
- *Influencia*: Se emplea la variable “Reproducciones” para este ranking.

## 3. Resultados

Con objeto de ilustrar la metodología presentada, vamos a evaluar los resultados de las universidades españolas en algunos de nuestros rankings. En concreto, hemos seleccionado dos RS de distinta naturaleza: *Twitter* (una red horizontal masiva) y *LinkedIn* (una red vertical dado su perfil más profesional y orientado); y dos rankings específicos que miden las mismas dimensiones para cada una de ellas, *Presencia* e *Influencia*.

Se ha trabajado con prácticamente toda la población de universidades españolas (78), tanto públicas como privadas (tabla 2). Tal y como se ha indicado a lo largo de este trabajo, nos pareció importante que se consideraran relevantes todas las universidades españolas en igualdad de condiciones para la definición de unos rankings verdaderamente útiles. Sólo se ha dejado fuera la *Universidad Europea del Atlántico* ya que, dada su reciente creación, presenta todavía una escasa presencia en RS.

El estudio se ha basado en la recogida de datos de varias fuentes. Los datos más significativos y numerosos se han extraído de las propias RS, aunque para *Twitter* se han empleado también las dos métricas que definen el ranking de influencia obtenidas de *Tweetreach.com*. Todas las cuentas de las dos RS consideradas en este estudio son las oficiales de las propias universidades recogidas en sus webs oficiales. En el caso de *LinkedIn* la cuenta de cada universidad se corresponde con el propio nombre de la misma. En el caso de *Twitter*, el *username* de cada universidad está incluido en la tabla 2.

La recogida de información se realizó entre los días 20 y 30 de agosto de 2015. Uno de los *handicaps* más importantes fue la falta de normalización del nombre de las universidades en cada una de las RS, lo que requirió un importante trabajo adicional.

Una vez establecidas las variables con toda su información, se procedió a la realización de los cálculos pertinentes para obtener los valores de los indicadores y a la ordenación de las instituciones con respecto a dichos valores en cada ranking. Además, se analizó la clasificación de las universidades en cuatro de las combinaciones por pares de los rankings definidos para proporcionar una visión más general de su posicionamiento.

### 3.1. Rankings de universidades en Twitter

#### 3.1.1. Ranking de Presencia en Twitter

Los resultados de este primer ranking están recogidos en la tabla 3. Según el indicador de presencia en *Twitter*, la *Universidad de Granada* estaría a la cabeza con casi 90.000 se-

guidores. El segundo y tercer lugar lo ocuparían otras dos universidades públicas, *UNED* y *Sevilla*, respectivamente. Además, podemos ver que el top ten de las universidades de esta clasificación es de carácter público, salvo por la presencia de la *Universidad Europea de Madrid*.

#### 3.1.2. Ranking de Influencia global en Twitter

El segundo de los rankings definidos sobre *Twitter* está relacionado con el grado de *Influencia global en la red*

Tabla 3. Ranking de Presencia en Twitter

Posición	Univ.	Seguidores	Posición	Univ.	Seguidores
1	UGR	89.471	40	UMH	14.252
2	UNED	81.381	41	UCAM	14.218
3	US	79.582	42	UIIMP	13.046
4	UMA	59.642	43	UZ	12.892
5	UM	51.054	44	UIC	12.871
6	UCM	43.701	45	UCJC	11.449
7	UA	43.299	46	UR	11.275
8	UV	37.640	47	UPN	11.159
9	UAM	35.512	48	UAN	10.523
10	UEM	34.540	49	UNIR	10.045
11	URJC	32.615	50	UJI	9.658
12	UJ	32.604	51	UNICAN	9.202
13	USAL	27.520	52	UIB	8.939
14	UAH	26.886	53	URV	8.592
15	UPV	26.596	54	UCST	8.111
16	UCA	26.363	55	UBU	7.792
17	UVA	26.259	56	UCH-CEU	7.663
18	UNA	25.995	57	UNIA	7.404
19	UOC	25.860	58	UDIMA	6.960
20	UCO	25.844	59	UFV	6.937
21	ULPGC	25.815	60	USP-CEU	6.758
22	UB	25.777	61	UVIC	6.644
23	ULL	24.102	62	IEU	6.467
24	UPM	23.724	63	UCV	6.390
25	UVIGO	22.607	64	UNIOVI	6.259
26	UC3M	22.197	65	URL	5.988
27	UGI	21.467	66	UMU	5.606
28	UPO	21.149	67	ULA	5.223
29	UAB	21.084	68	UEMC	4.997
30	UDC	20.464	69	UAX	4.871
31	UCLM	20.285	70	UIV	4.659
32	UHU	19.294	71	UPCO	4.484
33	UAL	18.646	72	USC	4.143
34	UPC	17.518	73	UAO-CEU	3.380
35	UDE	17.380	74	UPCA	3.037
36	ULE	17.242	75	UIIC	2.684
37	UPF	17.052	76	UPSAL	2.065
38	UEX	14.710	77	USJ	1.538
39	UPV-EHU	14.388	78	ULL	680

(IGR) que tienen las universidades en esta red social (tabla 4). Recordamos que este ranking se define a partir de la combinación de los indicadores “Exposición” y “Alcance estimado”. En esta ocasión es la UC3M la que ocupa el primer lugar, seguida muy de cerca la UIIC y la UFV. Merece la pena destacar que existe cierto equilibrio entre las universidades públicas y privadas en el top ten: seis públicas y cuatro privadas. Dado el menor número de universidades privadas y comparando con los resultados del resto de rankings presentados, parece ser que este

tipo de universidades realizan una apuesta por un modelo de comunicación basado en esta red social.

### 3.2. Rankings de universidades en LinkedIn

#### 3.2.1. Ranking de Presencia en LinkedIn

La UCM es la que lidera la clasificación de forma destacada con casi 250.000 seguidores, seguida de la UB y la UPM (tabla 5). Cabe destacar que las diez primeras posiciones de esta clasificación están ocupadas únicamente por universidades públicas.

Tabla 4. Ranking de *Influencia global* en Twitter

Posición	Univ.	IGR	Posición	Univ.	IGR
1	UC3M	0,9879	40	UPSAL	0,2813
2	UIIC	0,9642	41	UMU	0,2811
3	UFV	0,8227	42	UV	0,2779
4	USE	0,7424	43	UCH-CEU	0,2620
5	UDC	0,7390	44	UHU	0,2611
6	UCLM	0,5700	45	USP-CEU	0,2607
7	UIMP	0,5448	46	UAO-CEU	0,2599
8	UM	0,5291	47	UPF	0,2521
9	UPC	0,4968	48	UNIA	0,2418
10	UDE	0,4816	49	UJI	0,2413
11	UPN	0,4728	50	IEU	0,2407
12	UNA	0,4560	51	UCV	0,2394
13	UB	0,4558	52	UGI	0,2394
14	UNIR	0,4518	53	UCJC	0,2346
15	UAM	0,4471	54	UNICAN	0,2245
16	UPV	0,4336	55	UAH	0,2211
17	UEM	0,4330	56	ULE	0,2146
18	URV	0,4290	57	UPV-EHU	0,2144
19	URL	0,4279	58	UNED	0,2121
20	UGR	0,4204	59	UEX	0,2032
21	USAL	0,4181	60	UMA	0,2018
22	UCM	0,3702	61	UAL	0,1939
23	UPCA	0,3682	62	USC	0,1828
24	URJC	0,3674	63	UBU	0,1673
25	UZ	0,3551	64	ULA	0,1664
26	UJA	0,3539	65	UDIMA	0,1637
27	ULPGC	0,3454	66	UCAM	0,1625
28	UPM	0,3437	67	UEMC	0,1518
29	UAB	0,3434	68	ULL	0,1415
30	UR	0,3363	69	UA	0,1392
31	UIC	0,3270	70	UNIOVI	0,1385
32	UCA	0,3139	71	UVIGO	0,1365
33	UPCO	0,3046	72	UAX	0,1337
34	UIB	0,3039	73	UPO	0,1201
35	UVIC	0,3035	74	UMH	0,1119
36	UVA	0,3010	75	USJ	0,1030
37	UOC	0,2871	76	UIV	0,0907
38	UCST	0,2858	77	UAN	0,0732
39	UCO	0,2821	78	UL	0,0145

Tabla 5. Ranking de *Presencia* en LinkedIn

Posición	Univ	Seguidores	Posición	Univ	Seguidores
1	UCM	249.297	40	URV	18.082
2	UB	164.182	41	UNICAN	17.879
3	UPM	116.901	42	UPO	17.801
4	UAB	93.198	43	UPSAL	17.399
5	USE	91.316	44	ULE	17.334
6	UPC	86.424	45	UCO	17.207
7	UAM	78.508	46	UCA	16.847
8	UGR	74.702	47	UAN	16.516
9	UPV	62.219	48	UCJC	15.916
10	USAL	58.722	49	UAX	15.301
11	IEU	55.203	50	UMH	15.016
12	UPF	54.675	51	UIB	13.082
13	UV	54.257	52	UFV	12.648
14	UC3M	54.211	53	UJA	12.527
15	UOC	53.009	54	UHU	11.829
16	UNED	51.399	55	UPN	11.292
17	UZ	49.924	56	UL	11.251
18	URJC	44.492	57	UAL	11.004
19	UPV-EHU	43.467	58	UBU	10.529
20	USC	42.985	59	UIC	8.570
21	UA	41.770	60	UR	8.274
22	UNIOVI	40.308	61	UVIC	7.779
23	UVA	40.291	62	UPCA	7.716
24	UAH	39.296	63	UCAM	7.682
25	UMA	38.988	64	UMU	7.597
26	USP-CEU	36.305	65	UCV	6.668
27	UNA	35.816	66	UEMC	6.029
28	UDE	35.531	67	UAO-CEU	5.294
29	UPCO	33.115	68	UIMP	4.458
30	UM	31.659	69	UNIR	4.360
31	UVIGO	30.584	70	USJ	3.636
32	UEM	30.093	71	UNIA	2.828
33	URL	27.359	72	UDIMA	2.760
34	UCLM	25.866	73	UJI	2.503
35	UEX	22.938	74	UCST	2.446
36	ULL	21.895	75	UCH-CEU	1.103
37	UDC	21.110	76	ULA	966
38	ULPGC	20.320	77	UIV	740
39	UGI	19.329	78	UIIC	422

#### 4.2.2. Ranking de Influencia global en LinkedIn

En la clasificación de *Influencia global*, en la que consideraban las variables “*Interesados en asistir*” y “*Egresados*”, la *IEU* alcanza el primer puesto con un valor de 0,7992 (tabla 6). Le siguen a corta distancia la *UCM* y la *UB*. De nuevo, como en el ranking anterior, predominan las universidades públicas en el top ten, siendo la única de carácter privado la que ocupa el primer lugar.

Tabla 6. Ranking de *Influencia global* en LinkedIn

Posición	Univ	Influencia	Posición	Univ	Influencia
1	IEU	0,7992	40	UNIR	0,0647
2	UCM	0,7762	41	UCJC	0,0641
3	UB	0,6542	42	ULPGC	0,0640
4	UAB	0,4425	43	URV	0,0639
5	UAM	0,4287	44	UPO	0,0636
6	UPM	0,3965	45	UIB	0,0565
7	UGR	0,3419	46	ULE	0,0556
8	UPF	0,3371	47	UAN	0,0544
9	UPC	0,3201	48	UPSAL	0,0529
10	USE	0,3175	49	UFV	0,0528
11	UOC	0,2958	50	UCO	0,0528
12	UV	0,2871	51	UNICAN	0,0525
13	UNED	0,2747	52	UJA	0,0520
14	UPV	0,2726	53	UCAM	0,0493
15	USAL	0,2577	54	UEX	0,0482
16	UC3M	0,2478	55	UMH	0,0468
17	URJC	0,1785	56	UJI	0,0460
18	UAH	0,1771	57	UAX	0,0438
19	UEM	0,1628	58	UEMC	0,0437
20	UPV-EHU	0,1618	59	UMU	0,0436
21	UMA	0,1535	60	UAL	0,0417
22	UA	0,1496	61	UCH-CEU	0,0399
23	USC	0,1471	62	UCV	0,0394
24	UPCO	0,1401	63	UVIC	0,0391
25	UNA	0,1392	64	UPN	0,0386
26	UZ	0,1369	65	UIC	0,0372
27	UDE	0,1295	66	UBU	0,0334
28	UVA	0,1080	67	UHU	0,0300
29	UNIOVI	0,1069	68	UL	0,0295
30	UM	0,0933	69	UR	0,0294
31	USP-CEU	0,0910	70	UNIA	0,0241
32	UDIMA	0,0909	71	UPCA	0,0241
33	UVIGO	0,0866	72	ULA	0,0239
34	UCLM	0,0745	73	UIV	0,0217
35	UDC	0,0732	74	UAB-CEU	0,0203
36	ULL	0,0716	75	UIMP	0,0186
37	URL	0,0692	76	USJ	0,0174
38	UGI	0,0652	77	UCST	0,0148
39	UCA	0,0652	78	UIIC	0,0007

#### 4.3. Estudio conjunto de los resultados de los rankings

En esta sección pretendemos profundizar aún más en el posicionamiento de las universidades analizando algunos resultados combinados de los rankings que hemos definido. Para ello vamos a realizar unos gráficos de dispersión que representan la posición conjunta de cada universidad en dos rankings distintos. Por un lado, analizaremos conjuntamente los dos rankings definidos para cada red social mientras que por otro relacionaremos las dos dimensiones analizadas, presencia e influencia, en ambas RS. En los cuatro casos, realizaremos un análisis estadístico para determinar el grado de correlación lineal existente entre los dos indicadores considerados, empleando para ello el coeficiente *r* de Pearson.

##### 4.3.1. Estudio conjunto de los rankings de Presencia e Influencia en Twitter

En este primer caso, el valor del coeficiente *r* obtenido es de 0,198, lo que indica que existe una correlación muy poco significativa entre los dos indicadores considerados. De este modo, queda claro el hecho de que una presencia alta en *Twitter* no implica directamente una influencia importante en esta RS y viceversa.

Como se observa en la figura 1, la *Universidad de Sevilla* es la que tiene un equilibrio más destacable entre influencia y presencia en *Twitter*, siendo la tercera en presencia y la cuarta en influencia. Vemos que la primera y la segunda en presencia, *Granada* y la *UNED*, ocupan posiciones muy lejanas a la primera en *Influencia*, la *Universidad Carlos III de Madrid*. Esto puede deberse a que, a pesar de tener una fuerte actividad en la red y un gran volumen de seguidores, el alcance obtenido no se corresponde con su posición en términos de presencia.

La razón de ello puede estar en la “calidad” de los seguidores de esas universidades dado que la influencia depende del número de seguidores que tienen los seguidores de las universidades y de la actividad que estos ejerzan en la red social. *Isabel I de Castilla* y *Francisco de Vitoria* son universidades que, a pesar de tener una presencia muy escasa en *Twitter*, presentan una potente influencia ocupando el segundo y tercer lugar respectivamente. El caso de la propia *Carlos III* y la de *A Coruña* es similar, aunque en este caso la presencia es bastante más destacada.

##### 4.3.2. Estudio conjunto de los rankings de Presencia e Influencia en LinkedIn

Las dos variables presentan una correlación muy significativa en la que apenas se aprecian elementos demasiado desmarcados (figura 2). El valor del coeficiente *r* de Pearson obtenido en el análisis realizado es de 0,877, señalando una correlación positiva bastante alta. Podemos concluir, por tanto, que al contrario que en el caso de *Twitter*, en la red social *LinkedIn* la presencia y la influencia de las universidades españolas en la red están fuertemente relacionadas.

La *Universidad Complutense de Madrid* y la *Universitat de Barcelona* son dos claros ejemplos de equilibrio entre presencia e influencia en *LinkedIn*. Ambas universidades cuentan con un nutrido número de seguidores y además son instituciones consideradas muy positivamente por éstos. Cabe

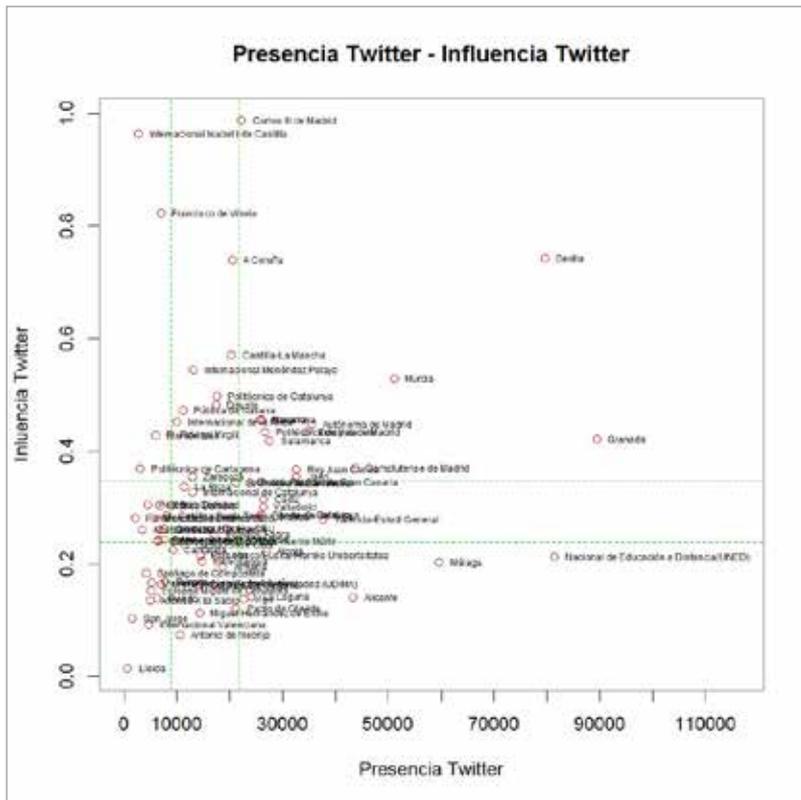


Figura 1. Gráfico de dispersión de los rankings de presencia e influencia en *Twitter*

destacar el caso de la *IE Universidad* que cuenta con una importante influencia en *LinkedIn* con una presencia poco relevante. Esto puede deberse a que es una universidad que tiene mucho prestigio y que, por tanto, cuenta con un alto número de personas que tienen la intención de asistir. En cambio, su actividad en la red es básicamente testimonial.

**4.3.3. Estudio conjunto de los rankings de Presencia en *Twitter* y *LinkedIn***

En este caso, el coeficiente de Pearson toma un valor de 0,47, lo que muestra una correlación positiva bastante moderada. Se puede concluir que la apuesta por la presencia conjunta de las universidades españolas en estas dos RS no es excesivamente importante.

La *Complutense de Madrid* es la que tiene la presencia combinada más destacable, siendo la primera en *LinkedIn* y la sexta en *Twitter* (véase la figura 3). La primera en *Twitter*, *Granada*, es la octava en *LinkedIn*, presentando también un buen equilibrio, al igual que *Sevilla* (tercera en *Twitter* y quinta en *LinkedIn*). El resto de universidades destacadas en alguna de las dos RS, no se encuentran en una posición tan señalada en la otra, como es el caso de *Barcelona*, *Politécnica de Madrid*, *Autònoma de Barcelona*, *Politécnica de Catalunya* (con mayor presencia en *LinkedIn*) y *UNED*, *Málaga*, *Murcia* y *Alicante* (más destacadas en *Twitter*).

**4.3.4. Estudio conjunto de los rankings de Influencia en *Twitter* y *LinkedIn***

En este caso no se aprecia una correlación significativa entre las influencias de las RS *Twitter* y *LinkedIn* (coeficiente  $r$ : 0,167). El grueso de las universidades que destacan en una de las dos, tiene una influencia inferior al 0,5 en la otra. Este hecho puede deberse a una posible focalización del esfuerzo de comunicación en una sola de estas dos RS por parte de las universidades españolas.

En los extremos se sitúan por un lado *IE Universidad* con una fuerte influencia en *LinkedIn*, la más importante, y una muy baja en *Twitter*. Por otro lado, la *Internacional Isabel I de Castilla*, situada en segunda posición en cuanto a influencia en *Twitter* y en cambio prácticamente inexistente en *LinkedIn*. En lo que respecta a universidades con un mejor equilibrio, podemos destacar a la *Carlos III*, primera en influencia en *Twitter* y decimosexta en *LinkedIn*; a la de *Sevilla*, cuarta en *Twitter* y décima en *LinkedIn*; a la de *Barcelona*, tercera en *LinkedIn* y decimotercera en *Twitter*; a la *Autónoma de Madrid*, quinta en *LinkedIn* y decimoquinta en *Twitter*; a la de *Granada*, séptima en *LinkedIn* y vigésima en *Twitter*, y a la *Politécnica de Catalunya*, novena en ambos rankings.

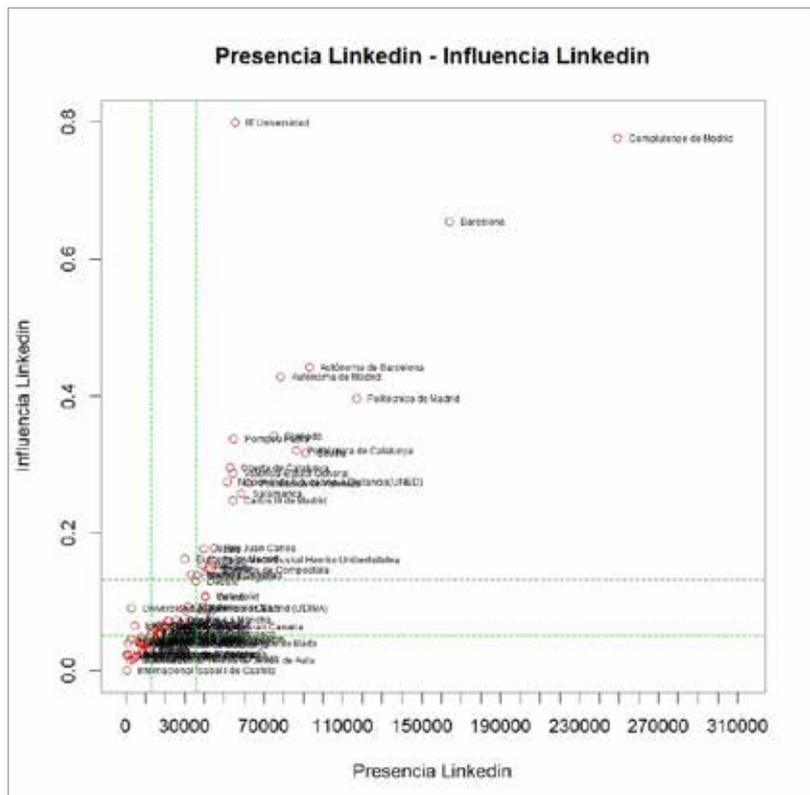


Figura 2. Gráfico de dispersión de los rankings de presencia e influencia en *LinkedIn*

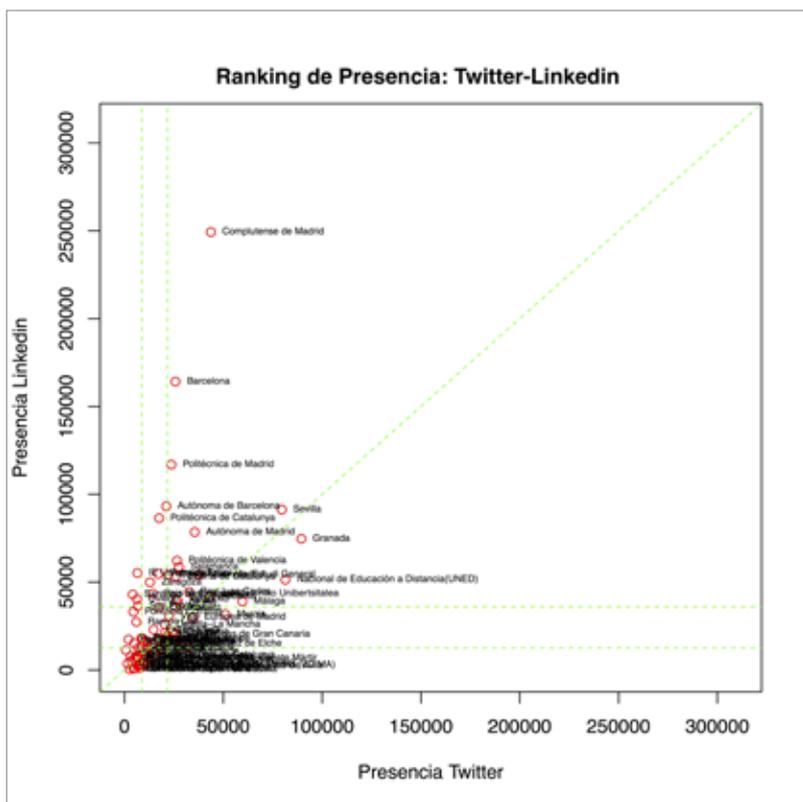


Figura 3. Gráfico de dispersión de los rankings de presencia en *Twitter* y *LinkedIn*

das las universidades españolas en todos los rankings realizados (siempre que, por supuesto, tengan presencia y existan datos de las mismas para la red social concreta), se explicitan claramente los indicadores empleados en cada caso, seleccionando aquellos más relevantes para el propósito de cada ranking de acuerdo con la naturaleza específica de la red social considerada, se señalan los mecanismos de agregación considerados para combinar los indicadores para la creación de cada ranking, y se realiza un análisis exhaustivo de los resultados obtenidos.

Del análisis de las cuentas institucionales de las universidades españolas a fecha del trabajo de campo (agosto de 2015), podemos concluir que las redes más extendidas por número de perfiles de universidades españolas en ese momento eran *Facebook*, *Twitter* y *LinkedIn* (todas ellas con 78 cuentas, resultando en un 93,9% de penetración), seguidas por *YouTube* (74 cuentas, 89,1%). Los indicadores seleccionados y clasificados en dimensiones de diferente naturaleza permiten diseñar una amplia batería de rankings para las cuatro RS consideradas. Se pueden proponer directa-

### 5. Conclusiones y líneas futuras de trabajo

En este trabajo hemos propuesto un marco general para la definición de rankings personalizados de medición del impacto de las universidades españolas en redes sociales (RS) con respecto a varias dimensiones. Usando este marco hemos construido rankings a partir de conjuntos de datos relevantes que permiten entender el funcionamiento de las universidades en RS a un nivel desconocido hasta ahora. La metodología seguida ha sido rigurosa y está alineada con las características que deben tener los rankings óptimos según la bibliografía (Merisotis, 2002).

Las propuestas presentadas superan las limitaciones de otros intentos realizados hasta la fecha que manifestaban notables carencias como la especificidad, al usar un número bajo de redes sociales (*Avanzaentucarrera.com*); la falta de cobertura al considerar un número limitado de universidades (varios de los rankings existentes); la falta de transparencia en la metodología de definición de los rankings (*Alianza, Blog Juan Carlos Martínez*), e incluso la definición de indicadores y la realización de una interpretación de datos excesivamente simples (*Avanzaentucarrera.com, IY Magazine.es*). En nuestro marco de trabajo se propone una metodología para cuatro de las RS más extendidas, se consideran to-

mente cuatro para *Twitter* (*Presencia en la red, Modelo de comunicación relacional, Modelo de comunicación tradicional e Influencia global en la red*), dos para *Facebook* (*Presencia e Internacionalización*), dos para *LinkedIn* (*Presencia*

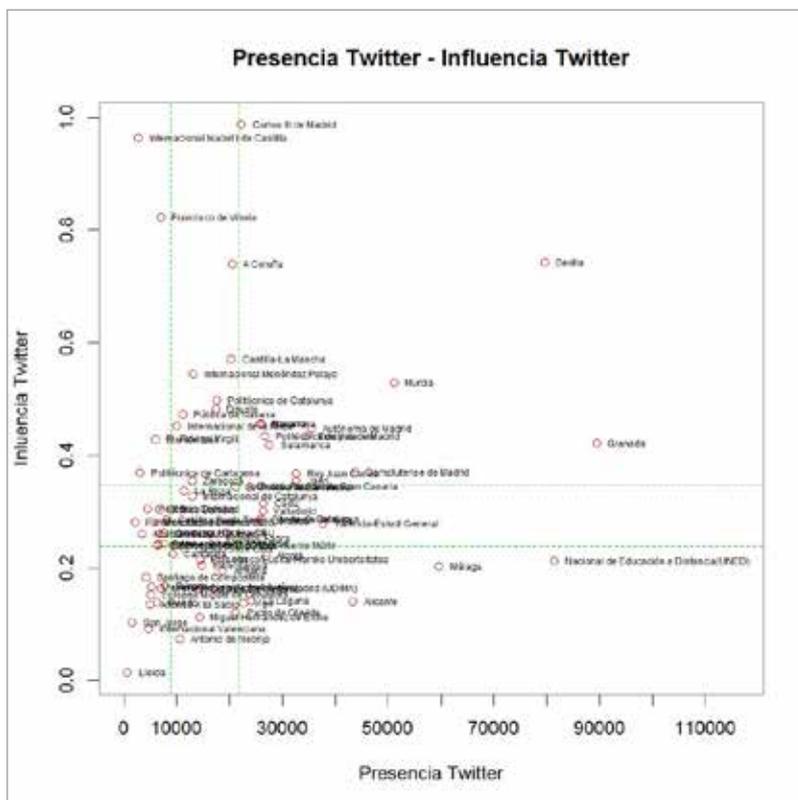


Figura 4. Gráfico de dispersión de los rankings de influencia en *Twitter* y *LinkedIn*

e *Influencia global*) y tres para YouTube (*Presencia, Modelo de comunicación relacional e Influencia*).

Por limitaciones de espacio, hemos desarrollado dos casos de estudio con datos actuales para dos de los rankings, *Presencia e Influencia*, en Twitter y LinkedIn. A la vista de los resultados obtenidos, hemos podido identificar que el ranking de *Influencia* de Twitter es el único que destaca a un grupo de universidades privadas en las primeras posiciones, aun teniendo una presencia menor que otras universidades. Los resultados asociados a la combinación de dos rankings distintos también han sido muy relevantes, mostrando por ejemplo como la *Presencia* y la *Influencia* presentan una mayor correlación en LinkedIn que en Twitter.

Este trabajo puede ayudar a los gestores universitarios a examinar su posición relativa en los distintos rankings de las RS en las que actúan y evaluar en qué medida están consiguiendo sus objetivos. En muchos casos la presencia de las universidades en RS no obedece a una estrategia de comunicación concreta e integrada con el resto de acciones de comunicación (Simón-Onieva, 2014), lo que sin duda afecta a la manera en la que actúan en las redes y a los resultados que obtienen. Son diversos los factores que pueden influir de manera positiva o negativa en el nivel de aceptación o *engagement* que reciben los mensajes que ponen las universidades en RS. Entre estos factores cabe destacar los formatos de contenidos utilizados, el momento de la publicación y la longitud de la misma (Cvijikj; Michahelles, 2013), a los que habría que unir el estilo de comunicación empleado centrado más en la persuasión o en el diálogo con los públicos objetivos.

La Universidad Complutense de Madrid y la Universitat de Barcelona son dos claros ejemplos de equilibrio entre presencia e influencia en LinkedIn

Como futuras líneas de trabajo nos planteamos realizar un estudio de correlación entre las posiciones que ocupan las universidades españolas en los rankings aquí presentados y su orden en otros rankings universitarios que miden otras dimensiones como los de investigación o los de docencia.

Por último, nos planteamos perfeccionar la definición de los rankings propuestos para considerar directamente el efecto del tamaño de la universidad en los valores de los indicadores. Como se indica en el informe del *Ranking Issue Fundación BBVA* (Pérez, 2013), el hecho de tener o no en cuenta el tamaño de las universidades a la hora de compararlas puede considerarse relevante y sería interesante analizar la relación entre el tamaño y la oferta formativa de las instituciones y su posición en nuestros rankings.

## Agradecimientos

Los autores desean agradecer la ayuda del profesor Luciano Sánchez, de la Universidad de Oviedo, en la generación de los gráficos de la sección 4.3.

Este trabajo está soportado por el Ministerio de Economía y Competitividad de España bajo el proyecto NEWSOCO (ref. TIN2015-67661-P, incluyendo Fondos Europeos de Desarro-

llo Regional FEDER) y por el Departamento de Innovación, Ciencia y Empresa de la Junta de Andalucía bajo el Proyecto de Excelencia P12-SEJ-2592.

## Notas

1. Existen países en los que estas redes no tienen tanta repercusión, como es el caso de China, que utiliza sus propias RS (*Sina Weibo, QZone, Renren, Youku*, etc.); Rusia, donde está muy extendida la red VK (antes *Vkontakte*); y la India, donde *Orkut* era mayoritaria hasta su desaparición en 2014
2. Según la Real Academia Española, se entiende por ranking una “clasificación de mayor a menor, útil para establecer criterios de valoración”.
3. También se ha presentado recientemente un estudio de presencia de las universidades españolas en otro tipo de RS más académicas y profesionales como *ResearchGate* y *Academia.edu* en González-Díaz, Iglesias-García y Codina (2015).

Una Presencia alta en Twitter no implica directamente una Influencia importante en esta red social y viceversa

## Bibliografía

- AIMC (2015). *Estudio general de medios, 3ª ola 2015* (abril-mayo). <http://www.aimc.es/-Audiencia-de-Internet-en-el-EGM-.html>
- Bortree, Denise S.; Seltzer, Trent (2009). “Dialogic strategies and outcomes: An analysis of environmental advocacy groups’ Facebook profiles”. *Public relations review*, v. 35, n. 3, pp. 317-319. <http://dx.doi.org/10.1016/j.pubrev.2009.05.002>
- Buela-Casal, Gualberto; Bermúdez, María-de-la-Paz; Sierra, Juan-Carlos; Quevedo-Blasco, Raúl; Castro, Ángel (2009). “Ranking de 2008 en productividad en investigación de las universidades públicas españolas”. *Psicothema*, v. 21, n. 2, pp. 309-317. <http://www.ugr.es/~aepc/articulo/ranking08.pdf>
- Campos-Freire, Francisco; Rivera-Rogel, Diana; Rodríguez-Hidalgo, Claudia (2014). “Presence and impact of Andean universities in online social networks”. *Revista latina de comunicación social*, n. 69, pp. 571-592. <http://dx.doi.org/10.4185/RLCS-2014-1025en>
- Cvijikj, Irena P.; Michahelles, Florian (2013). “Online engagement factors on Facebook brand pages”. *Social network analysis and mining*, v. 3, n. 4, pp. 843-861. [http://cocoa.ethz.ch/downloads/2013/07/1253\\_10.1007\\_s13278-013-0098-8.pdf](http://cocoa.ethz.ch/downloads/2013/07/1253_10.1007_s13278-013-0098-8.pdf) <http://dx.doi.org/10.1007/s13278-013-0098-8>
- Del-Barrio-García, Salvador; Arquero, José-Luis; Romero-Frías, Esteban (2015). “Personal learning environments acceptance model: The role of need for cognition, e-learning satisfaction and students’ perceptions”. *Educational technology & society*, v. 18, n. 3, pp. 129-141. <http://digibug.ugr.es/handle/10481/37292#.V0coPZOLQUE>

- Docampo, Domingo** (2012). "Reproducibility of the Shanghai Academic Ranking of World Universities results". *Scientometrics*, v. 94, n. 2, pp. 567-587.  
<http://dx.doi.org/10.1007/s11192-012-0801-y>
- Docampo, Domingo; Herrera, Francisco; Luque-Martínez, Teodoro; Torres-Salinas, Daniel** (2012). "Efecto de la agregación de universidades españolas en el *Ranking de Shanghai (ARWU)*: Caso de las comunidades autónomas y los campus de excelencia". *El profesional de la información*, v. 21, n. 4, pp. 428-432.  
<http://dx.doi.org/10.3145/epi.2012.jul.16>
- Duncan, Tom; Moriarty, Sandra E.** (1998). "A communication-based marketing model for managing relationships". *Journal of marketing*, v. 62, n. 2, pp. 1-13.  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.461.5667&rep=rep1&type=pdf>
- Gómez-Calderón, Bernardo J.; Paniagua-Rojano, Francisco-Javier** (2014). "Las universidades españolas en Twitter: mensajes, contenidos y públicos". *Historia y comunicación social*, v. 19, Extra 1, pp. 681-694.  
<https://revistas.ucm.es/index.php/HICS/article/viewFile/44994/42365>
- González-Díaz, Cristina; Iglesias-García, Mar; Codina, Lluís** (2015). "Presencia de las universidades españolas en las redes sociales digitales científicas: caso de los estudios de comunicación". *El profesional de la información*, v. 24, 5, pp. 1699-2407.  
<http://dx.doi.org/10.3145/epi.2015.sep.12>
- Gonzalo-Brito, Julio; Laaser, Wolfram; Toloza, Eduardo** (2012). "El uso de redes sociales por parte de las universidades a nivel institucional. Un estudio comparativo". *RED. Revista de educación a distancia*, n. 32, pp. 1-38.  
[http://www.um.es/ead/red/32/laaser\\_et\\_al.pdf](http://www.um.es/ead/red/32/laaser_et_al.pdf)
- GWI** (2016). *GWI social. Q1 2016*.  
<http://insight.globalwebindex.net/social>
- Holmberg, Kim** (2015). "Online attention of universities in Finland: Are the bigger universities bigger online too?". En: *Procs of ISSI 2015-15<sup>th</sup> Intl conf of the Intl Society for Scientometrics and Informetrics*, pp. 83-88.  
<http://www.issi2015.org/files/downloads/all-papers/0083.pdf>
- IAB** (2015). *VI Estudio redes sociales de IAB Spain*.  
<http://www.iabspain.net/wp-content/plugins/download-monitor/download.php?id=176>
- Liu, Nian C.** (2015). "The story of Academic Ranking of World Universities". *International higher education*, n. 54, pp. 2-3.  
<http://ejournals.bc.edu/ojs/index.php/ihe/article/view/8409>
- Merisotis, Jamie P.** (2002). "On the ranking of higher education institutions". *Higher education in Europe*, n. 27, n. 4, pp. 361-363.  
<http://dx.doi.org/10.1080/0379772022000071832>
- O'Keeffe, Gwenn S.; Clarke-Pearson, Kathleen** (2011). "The impact of social media on children, adolescents, and families". *Pediatrics*, v. 127, n. 4, pp. 800-804.  
[http://dx.doi.org/10.1593/0127-1674\(2011\)127\[800:IMPMSM\]2.0.CO;2](http://dx.doi.org/10.1593/0127-1674(2011)127[800:IMPMSM]2.0.CO;2)
- <http://dx.doi.org/10.1542/peds.2011-0054>
- Ordorika, Imanol; Rodríguez-Gómez, Roberto** (2010). "El ranking Times en el mercado del prestigio universitario". *Perfiles educativos*, v. 32, n. 19, pp. 7-29.  
[http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0185-26982010000300002](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-26982010000300002)
- Otto, Francis; Williams, Shirley** (2014). "Official use of social network sites by African universities". En: *Procs of Pan African intl conf on information science, computing and telecommunications (PACT)*, Arusha, Tanzania, pp. 46-51.  
<http://dx.doi.org/10.1109/SCAT.2014.7055135>
- Paniagua-Rojano, Francisco-Javier; Gómez-Calderón, Bernardo-José** (2012). "Hacia la comunicación 2.0. El uso de las redes sociales por parte de las universidades españolas". *Revista Icono14. Revista científica de comunicación y tecnologías emergentes*, v. 10, n. 3  
<http://www.icono14.net/ojs/index.php/icono14/article/view/473>
- Pérez, Francisco** (dir.) (2013). *Rankings Issue*.  
<http://www.u-ranking.es/analisis.php>
- Permatasari, Hanunm P.; Harlena, Silvia; Erlangga, Donny; Chandra, Reza** (2013). "Effect of social media on website popularity: Differences between public and private universities in Indonesia". *World of computer science & information technology journal*, v. 3, n. 2, pp. 32-37.  
<http://arxiv.org/pdf/1403.1956.pdf>
- Petty, Richard E.; Cacioppo, John T.** (1986). *The elaboration likelihood model of persuasion*. New York: Springer.  
[http://dx.doi.org/10.1007/978-1-4612-4964-1\\_1](http://dx.doi.org/10.1007/978-1-4612-4964-1_1)
- Petty, Richard E.; Cacioppo, John T.** (2011). *Communication and persuasion: Central and peripheral routes to attitude change*. Springer Science & Business Media. ISBN: 978 1 4612 4964 1
- Roblyer, Margaret D.; McDaniel, Michelle; Webb, Marseena; Herman, James; Witty, James-Vince** (2010). "Findings on Facebook in higher education: A comparison of college faculty and student uses and perceptions of social networking sites". *The internet and higher education*, v. 13, n. 3, pp. 134-140.  
<http://dx.doi.org/10.1016/j.iheduc.2010.03.002>
- Rybalko, Svetlana; Seltzer, Trent** (2010). "Dialogic communication in 140 characters or less: How Fortune 500 companies engage stakeholders using Twitter". *Public relations review*, v. 36, n. 4, pp. 336-341.  
<http://dx.doi.org/10.1016/j.pubrev.2010.08.004>
- Simón-Onieva, José-Eduardo** (2014). "El uso de las redes sociales en el ámbito de la comunicación universitaria andaluza". *Revista internacional de relaciones públicas*, v. 4, n. 8, pp. 139-160.  
<http://revistarelacionespublicas.uma.es/index.php/revrrpp/article/view/285>
- The Cocktail Analysis* (2015). *Sexta oleada del observatorio de redes sociales*.  
<http://es.slideshare.net/TCAAnalysis/6-oleada-observatorio-redes-sociales>

**Tomàs-Folch, Marina; Feixas, Monica; Bernabeu, Maria-Dolors; Ruiz, José-María** (2015). "La literatura científica sobre rankings universitarios: una revisión sistemática". *REDU. Revista de docencia universitaria*, v. 13, n. 3, pp. 33-54.  
<http://red-u.net/redu/index.php/REDU/article/view/828>

**Torres-Salinas, Daniel; Delgado-López-Cózar, Emilio; García-Moreno-Torres, José; Herrera, Francisco** (2011). "Rankings ISI de las universidades españolas según campos científicos: descripción y resultados". *El profesional de la información*, v. 20, n. 1, pp. 111-118.

<http://recyt.fecyt.es/index.php/EPI/article/view/epi.2011.ene.14>  
<http://dx.doi.org/10.3145/epi.2011.ene.14>

**Valerio, Gabriel; Herrera-Murillo, Dagoberto-José; Villanueva-Puente, Fernando; Herrera-Murillo, Natalia; Rodríguez, María-del-Carmen** (2015). "En Facebook el tamaño sí importa. Engagement y el impacto de la longitud del mensaje en las fanpages de las universidades mexicanas". *Revista digital universitaria*, v. 12, n. 1.  
<http://dx.doi.org/10.7238/rusc.v12i1.1887>

## Colección EPI Scholar

Libros científicos de Información, Documentación y Comunicación



Información: Isabel Olea  
[epi.iolea@gmail.com](mailto:epi.iolea@gmail.com)

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

# EVALUADORES DE LA REVISTA

## *El profesional de la información 2014-2015*

**Abad-García, Francisca.** *Departament d'Història de la Ciència i Documentació, Universitat de València.*

**Aceituno-Aceituno, Pedro.** *Departamento de Administración de Empresas, Universidad a Distancia de Madrid.*

**Aguillo, Isidro F.** *Cybermetrics Lab, Instituto de Políticas y Bienes Públicos, Consejo Superior de Investigaciones Científicas, Madrid.*

**Alberich-Pascual, Jordi.** *Facultad de Comunicación y Documentación, Universidad de Granada.*

**Aleixandre-Benavent, Rafael.** *Instituto de Historia de la Ciencia y Documentación López Piñero, Valencia.*

**Alonso-Gamboa, José-Octavio.** *Dirección General de Bibliotecas, Universidad Nacional Autónoma de México.*

**Amat, Carlos B.** *Ingenio (CSIC-UPV), Institute of Innovation and Knowledge Management, Universitat Politècnica de València.*

**Andreu-Sánchez, Celia.** *Departament de Filologia i Comunicació, Universitat de Girona.*

**Arroyo-Vázquez, Natalia.** *Universidad Isabel I, Burgos.*

**Astigarraga, Eneko.** *CodeSyntax, Universidad de Deusto, Bilbao.*

**Baiget, Tomàs.** *El Profesional de la Información, Barcelona.*

**Benavides, David.** *Escuela Técnica Superior de Ingeniería Informática, Departamento de Lenguajes y Sistemas Informáticos, Universidad de Sevilla.*

**Bermejo-Berros, Jesús.** *Facultad de Ciencias Sociales, Jurídicas y de la Comunicación, Universidad de Valladolid.*

**Borrego, Ángel.** *Facultat de Biblioteconomia i Documentació, Universitat de Barcelona.*

**Bueno-de-la-Fuente, Gema.** *Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid.*

**Bustamante-Díaz, Jesús.** *European Commission, Centro Europeo para el Desarrollo de la Formación Profesional (Cedefop), Tesalónica, Grecia.*

**Calderón-Rehecho, Andoni.** *Biblioteca, Universidad Complutense de Madrid.*

**Caldera-Serrano, Jorge.** *Facultad de Ciencias de la Documentación y la Comunicación, Universidad de Extremadura.*

**Campillo-Alhama, Concepción.** *Departament de Comunicació i Psicologia Social, Universitat d'Alacant.*

**Campos-Freire, Francisco.** *Facultad de Ciencias de la Comunicación, Universidade de Santiago de Compostela.*

**Cañabate-Carmona, Antonio.** *Departamento de Organización de Empresas, Universitat Politècnica de Catalunya.*

**Canals, Agustí.** *Estudis de Ciències de la Informació i la Comunicació, Universitat Oberta de Catalunya, Barcelona.*

**Canavilhas, João.** *Labcom - Laboratório de Comunicação OnLine, Universidade da Beira Interior, Portugal.*

**Casero, Andreu.** *Departament de Ciències de la Comunicació, Universitat Jaume I, Castelló de la Plana.*

**Castelló-Martínez, Araceli.** *Departament de Comunicació i Psicologia Social, Universitat d'Alacant.*

**Cerdá-Díaz, Julio.** *Ayuntamiento de Arganda del Rey (Madrid), Universidad Carlos III de Madrid.*

**Cobarsí-Morales, Josep.** *Estudis de Ciències de la Informació i la Comunicació, Universitat Oberta de Catalunya, Barcelona.*

**Codina, Lluís.** *Departament de Comunicació, Universitat Pompeu Fabra, Barcelona.*

**Córdoba, Saray.** *Vicerrectoría de Investigación, Universidad de Costa Rica.*

**Cordón-García, José-Antonio.** *Departamento de Biblioteconomía y Documentación, Universidad de Salamanca.*

**Costas-Comesaña, Rodrigo.** *Centre for Science and Technology Studies (CWTS), Leiden University, Holanda.*

**D'Alòs-Moner, Adela.** *Fundación Biblioteca Social, Barcelona.*

**De-la-Cuadra-Colmenares, Elena.** *Facultad de Ciencias de la Información, Universidad Complutense de Madrid.*

**Del-Fresno-García, Miguel.** *Departamento de Trabajo Social, Universidad Nacional de Educación a Distancia (UNED), Madrid.*

**Del-Valle-Gastaminza, Félix.** *Facultad de Ciencias de la Información, Universidad Complutense de Madrid.*

**Díaz-Noci, Javier.** *Departament de Comunicació, Universitat Pompeu Fabra, Barcelona.*

**Domínguez-Martín, Eva.** *Departament de Comunicació, Universitat Pompeu Fabra, Barcelona.*

**Eito-Brun, Ricardo.** *Grupo GMV, Madrid.*

**Estivill-Rius, Assumpció.** *Facultat de Biblioteconomia i Documentació, Universitat de Barcelona.*

**Feijoo, Claudio.** *European Commission, Institute for Prospective Technological Studies, Joint Research Centre, Sevilla.*

**Fernández-Molina, Juan-Carlos.** *Facultad de Comunicación y Documentación, Universidad de Granada.*

**Fernández-Sande, Manuel.** *Facultad de Ciencias de la Información, Universidad Complutense de Madrid.*

**Fernández-Travieso, Carlota.** *Facultad de Filología, Universidade da Coruña.*

**Ferran-Ferrer, Núria.** *Estudis de Ciències de la Informació i la Comunicació, Universitat Oberta de Catalunya, Barcelona.*

**Ferraro, Gabriela.** *College of Engineering and Computer Science, Australian National University, Canberra, Australia.*

**Ferreras-Rodríguez, Eva.** Consultora en periodismo y comunicación, Madrid.

**Ferrer-Sapena, Antonia.** *Departament de Comunicació Audiovisual, Documentació i Història de l'Art, Universitat Politècnica de València.*

**Freixa-Font, Pere.** *Departament de Comunicació, Universitat Pompeu Fabra, Barcelona.*

**Galán-Cubillo, Esteban.** *Universitat Jaume I de Castelló.*

**Gallardo-Rodríguez, Xavier.** *Universitat Autònoma de Barcelona.*

**García-de-Torres, Elvira.** *Universidad CEU Cardenal Herrera, Valencia.*

**García-Marco, Francisco-Javier.** *Departamento de Ciencias de la Documentación, Universidad de Zaragoza.*

**Gifreu-Castells, Arnau.** *Universidad de Vic - Universidad Central de Cataluña.*

**Giménez-Rayó, Mabel.** *SPW, Madrid.*

**Gómez-Hernández, José-Antonio.** *Facultad de Comunicación y Documentación, Universidad de Murcia.*

**González-Albo-Manglano, Borja.** *Centro de Ciencias Humanas y Sociales, Consejo Superior de Investigaciones Científicas, Madrid.*

**González-Caro, Cristina.** *Ingeniería de Sistemas, Universidad Autónoma de Bucaramanga, Colombia.*

**González-Fernández-Villavicencio, Nieves.** *Área de Biblioteconomía y Documentación, Departamento de Geografía, Historia y Filosofía, Universidad Pablo de Olavide, Sevilla.*

**González-Teruel, Autora.** *Departament d'Història de la Ciència i Documentació, Universitat de València.*

**Gorbea-Portal, Salvador.** *Instituto de Investigaciones Bibliotecológicas y de la Información, Universidad Nacional Autónoma de México.*

**Granell-Canut, Carlos.** *European Commission, Joint Research Centre, Ispra, Italia.*

**Guallar, Javier.** *Facultat de Biblioteconomia i Documentació, Universitat de Barcelona.*

**Hassan, Yusef.** *Grupo SCImago / Universidad de Granada.*

**Hernández-Pérez, Tony.** *Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid.*

**Huertas-Roig, Assumpció.** *Estudis de Comunicació, Universitat Rovira i Virgili, Tarragona.*

**Izquierdo-Castillo, Jéssica.** *Departament de Ciències de la Comunicació, Universitat Jaume I, Castelló de la Plana.*

**Jaramillo, Orlanda.** *Escuela Interamericana de Bibliotecología, Universidad de Antioquia, Colombia.*

**Jiménez-Morales, Mònika.** *Departament de Comunicació, Universitat Pompeu Fabra, Barcelona.*

**Juanatey-Boga, Óscar.** *Facultade de Economía e Empresa, Universidade da Coruña.*

**Keefer-Riva, Alice.** *Facultat de Biblioteconomia i Documentació, Universitat de Barcelona.*

**Labastida, Ignasi.** *Oficina de Difusión del Conocimiento, Universidad de Barcelona.*

**Lacasa-Mas, Iván.** *Facultad de Ciencias de la Comunicación, Universidad Internacional de Catalunya, Barcelona.*

**Lascurain, María-Luisa.** *Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid.*

**Leiva-Aguilera, Javier.** *Universitat de Vic – Universitat Central de Catalunya, Vic (Barcelona).*

**Llorens-Maluquer, Carles.** *Facultat de Comunicació, Universitat Autònoma de Barcelona.*

**López-Cepeda, Ana-María.** *Facultad de Periodismo, Universidad de Castilla-La Mancha.*

**López-del-Ramo, Joaquín.** *Facultad de Ciencias de la Comunicación, Universidad Rey Juan Carlos.*

**López-de-Solís, Iris.** *RTVE, Madrid.*

**López-García, Guillermo.** *Departamento de Teoría de los Lenguajes y Ciencias de la Comunicación, Universitat de València.*

**López-García, Xosé.** *Departamento de Ciencias de la Comunicación, Universidad de Santiago de Compostela, La Coruña.*

**Manfredi Sánchez, Juan.** *Facultad de Periodismo, Universidad de Castilla-La Mancha, Cuenca.*

**Mañá-Terré, Teresa.** *Facultat de Biblioteconomia i Documentació, Universitat de Barcelona.*

**Mañana-Rodríguez, Jorge.** *Centro de Ciencias Humanas y Sociales, Consejo Superior de Investigaciones Científicas, Madrid.*

**Marcos, Mari-Carmen.** *Departament de Comunicació, Universitat Pompeu Fabra, Barcelona.*

**Martínez-Fernández, Valentín-Alejandro.** *Facultade de Ciencias da Comunicación, Universidade da Coruña.*

**Martínez-Martínez, Silvia.** *Estudis de Ciències de la Informació i la Comunicació, Universitat Oberta de Catalunya, Barcelona.*

**Martínez-Urbe, Luis.** *Instituto Juan March de Estudios e Investigaciones, Madrid.*

**Mas, María-Victoria.** *Facultat de Ciències de la Comunicació, Universitat Internacional de Catalunya, Barcelona.*

**Masip, Pere.** *Facultat de Comunicació Blanquerna, Universitat Ramon Llull, Barcelona.*

- Mendo-Carmona, Concepción.** *Facultad de Ciencias de la Documentación, Universidad Complutense de Madrid.*
- Mendoza-García, Isabel.** *Biblioteca, Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid.*
- Micó-Sanz, Josep-Lluís.** *Facultat de Comunicació Blanquerna, Universitat Ramon Llull, Barcelona.*
- Mínguez, Norberto.** *Facultad de Ciencias de la Información, Universidad Complutense de Madrid.*
- Morales-García, Ana-María.** *Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid.*
- Morato, Jorge.** *Departamento de Informática, Universidad Carlos III de Madrid.*
- Moreiro-González, José-Antonio.** *Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid.*
- Moya, Eva.** *S21sec, Instituto de Ciencias Forenses, Universidad Autónoma de Madrid, y Universitat Oberta de Catalunya.*
- Muñoz-Cañavate, Antonio.** *Facultad de Biblioteconomía y Documentación, Universidad de Extremadura.*
- Muriel-Torrado, Enrique.** *Departamento de Ciência da Informação, Universidade Federal de Santa Catarina, Brasil.*
- Noguera-Vivo, José-Manuel.** *Facultad de Ciencias Sociales y de la Comunicación, Universidad Católica San Antonio, Murcia.*
- Ojer-Goñi, Teresa.** *Facultad de Comunicación, Universidad San Jorge, Villanueva de Gállego (Zaragoza).*
- Olabe-Sánchez, Miguel-Fernando.** *Facultad de Humanidades y Ciencias de la Comunicación, Universidad CEU Cardenal Herrera, Valencia.*
- Olea, Isabel.** *El Profesional de la Información, León.*
- Olmeda-Gómez, Carlos.** *Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid.*
- Orduña-Malea, Enrique.** *Escuela Técnica Superior de Informática, Universitat Politècnica de València.*
- Ortiz-Repiso, Virginia.** *Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid.*
- Ortiz-Sobrino, Miguel-Ángel.** *Facultad de Ciencias de la Información, Universidad Complutense de Madrid.*
- Ortoll, Eva.** *Estudis de Ciències de la Informació i la Comunicació, Universitat Oberta de Catalunya, Barcelona.*
- Palacios-Madrid, Vicente.** *Departamento de Informática, Universidad Carlos III de Madrid.*
- Palafox-Parejo, Manuela.** *Biblioteca, Universidad Complutense de Madrid.*
- Paniagua-Rojano, Francisco-Javier.** *Departamento de Periodismo, Universidad de Málaga.*
- Pascual-Miguel, Félix-José.** *Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid.*
- Pascual, Víctor.** *Consultor en ciencias de la computación y comunicación digital.*
- Pastor-Sánchez, Juan-Antonio.** *Facultad de Comunicación y Documentación, Universidad de Murcia.*
- Pedraza-Gracia, Manuel-José.** *Departamento de Ciencias de la Documentación, Universidad de Zaragoza.*
- Pérez-González, Daniel.** *Departamento de Administración de Empresas, Universidad de Cantabria.*
- Pérez-Montoro, Mario.** *Facultat de Biblioteconomia i Documentació, Universitat de Barcelona.*
- Pérez-Soler, Susana.** *Fundació Tecnocampus Mataró-Maresme / Facultat de Comunicació Blanquerna, Universitat Ramon Llull, Barcelona.*
- Perianes-Rodríguez, Antonio.** *Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid.*
- Peset, Fernanda.** *Departament de Comunicació Audiovisual, Documentació i Història de l'Art, Universitat Politècnica de València.*
- Pinto, Adilson-Luiz.** *Departamento de Ciência da Informação, Universidade Federal de Santa Catarina, Brasil.*
- Ramírez-Velázquez, César-Augusto.** *Instituto de Investigaciones Bibliotecológicas y de la Información, Universidad Nacional Autónoma de México.*
- Robinson-García, Nicolás.** *Facultad de Comunicación y Documentación, Universidad de Granada.*
- Robledano-Arillo, Jesús.** *Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid.*
- Rodríguez-Ruiz, Óscar.** *Departamento de Organización de Empresas, Universidad Complutense de Madrid.*
- Rodríguez-Yunta, Luis.** *Centro de Ciencias Humanas y Sociales, Consejo Superior de Investigaciones Científicas, Madrid.*
- Roig-Telo, Antoni.** *Estudis de Ciències de la Informació i la Comunicació, Universitat Oberta de Catalunya, Barcelona.*
- Rovira, Cristòfol.** *Departament de Comunicació, Universitat Pompeu Fabra, Barcelona.*
- Rueda-Ramírez, Pedro.** *Facultat de Biblioteconomia i Documentació, Universitat de Barcelona.*
- Salaberria-Lizarazu, Ramón.** *Biblioteca Vasconcelos, México DF.*
- Salaverría, Ramón.** *Facultad de Comunicación, Universidad de Navarra.*
- Salgado-de-Dios, Francesc.** *Departament de Comunicació, Universitat Pompeu Fabra, Barcelona.*
- San-José-Montano, Blanca.** *Biblioteca, Hospital Universitario de Móstoles, Madrid.*
- Sánchez-Jiménez, Rodrigo.** *Facultad de Ciencias de la Documentación, Universidad Complutense de Madrid.*
- Sánchez-Navarro, Jordi.** *Estudis de Ciències de la Informació i la Comunicació, Universitat Oberta de Catalunya, Barcelona.*

**Sanz-Casado, Elías.** *Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid.*

**Scolari, Carlos.** *Departament de Comunicació, Universitat Pompeu Fabra, Barcelona.*

**Sellas, Toni.** *Departament de Filologia i Comunicació, Universitat de Girona.*

**Sevillano-Pintado, Olga.** *Museo Nacional Centro de Arte Reina Sofía, Madrid.*

**Spinak, Ernesto-Luis.** *Consultor, Montevideo, Uruguay.*

**Tardón, Eugenio.** *Biblioteca, Universidad Complutense de Madrid.*

**Tejedor, Santiago.** *Departament de Periodisme y Ciències de la Comunicació, Universitat Autònoma de Barcelona.*

**Tejada-Artigas, Carlos.** *Facultad de Ciencias de la Documentación, Universidad Complutense de Madrid.*

**Torres-Padrosa, Víctor.** *Institute of Informatics and Applications, Universitat de Girona.*

**Torres-Salinas, Daniel.** *Centro de Investigación Médica Apli-*

*cada, Universidad de Navarra.*

**Tramullas, Jesús.** *Departamento de Ciencias de la Documentación, Universidad de Zaragoza.*

**Trujillo, Javier.** *Consultoría informática, Madrid.*

**Tur-Viñes, Victoria.** *Facultat de Ciències Econòmiques i Empresarials, Universitat d'Alacant.*

**Valls-Pasola, Anna.** *Biblioteca, Institut del Teatre, Barcelona.*

**Villaseñor, Isabel.** *Facultad de Ciencias de la Documentación, Universidad Complutense de Madrid.*

**Vivas-Moreno, Agustín.** *Facultad de Biblioteconomía y Documentación, Universidad de Extremadura.*

**Xifra-Triadú, Jordi.** *Departament de Comunicació, Universitat Pompeu Fabra, Barcelona.*

**Yegros, Alfredo.** *Centre for Science and Technology Studies (CWTS), Leiden University, Holanda.*

**Zozaya-Montes, Leonor.** *Fundação para a Ciência e a Tecnologia, Coimbra, Portugal.*



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

RedIRIS

IWETEL

Foro para profesionales de bibliotecas y documentación

Con unos 6.000 miembros, *IweTel* es la mayor lista de distribución en castellano para debatir y estar al día sobre temas de biblioteconomía y documentación.

Fue creada en 1993 por Tomàs Baiget, como complemento de *Information World en Español (IWE)*, revista que en 1999 pasó a denominarse *El profesional de la información (EPI)*.

Desde 1998 *IweTel* está alojada en el servicio de listas de RedIRIS, siendo posible consultar en sus archivos estos 18 años de la historia de la documentación en España:  
<https://listserv.rediris.es/cgi-bin/wa?A0=IWETEL>

La lista cuenta con 4 moderadores que permanentemente filtran los mensajes para evitar spam, mensajes inapropiados, anuncios, mensajes repetidos, etc.:

**David Gómez** (*Observatorio de la Infancia de Andalucía*),

**Isabel Olea** (*EPI, León*),

**Nieves González-Fernández-Villavicencio** (*Universidad Pablo de Olavide, Sevilla*),

**Tomàs Baiget** (*EPI, Barcelona*).

Puedes suscribirte a *IweTel* en:

<https://listserv.rediris.es/cgi-bin/wa?A0=IWETEL>

**julio 2016-junio 2017**

**25-28 de julio de 2016**

3<sup>RD</sup> ANNUAL INTL CONF ON LIBRARY AND INFORMATION SCIENCE  
Atenas  
*Athens Institute for Education and Research (Atiner)*  
<http://www.atiner.gr/library>

**3-5 de agosto de 2016**

VI SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO (SECIN)  
Londrina, Brasil  
*Universidade Estadual de Londrina; Departamento de Ciência da Informação; Programa de Mestrado em Ciência da Informação.*  
<http://www.uel.br/eventos/cinf/index.php/secin2016/secin2016>

**13-19 de agosto de 2016**

IFLA WLIC 2016. World library and information congress  
80<sup>th</sup> IFLA General conf and assembly  
Columbus, Ohio  
<http://conference.ifla.org/past-wlic/2014/ifla80/node/887.html>



**18 de agosto de 2016**

AICSSH. Academic intl conf on social sciences and humanities  
Oxford  
<http://www.flelearning.co.uk/aicssh>

**5-9 de septiembre de 2016**

20<sup>TH</sup> INTL CONF ON THEORY AND PRACTICE OF DIGITAL LIBRARIES. Overcoming the limits of digital archives  
Hannover  
<http://www.tpd12016.org>

**7-9 de septiembre de 2016**

II CONGRESO INTL SOBRE LIBRO MEDIEVAL Y MODERNO  
Zaragoza  
<http://congresolibroantiguo.weebly.com>

**9 de septiembre de 2016**

REPSCIENCE 2016. First intl workshop on reproducible open science. Celebrado conjuntamente con TPD1 2016  
Hannover  
*Research Data Alliance Europe and OpenAIRE2020*  
<http://repscience2016.research-infrastructures.eu>

**11-12 de septiembre de 2016**

INTL CONF ON HUMANITIES AND SOCIAL SCIENCES  
Venecia  
*The Institute of Science, Technology & Development Studies*  
<http://www.istdst.org/HSS>

**12-14 de septiembre de 2016**

3<sup>ER</sup> ENCUENTRO DE HUMANISTAS DIGITALES  
México DF  
*Red de Humanidades Digitales (RedHD), El Colegio de México, Universidad Nacional Autónoma de México (UNAM)*  
<http://www.humanidadesdigitales.net/3ehd>



**14-16 de septiembre de 2016**

ALPSP Conf  
Londres  
*Association of Learned and Professional Society Publishers*  
<http://www.alpsp.org/Conference>

**18-20 de septiembre de 2016**

MERCOSUR 2016. XI Encontro de diretores e X Encontro de docentes de escolas de biblioteconomia e ciência da informação do Mercosul  
Belo Horizonte, Brasil  
*Escola de Ciência da Informação da Universidade Federal de Minas Gerais*  
<http://mercosur2016.eci.ufmg.br>

**19-21 de septiembre de 2016**

ICDIM 2016. 11<sup>th</sup> Intl conf on digital information management  
5<sup>th</sup> Workshop on emerging problem- specific crowdsourcing technologies  
5<sup>th</sup> Workshop on advanced techniques on data analytics and data visualization  
4<sup>th</sup> IEEE Intl workshop on data management  
2<sup>nd</sup> Workshop on internet of things  
2<sup>nd</sup> Workshop on big data mining  
2<sup>nd</sup> Workshop on cluster computing  
2<sup>nd</sup> Workshop on intelligent information systems  
Porto, Portugal  
*IEEE Technology Engineering Management Society*  
<http://www.icdim.org>



**21-23 de septiembre de 2016**

X EDICIC. Encontro da Associação de Educadores e Pesquisadores de Biblioteconomia, Arquivologia, Ciências da Informação e Documentação da Iberoamérica e Caribe Pampulha, Belo Horizonte, Minas Gerais  
*Escola de Ciência da Informação (ECI), Universidade Federal de Minas Gerais (UFMG)*  
<http://edicic2016.eci.ufmg.br>

**27-29 de septiembre de 2016**

14<sup>TH</sup> ISKO INTL CONF  
Rio de Janeiro  
*Getulio Vargas Foundation*  
*iskorio@gmail.com*  
[http://isko-brasil.org.br/?page\\_id=711](http://isko-brasil.org.br/?page_id=711)

**28-29 de septiembre de 2016**

3<sup>RD</sup> ALTMETRICS CONFERENCE (3:AM)  
Bucarest  
Organizers: *DataCite, Wellcome Trust, Altmetric, Elsevier y Crossref*  
Sponsors: *Plum Analytics y Frontiers*  
<http://altmetricsconference.com>



**29-30 de septiembre de 2016**

XV WORKSHOP DE REBIUN. Gestión de datos  
Castelló de la Plana  
*Red de Bibliotecas Universitarias Españolas; Universitat Jaume I*  
*Lidón París-Folch, paris@uji.es*

**7-8 de octubre de 2016**

FUENTES DIGITALES E HISTORIA DE LA LENGUA  
San Millán de la Cogolla, La Rioja  
*Instituto Historia de la Lengua de Cilengua; Fundación San Millán de la Cogolla*  
<http://www.cilengua.es>

**12-13 de octubre de 2016**

ECIL 2016. European conf on information literacy  
Praga  
*Department of Information Management, Hacettepe University; Department of Information and Communication Sciences, Zagreb University; Association of Libraries of Czech Universities (ALCU)*  
<http://ecil2016.ilconf.org>



**12-14 de octubre de 2016**

LIBER. Feria internacional del libro  
Barcelona  
<http://www.salonliber.es>  
<http://federacioneditores.org>

**13-14 de octubre de 2016**

I CONGRESSO CIENTÍFICO BRASILEIRO DA WIKIPÉDIA. Wikipédia na disseminação do conhecimento científico  
Rio de Janeiro  
*Universidade de Rio de Janeiro*  
<http://ccbwiki.wikimedia.org.br>

**13-16 de octubre de 2016**

DC-2016 METADATA SUMMIT. Intl conf & annual meeting  
Copenhage  
Simultáneo al ASIS&T Annual Meeting  
<http://dcevents.dublincore.org/index.php/IntConf/dc-2016/schedConf>



**14 de octubre de 2016**

1r Congreso de Gestores de Información Científica  
VIII Seminario de Bibliotecas Biomédicas  
Chile  
*Sociedad Evidenti@, Corporación Cultural de Las Condes; Colegio de Bibliotecarios de Chile*  
*bibliotecasbiomedicas@gmail.com*

**14-18 de octubre de 2016**

ASIS&T Annual Meeting  
Copenhage  
*Association for Information Science and Technology*  
<https://www.asist.org/events/annual-meeting/annual-meeting-2016>

**17-19 de octubre de 2016**

BIREDIAL-ISTEC'16. VI Conf Intl sobre Bibliotecas y Repositorios Digitales de América Latina  
SIBD'16. XI Simposio Intl de Biblioteca Digitales  
San Luis Potosí, México  
*Consorcio Iberoamericano para Educación en Ciencia y Tecnología (ISTEC), etc.*  
<http://congresos.unlp.edu.ar/public/conferences/35/boletin/1>

**18 de octubre de 2016**

THE MARKETS. Pre-fair Buchmesse conference. Publishing perspectives. Global publishing summit. Focus countries: Brazil, Flanders & The Netherlands, Philippines, Poland, Spain, United Arab Emirates and the United Kingdom.  
Frankfurt  
<http://www.markets2016.com>

**18-19 de octubre de 2016**

INTERNET LIBRARIAN INTERNATIONAL (ILI)

Londres

Tel.: +44-1865 327 813

[info.uk@infotoday.com](mailto:info.uk@infotoday.com)<http://www.internet-librarian.com/2016>**19 de octubre de 2016**

II JORNADAS TÉCNICAS DE BIBLIOTECAS. Marketing bibliotecario a través de las redes sociales

Cádiz

*Asociación Andaluza de Bibliotecarios; Dirección General de Innovación Cultural y del Libro de la Junta de Andalucía.*<http://www.aab.es/jornadas/ii-jornadas-t%C3%A9cnicas-de-bibliotecas>**19-23 de octubre de 2016**

BUCHMESSE. Frankfurt book fair

Frankfurt

<http://www.buchmesse.de/en/fbf>**20 de octubre de 2016**

II JORNADAS TÉCNICAS DE BIBLIOTECAS. Marketing bibliotecario a través de las redes sociales

Almería

*Asociación Andaluza de Bibliotecarios; Dirección General de Innovación Cultural y del Libro de la Junta de Andalucía.*<http://www.aab.es/jornadas/ii-jornadas-t%C3%A9cnicas-de-bibliotecas>**21 de octubre de 2016**

AIB CILW 2016 CONFERENCE

Roma

*Associazione Italiana Biblioteche, Sezione Lazio; Biblioteca nazionale centrale di Roma*<http://www.aib.it/attivita/congressi/c2016/giornata-studi-aib-cilw-2016>**24-26 de octubre de 2016**

IBERSID 2016. XXI Encuentros Internacionales sobre Sistemas de Información y Documentación

Zaragoza

*Francisco-Javier García-Marco, Depto. de Ciencias de la Documentación e Historia de la Ciencia, Fac. de Filosofía y Letras, Univ. de Zaragoza*<http://www.ibersid.org>**29 de octubre de 2016**ICICM 2016. 6<sup>th</sup> Intl conf on information communication and management

Hertfordshire, UK

<http://www.icicm.org>**31 de octubre-4 de noviembre de 2016**

INFO'2016. XIV Congreso Internacional de Información. Información, Evolución y Sostenibilidad

II Simposio sobre la Conservación del Patrimonio Documental La Habana

*Instituto de Información Científica y Tecnológica, del Ministerio de Ciencia, Tecnología y Medio Ambiente*<http://www.congreso-info.cu>**2-3 de noviembre de 2016**

VIII SEMINARIO DE ESTUDIOS CUANTITATIVOS Y CUALITATIVOS DE LA CIENCIA Y LA TECNOLOGÍA. INFO'2016

La Habana

<http://www.congreso-info.cu>**3-4 de noviembre de 2016**THE 21<sup>ST</sup> NORDIC WORKSHOP ON BIBLIOMETRICS AND RESEARCH POLICY

Copenhagen

*Aalborg University*<http://nwb.aau.dk>**7-9 de noviembre de 2016**

AAHD 2016. Congreso intl - Humanidades digitales: construcciones locales en contextos globales

Buenos Aires

<http://www.academica.org/aahd.congreso>**8-9 de noviembre de 2016**

CARA 2016. II Colóquio de Análise de Redes Aplicada Rio de Janeiro

*Fundação Oswaldo Cruz*<http://cara2016.fiocruz.br>

**9-11 de noviembre de 2016**

JORNADAS ARCHIVANDO 2016. Usuarios, retos y oportunidades

León

Fundación Sierra-Pambley

<http://jornadasarchivando.sierrapambley.org>

**9-11 de noviembre de 2016**

XIX CONGRESO INTL EDUTEC

Alicante

Universidad de Alicante

<http://edutec.ddgde.ua.es>



**11 de noviembre de 2016**

VII JORNADA TEMAS ACTUALES EN BIBLIOTECOLOGÍA (TAB)

Mar del Plata

Biblioteca del Centro Médico de Mar del Plata

[bibliotecamp@gmail.com](mailto:bibliotecamp@gmail.com)

**11 de noviembre de 2016**

#3WPGCI. III Workshop de pós-graduação em ciência da informação

Lisboa

GT de Ciência da Informação da SOPCOM; Faculdade de Letras da Universidade de Lisboa

<http://eseig.ipp.pt/conferencias/index.php/wpgci/2015/index>

**14-17 de noviembre de 2016**

V SEMINARIO HISPANO BRASILEÑO. Investigación en información, documentación y sociedad

Madrid

Facultad Ciencias de la Documentación, Universidad Complutense de Madrid

<http://seminariohispano-brasileiro.org.es/ocs/index.php/shb/2016>



**16-17 de noviembre de 2016**

VIII CONGRESO NACIONAL DE BIBLIOTECAS PÚBLICAS

Toledo

Ministerio de Educación, Cultura y Deporte, Subdirección General de Coordinación Bibliotecaria

<http://www.mecd.gob.es/cultura-mecd/areas-cultura/bibliotecas/novedades/viii-congreso-bp.html>

**16-18 de noviembre de 2016**

14ES JORNADES IMATGE I RECERCA

Girona

Centre de Recerca i Difusió de la Imatge (Ajuntament de Girona); Associació d'Arxivers – Gestors de Documents de Catalunya (AAC-GD)

[http://www.girona.cat/sgdap/cat/jornades\\_presentacio.php](http://www.girona.cat/sgdap/cat/jornades_presentacio.php)

**17-19 de noviembre de 2016**

AHLIST 2016. Centros y periferias en movimiento: confluencia, fortalecimiento e innovación

Las Palmas

Association of History, Literature, Science and Technology

Universidad de Las Palmas de Gran Canaria

<http://www.ahlist.org>

**22 de noviembre de 2016**

4<sup>th</sup> INTL SEMANTIC WEB AND LINKED OPEN DATA WORKSHOP (ISW-LOD). Co-located with IBERAMIA 2016

San José, Costa Rica

Sociedad Iberoamericana de Inteligencia Artificial

<http://ciep.ing.uaslp.mx/sw-lod/index.html>



**22-25 de noviembre de 2016**

MTSR'16. 10<sup>th</sup> Intl conf on metadata and semantics research

Roma

Food and Agriculture Organization (FAO)

<http://www.mtsr-conf.org>



**28-29 de noviembre de 2016**

SWIB16. Semantic web in libraries conference

Bonn

ZBW - German National Library of Economics, Leibniz Information Centre for Economics; North Rhine-Westphalian Library Service Centre (hbz)

<http://swib.org/swib16>



**1-3 de diciembre de 2016**

ICOA 2016. Intl conf on open access. OA to scientific publications: Between use and preservation of the digital memory  
Túnez

*Institut Supérieur de Documentation, University of Manouba; The National University Center for Scientific and Technical Documentation (Cnudst), Túnez*  
<http://icoa2016.sciencesconf.org>



**6-7 de diciembre de 2016**

LONDON INFO INTERNATIONAL

Londres

<http://londoninfointernational.com>

**12-15 de diciembre de 2016**

12<sup>TH</sup> INTL CONF ON WEBOMETRICS, INFORMETRICS AND SCIENTOMETRICS (WIS) & 17<sup>TH</sup> COLLNET MEETING

Nancy, Francia

<http://www.slp.org.in/collnet2016>

**15-18 de febrero de 2017**

WRITING RESEARCH ACROSS BORDERS (WRAB) IV  
Bogotá

*Sociedad Internacional para el Avance de la Investigación en Escritura (ISAWR). Pontificia Universidad Javeriana*  
<http://wrab2017.com/javeriana>



**10-12 de abril de 2017**

40<sup>TH</sup> UKSG ANNUAL CONFERENCE

Harrogate, Yorkshire, UK

<http://www.uksg.org/event/conference17>



**4-5 de mayo de 2017**

7<sup>a</sup> CRECS. Conf intl de revistas de ciencias sociales y humanidades

Cuenca

*Facultad de Ciencias de la Educación y Humanidades. Universidad de Castilla La Mancha, Campus de Cuenca.*

*El Profesional de la Información; Grupo ThinkEPI; Universidad de Castilla La Mancha*

<http://crecs.info>



**25-26 de mayo de 2017**

FESABID 2017. XV Jornadas Españolas de Documentación  
Pamplona

*Asociación Navarra de Bibliotecarios (Asnabi); Federación Española de Sociedades de Archivística, Biblioteconomía, Documentación y Museística*

<http://www.asnabi.com>

<http://www.fesabid.org>



**12-16 de junio de 2017**

EAHIL 2017. 12<sup>th</sup> Intl congress on medical librarianship (ICML)

2017 European Association for Health Information and Libraries (EAHIL) Workshop

Dublin

<http://eahil2017.net>



# PUBLICACIONES EL PROFESIONAL DE LA INFORMACIÓN

## REVISTA EL PROFESIONAL DE LA INFORMACIÓN

Desde 2015 sólo online

- Suscripción anual institucional ..... 137 €  
 Suscripción anual individual ..... 84 €

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

## ANUARIO THINKEPI (versión online)

### Tarifas institucionales

- Anuario 2016 ..... 80 €  
 Anuario 2015 ..... 75 €

### Tarifas individuales (particulares)

- Anuario 2016 ..... 48 €  
 Anuario 2015 ..... 45 €

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

## EPI + ANUARIO THINKEPI (suscripción conjunta)

- Suscripción anual institucional ..... 202 €  
 Suscripción anual individual ..... 116 €

## COLECCIÓN LIBROS EPI SCHOLAR (Editorial UOC)

1. Cibermetría. Midiendo el espacio red ..... 20,00€  
Enrique Orduña-Malea e Isidro F. Aguillo
2. La web social como nuevo medio de comunicación  
y evaluación científica ..... 21,00€  
Amalia Mas-Bleda e Isidro F. Aguillo
3. SEO. Introducción a la disciplina del posicionamiento  
en buscadores ..... 20,00€  
Jorge Serrano Cobos
4. Métricas de la web social para bibliotecas ..... 26,00€  
Nieves González-Fernández-Villavicencio
5. Calidad en sitios web. Métodos de análisis general,  
e-commerce, imágenes, hemerotecas y turismo ..... 19,50€  
Rafael Pedraza-Jiménez, Lluís Codina, Javier Guallar
6. Gestión de datos de investigación ..... 17,00€  
Fabiano Couto Corrêa

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

## LIBROS EL PROFESIONAL DE LA INFORMACIÓN

(Editorial UOC)

1. Tecnologías de la web semántica ..... 12,00 €  
Juan-Antonio Pastor
2. La revolución del libro electrónico ..... 11,50 €  
José-Antonio Cordón
3. Sistemas de información en la empresa ..... 11,50 €  
Josep Cobarsí-Morales
4. Información en el móvil ..... 12,00 €  
Natalia Arroyo-Vázquez
5. Acceso abierto a la ciencia ..... 12,00 €  
Ernest Abadal
6. Fuentes de información médica ..... 12,00 €  
Pablo Medina-Aguerrebere
7. Gestión de la reputación online ..... 11,50 €  
Javier Leiva-Aguilera
8. Wikipedia de la A a la W ..... 12,50 €  
Tomás Saorín
9. Etiquetar en la web social ..... 12,00 €  
Raquel Gómez-Díaz
10. Mejorar las búsquedas de información ..... 12,00 €  
Silvia Argudo y Amadeu Pons
11. Clubes de lectura ..... 11,50 €  
Óscar Carreño
12. Plan social media y community manager ..... 13,50 €  
Julián Marquina-Arenas
13. Documentación audiovisual en televisión ..... 12,00 €  
Jorge Caldera y Pilar Arranz
14. Gestión de documentos en la e-administración ..... 12,00 €  
Elisa García-Morales
15. El film researcher ..... 12,00 €  
Iris López-de-Solis
16. Preservación digital ..... 12,00 €  
Miquel Térmens
17. Gestión de contenidos ..... 13,00 €  
Ricardo Eito-Brun
18. Documentación fotográfica ..... 14,00 €  
Juan-Miguel Sánchez-Vigil y Antonia Salvador-Benitez
19. Documentación cinematográfica ..... 12,00 €  
Elena De la Cuadra
20. Archivos ..... 14,00 €  
Ramón Alberch-Fugueras
21. Inteligencia en redes sociales ..... 11,00 €  
Eva Moya
22. Bibliotecas escolares ..... 10,00 €  
Concepción Mª Jiménez-Fernández y Raúl Cremades-García
23. Marca y comunicación empresarial ..... 11,00 €  
Pablo Medina-Aguerrebere
24. El content curator ..... 14,00 €  
Javier Guallar y Javier Leiva-Aguilera
25. Gestión de la calidad en la biblioteca ..... 13,50 €  
Nuria Balagué y Jarmo Saarti
26. Innovación en bibliotecas ..... 13,50 €  
Isabel Riaza
27. La intranet social ..... 13,50 €  
Ana Carrillo Pozas
28. Los impresos antiguos ..... 12,50 €  
Jon Zabala
29. Los sexenios de investigación ..... 12,00 €  
Álvaro Cabezas-Clavijo y Daniel Torres-Salinas
30. Geobibliotecas ..... 11,50 €  
Estefanía Aguilar-Moreno y Carlos Granell-Canut
31. Biblioteca pública. Mientras llega el futuro ..... 14,50 €  
Fernando Juárez-Urquijo
32. Geolocalización online. La importancia del dónde ..... 12,00 €  
Gersón Beltrán
33. Lectura digital infantil. Dispositivos, aplicaciones y contenidos... 13,00 €  
Araceli García-Rodríguez y Raquel Gómez Díaz



En el caso de las publicaciones en papel hay que  
añadir los gastos de envío

**Información y pedidos:**

**Isabel Olea**

[epi.iolea@gmail.com](mailto:epi.iolea@gmail.com)

Tel.: +34-608 491 521

<http://www.elprofesionaldelainformacion.com>

<http://www.thinkepi.net>

# Nada se sabe bien sino por medio de la experiencia. *open*

Sir Francis Bacon

## KOHA en el MUNDO

- + 3000 instalaciones
- + 35 instalaciones de Koha-kobli
- + 350 desarrolladores
- + 80 idiomas

## MASmedios y KOHA

- +35 instalaciones
- Desde 2007 desarrollando en Koha
- Empresa soporte reconocido en Koha Community
- Implantadores de Koha-Kobli
- 90% de los proyectos incluyen migración de otros sistemas
- Especialistas en Open Source Library System
- Expertos en integración e interoperabilidad

MASmedios desarrolladores de Sistemas abiertos:



Deseo suscribirme a la revista EPI a partir del mes de enero del año 20  Las suscripciones van por años naturales, de enero a diciembre

Suscripción  Institucional  Personal

Nombre  Institución

(Los suscriptores individuales no han de escribir ningún nombre de institución, sólo indicar la dirección particular)

Departamento  NIF institucional

Dirección  Código postal  Ciudad  País

Teléfono  Fax  Correo-e

#### Método de pago

Tarjeta de crédito  VISA  Master Card

Titular de la tarjeta  Código de seguridad CVC2

Número de tarjeta  Caducidad (mm/aaaa)

Cheque nominativo en euros a nombre de *El profesional de la información*

Transferencia bancaria a la cuenta de La Caixa **ES95 2100 0818 93 0200745544** Enviar, fotocopiado o escaneado, el resguardo de la transferencia.

Giro postal al apartado de correos 32.280 de Barcelona Enviar, fotocopiado o escaneado, el resguardo del giro.

Domiciliación en cuenta bancaria

Entidad  Oficina  DC  Núm

Titular de la cuenta

PayPal a la cuenta EPISCP@gmail.com



Enviar el boletín cumplimentado por correo electrónico, o postal al APARTADO 32.280 - 08080 BARCELONA - ESPAÑA.  
Consultas: suscripciones@elprofesionaldelainformacion.com o +34 609 352 954

**Open choice.** Los autores pueden liberar su artículo en open access en las webs de EPI mediante el pago de 500 €

## Precios 2016

**Importante: Desde enero de 2015 EPI sólo se publica online**

Suscripción institucional EPI: 113,23 € + 21% IVA = 137 €

Suscripción institucional conjunta EPI + Anuario ThinkEPI: 166,95 € + 21% IVA = 202 €

Número suelto: 35,57 € + IVA = 37 €

Suscripción personal: 69,43 € + 21% IVA = 84 €

Fuera de Europa no se cobra el IVA

Suscripción personal conjunta EPI + Anuario ThinkEPI: 95,87 € + 21% IVA = 116 €

## NORMAS PARA LOS AUTORES

*El profesional de la información* tiene dos secciones principales:

ARTÍCULOS: Trabajos de investigación y temas analizados en profundidad.

ANÁLISIS: Experiencias, estudios de casos, análisis de productos, reseñas, etc.

Las contribuciones han de ser originales e inéditas, no pueden haberse publicado previamente en soporte papel o electrónico. El tamaño ideal es de 3.500 palabras, aunque en algunos casos la Redacción puede autorizar una mayor extensión.

El texto ha de enviarse en Word, rtf u odt. Las tablas deberán ir pegadas en el mismo documento. Todos los materiales gráficos (diagramas, fotografías, capturas de pantalla, etc.) deben pegarse en el Word y además enviarse en ficheros independientes (en formatos xls, jpg, pdf, etc.). Las imágenes jpg deben tener una resolución de al menos 300 pp (unos 200 KB cada una).

El texto debe presentarse completamente plano, sin autoformatos ni automatismos de Word (subsecciones, viñetas, citas enlazadas, pies de página, sangrías, tabulaciones, colores, etc.), pero debe seguir el estilo de EPI en cuanto a **negritas** (nombres de los autores citados), *cursivas* (instituciones, títulos de revista, marcas) y mayúsculas. Los urls deben estar sin hipervínculo.

Las citas bibliográficas en el texto se realizarán de la forma: (**Apellido**, año) o (**ApellidoAutor1**; **ApellidoAutor2**, año).

La redacción debe ser concisa y precisa, evitando la retórica.

Los trabajos deben incluir: a) título, b) resumen de 100-150 palabras, c) 5-10 palabras clave, d) title, e) abstract de 100-150 palabras, f) 5-10 keywords.

Aparte se incluirá el nombre de los autores, su lugar de trabajo y dirección (postal y electrónica), su foto tipo carnet superior a 100 KB en jpg, un currículum de unas 70 palabras, y su orcid

#### EVALUACIÓN

Los trabajos son revisados según el sistema tradicional "peer review" en doble ciego por al menos dos expertos en el tema, del Consejo Asesor de la revista y/o externos. La revista se compromete a informar del resultado a los autores.

#### REFERENCIAS BIBLIOGRÁFICAS

Ordenadas alfabéticamente por autor, se limitarán a las obras citadas en el artículo. No se acepta bibliografía de relleno.

Artículos de una publicación periódica:

**Apellido, Nombre; Apellido2, Nombre2.** "Título del artículo".

*Título de la publicación periódica*, año, v., n., pp. xx-yy.

*Dirección url iniciada en nueva línea sin barra y sin punto finales*

<http://dx.doi.org/10.xxxx/doi>

Ponencia presentada en un congreso:

**Apellido, Nombre; Apellido2, Nombre2.** "Título de ponencia".

En: *Nombre del congreso*, año, pp. xx-yy.

*Dirección url iniciada en nueva línea sin barra y sin punto finales*

<http://dx.doi.org/10.xxxx/doi>

Monografías e informes:

**Apellido, Nombre; Apellido2, Nombre2.** *Título del trabajo.*

Lugar de publicación: editor, fecha, ISBN: espacios, sin guiones

*Dirección url iniciada en nueva línea sin barra y sin punto finales*

<http://dx.doi.org/10.xxxx/doi>

Capítulo de una monografía:

**Apellido, Nombre; Apellido2, Nombre2.** "Título del capítulo".

En: Apellido, Nombre; Apellido2, Nombre2. *Título de la monografía.* Lugar de publicación: editor, fecha, pp. xx-yy. ISBN: espacios, sin guiones

*Dirección url iniciada en nueva línea sin barra y sin punto finales*

Recurso en línea:

**Apellido, Nombre; Apellido2, Nombre2.** *Título del recurso (sin fecha de la consulta).*

*Dirección url iniciada en nueva línea sin barra y sin punto finales*

Todas las contribuciones se tienen que enviar a la sección EPI de la plataforma OJS del Repositorio Español de Ciencia y Tecnología (Recyt) de la Fecyt:

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

Previamente los autores deben registrarse en:

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