

# Una perspectiva documental y bibliotecológica sobre el *big data* y el periodismo de datos

Antonio García Jiménez\*  
Beatriz Catalina García\*

Artículo recibido:  
3 de agosto de 2015  
Artículo aceptado:  
29 de abril de 2016

## RESUMEN

La aplicación de nuevos sistemas para la creación, almacenamiento y difusión de la información ha tenido como uno de los factores clave el exponencial desarrollo de la apertura de datos a través de mecanismos técnicos y computacionales. El *open data* y el *big data*, así como las redes sociales, han transformado el trabajo en los medios de comunicación en un mundo informativamente globalizado. Esta transformación conlleva también cambios en las tareas de documentalistas y periodistas. Mediante el análisis sobre la literatura existente, se analizan los procesos, métodos

\* Universidad Rey Juan Carlos, España. antonio.garcia@urjc.es  
beatriz.catalina@urjc.es

e instrumentos aplicados en la documentación de medios y en el periodismo a partir de la perspectiva del *big data*, evidenciando casos particulares. Aparte de los cambios rutinarios de los profesionales, se observa una serie de contradicciones entre la web semántica y el *big data* conforme a los objetivos de transparencia y comprensión de la información. Finalmente es necesario plantear el valor real que la difusión de datos aporta frente al filtrado de éstos desde el plano puramente documental.

**Palabras clave:** *Open data*; *Big data*; Redes sociales; Medios de comunicación; Documentación; Bibliotecología; Periodismo

### **A documentary and library science perspective about big data and data journalism**

*Antonio García-Jiménez and Beatriz Catalina-García*

#### **ABSTRACT**

The implementation of new systems for the creation, storage and dissemination of information is a key factor in the rapid growth of open data access through technical and computational mechanisms. Open data, big data, and social networks have transformed the work in mass media, driving the creation of a world of globalized information. This transformation also involves changes in the work of documentalists and journalists. Through a review of current research, including case studies, this paper analyzes from the standpoint of big data the processes, methods and tools oriented to media documentation. In addition to changes in professional routines, a set of contradictions between the semantic web and big data with regard to the objectives of transparency and understanding of information. Finally, from the documentary level, it is important to weigh the benefits reaped by broad data dissemination against those provided by filtering.

**Keywords:** Open data; Big data; Social Networking; Media; Documentation; Library and Information Science; Journalism

## INTRODUCCIÓN

Enmarcados en los cambios culturales y tecnocomunicativos a los que asistimos, este trabajo pretende detectar las consecuencias teóricas y prácticas para la bibliotecología y la documentación, en el marco de las prácticas periodísticas, de la expansión de técnicas, programas y procedimientos al amparo de los desarrollos derivados del *big data* y las redes sociales, ámbitos diferenciados pero claramente unidos.

Tomando el *big data* como eje fundamental del texto, debemos recordar que está presente en múltiples acciones de la vida cotidiana, como las vinculadas a la búsqueda de información, a la personalización de información y de la publicidad, al reconocimiento de patrones de información, al teclado predictivo, o al desarrollo de la investigación de dominios como el de las *smart cities*.

Tal y como apunta Fondevila Gascón (2013), la clave actual del ecosistema dominante reside en el flujo constante, y en aumento, de contenidos, que proceden de fuentes diversas, incluidos los medios de comunicación, los blogs, los *social media*, etc. En el mismo sentido, Lewis (2014) destaca el volumen y la variedad de la información digital producida, junto al crecimiento de la ubicuidad de los dispositivos móviles y al rápido avance y difusión de los procesos informáticos, caracterizados por las diferentes formas de aprendizaje de las máquinas y el uso continuado de algoritmos.

Esta realidad afecta directamente las tareas de los gobiernos, las empresas o los investigadores y reafirma la importancia de las operaciones de recopilación, análisis y representación de la información, asociadas con técnicas basadas en el cálculo y la cuantificación, y en donde adquieren mayor trascendencia la intersección entre medios de comunicación, tecnología y sociedad. Para entender este fenómeno, Lewis (2014) apunta la conveniencia de profundizar en el análisis de casos empíricos, tanto en el nivel micro (local) como en el nivel macro (institucional), en la generación de un marco conceptual para organizar, interpretar y teorizar esta cuestión y en la aplicación de perspectivas críticas que nos ayuden a interpretarla.

En este caso, situamos el punto de partida en la documentación o bibliotecología. Sugimoto, Ding y Thelwall (2012) afirman que ya en el marco de nuestro campo se ha intentando comprender esta realidad a partir de la *webometrics*, orientada a recopilar y analizar datos en la web en una gran escala en términos de desarrollo de software para descargar datos de la web y de métodos específicos para analizar la información (por ejemplo, análisis de enlaces). A esto se añade el papel de las teorías documentales centradas en la información. Otro de los esfuerzos documentales es el de la normalización de prácticas para representar e integrar datos, dentro de los que se puede

destacar la integración de la web semántica y los *Linked Open data*. Por último, y en su vinculación con los *social media*, sobresalen las herramientas, técnicas y teorías bibliométricas, que también cuentan con nuevas perspectivas como la *altmetrics*.

El objetivo de este trabajo es la descripción de lo que supone el *big data* para la documentación en el contexto de los medios de comunicación y del periodismo. Partiendo de la bibliografía sobre el particular, revisamos en primer lugar la conexión mediática con el *big data*. En segundo lugar, se abordan las diferentes formas y productos que adopta el periodismo ante este fenómeno, tomando en consideración también la perspectiva del *open data* y su conexión con la web semántica. Para finalizar, y con la mirada puesta en las redes sociales, tratamos algunos de los instrumentos documentales utilizados así como las aplicaciones que los periodistas están empleando, siempre primando el enfoque bibliotecológico y documental.

#### BIG DATA Y PERIODISMO

La conexión entre *big data* y periodismo ofrece diversos enfoques. A juicio de Peiró y Guallar (2013), el periodismo de datos está básicamente en relación con el periodismo de investigación, el periodismo de precisión, la documentación, la infografía, el *fact checking* y el *open data* y el *big data*.

Concretamente el periodismo de datos nacería de la aplicación del *open data* y del *big data*. Por un lado, al hablar de datos en abierto se conectaría con todas las opciones que existen para la recopilación, tratamiento, reutilización y difusión de los datos sin restricciones. En efecto, esta idea estaría relacionada con todos aquellos que apoyan el libre acceso al conocimiento, a los programas informáticos, etc. Por otro lado, la idea de *big data* tiene que ver con los procedimientos, técnicas y herramientas dispuestas para gestionar grandes cantidades de datos. En cualquier caso, el dato siempre necesitará una historia que contar, un contexto que facilite la comprensión, y profesionales que realicen esta conjunción.

Rodríguez Brito y García Chico (2013) enfatizan la relevancia que tiene en la actualidad la filosofía del *open data*, fundamentalmente en lo que se refiere a la transparencia en el acceso, desde la óptica profesional del periodista. Nos situamos en un momento en el que surgen también nuevas tareas dado el elevado nivel de información que debe ser filtrado gracias a herramientas estadísticas y de visualización.

En definitiva, es esta una etapa marcada por figuras como Adrian Holovaty, con su proyecto *Chicago Crime*, que se convirtió en uno de los primeros

casos de proyección de la implicación de bases de datos públicas (Flores y Salinas, 2012). De este tipo de experiencias se extrae la importancia del uso de base de datos, así como su aplicación a la visualización de información para mejorar el acceso a la información periodística.

Algunos casos fueron:

- El diario argentino *La Nación*, en el que un equipo de periodistas construyó una base de datos (con alrededor de 285 000 documentos) sobre los subsidios a colectivos y compañías desde el año 2006. La clave se situó en la disposición pública de datos, la sistematización del tratamiento de la información, y la proyección y entrecruzamiento con estadísticas oficiales, dando todo este esfuerzo como resultado un trabajo de interés periodístico.
- En el caso de la publicación brasileña *Gazeta do Povo*, la aplicación “Retratos Paraná” hizo posible el acceso a estadísticas de las ciudades del Estado. Fundamentalmente se difundió información sobre el desarrollo del Estado a partir de indicadores sobre la sociedad, la economía, la política, el medio ambiente, la educación y cultura. A esto se añadió una serie de reportajes que complementaron el material.
- El ejemplo del “Government Employee Salary Database” del *Texas Tribune*, al poner a disposición del público el salario de los empleados públicos del Estado, lo que permitió consultar, a partir del nombre, la agencia empleadora o la función.
- Por su parte, el *Chicago Tribune* publicó en 2011 el “Illinois School Report Cards”, que tenía como base el análisis de las escuelas públicas en Illinois, con un claro componente visual que permite conocer cada uno de los centros. También contó con el apoyo de los miembros del equipo de educación del diario.
- La propuesta “Terrorists for the FBI”, desarrollada por el Programa de Periodismo de Investigación de la Universidad de Berkeley-California, y la revista *Mother Jones*, que intentaba profundizar en el papel del FBI en las redes sociales. Con este trabajo, como gran artículo de investigación, se pudieron encontrar patrones de comportamiento.

Más recientemente, Stone (2014) recopila algunos casos de aplicación del *big data* en medios como *The Huffington Post*, *Financial Times*, *Sacramento Bee*, *BBC*, *CNN*, con tecnologías como Cxense, Neodata, Lotame, Jaspersoft, Gígya y Tableau. En resumen, y junto a los diferentes formatos de visualización, se empleó para:

- Medición del éxito de los titulares y la adaptación a las audiencias, también en función del dispositivo empleado. Destacan el análisis de comentarios (a partir de técnicas estadísticas), el análisis de calidad de productos y servicios, el análisis de los intereses de la audiencia conforme al comportamiento de navegación de los usuarios, la aplicación de métricas para mejorar el trabajo de los periodistas (con datos sobre usuarios, páginas fuentes, comportamiento, etc.), el estudio sobre la preferencia de contenidos y la delimitación de temas para publicación.
- Personalización de los contenidos periodísticos (también los anuncios, en función de los intereses de los lectores), la segmentación de la audiencia en general, o bien conforme a grupos concretos como los formados por los usuarios influyentes o los grandes consumidores de noticias.
- Seguimiento de noticias de última hora, detección de uso de tecnología y de consumo de noticias en tiempo real. Mediante técnicas de minería de datos, y tomando en consideración los tuits, por ejemplo, se genera un sistema de alerta temprana de noticias. A través de determinados algoritmos se identifican, clasifican y alertan a los clientes de información clave en tiempo real. También se emplea para la verificación de información previa a su publicación.
- Generación de noticias, gracias a técnicas que permiten condensar/destilar un conjunto de datos y presentarlo en un formato de noticias de periodismo de datos. Debemos recordar que determinados sistemas pueden combinar información tan diferenciada como pueden ser los datos estructurados en forma de base de datos, junto a tuits y datos sobre audiencias. Incluso hay herramientas, como Woch.it, que permite la creación de videos “sobre la marcha”.

#### DEL PERIODISMO DE DATOS AL PERIODISMO COMPUTACIONAL

Al enfrentarnos a la idea de periodismo de datos, se advierte cierta polémica que puede tener consecuencias prácticas y teóricas. Autores como Stavelin (2013) prefieren la idea de “periodismo computacional”: un periodismo que emerge de la tradición de periodismo *software-oriented* y que se cimenta en la mezcla de informática y de periodismo. De todas formas, este tipo de periodismo suele tratar con diferentes tipos de datos públicos como los registros, los presupuestos, datos de censos y también con información privada como los tuits y los videos que circulan en los *social media*, sin olvidar los registros de transacciones o los datos filtrados (como los empleados en el Wikileaks).

En los últimos años, podemos encontrar diferentes tipos y denominaciones de periodismo vinculados a esta esfera. El periodismo de precisión hace hincapié en el uso de métodos científicos, el *computer-assisted reporting* (CAR) enfatiza el uso de herramientas digitales, el periodismo de base de datos pone en valor la importancia del almacenamiento y la recuperación de información, el periodismo de datos y el *data-driven* hace hincapié en la búsqueda de historias a partir de conjuntos de datos, mientras que el periodismo computacional enfatiza la fusión de los valores informáticos y periodísticos tanto en la aplicación cotidiana como en la creación de herramientas (Stavelin, 2013).

Independientemente de la denominación, una sociedad en la que aumenta la información y en la que crece la necesidad de transparencia se postula como el marco ideal para que los periodistas utilicen y expliquen los datos, aunque sean complejos, por técnicas como la visualización. En el momento actual, en muchos países se está expandiendo el término *periodismo de datos*, que sugiere una forma especializada de recogida y análisis de datos en la línea del trabajo realizado por analistas o investigadores, en donde el componente informático es una parte central de la elaboración, en ocasiones automatizada, de la producción de noticias.

Se trata de un fenómeno complejo, en el que el periodista de datos computacional se debe desenvolver tomando en consideración lo siguiente (Ferrerías Rodríguez, 2012):

- Recopilación de datos e información, trabajando con fuentes informativas (públicas y privadas) en clara relación con el *big data* y el *open government*. Aparte del acceso a bases de datos abiertas al público, o mediante petición de información, se emplean técnicas de *scrapping* para obtener información de sitios web (*Needlebase*, *OutWit Hub* o *Scraperwiki*).
- Limpieza y filtrado de información, por ejemplo, detectando errores o bien observando la calidad de la información. Una de las herramientas que se puede utilizar es Google Refine, o incluso Excel, que permite elaborar porcentajes o patrones.
- Contextualización y combinación, tanto en el sentido del procesamiento y la recolección de los datos como en lo que se refiere al cruce con otras posibles fuentes de información (por ejemplo, bases de datos), lo que también permite delimitar su relevancia. Sería el caso de la combinación de datos con diferentes formatos de geolocalización.
- Comunicación. La exposición de los resultados muchas veces descansa en la visualización de los datos mediante tablas, gráficos, mapas, etc. Esto se puede hacer con todo tipo de herramientas y aplicaciones (Tableau, CartoDB, Google Fusion, Many Eyes, etc.).

Se trata de acciones concretas que conducen a productos concretos y que, según Peiró y Guallar (2013), son:

- Artículos con datos. Parece evidente que un texto de calidad debe aportar datos. La novedad se encuentra en la proliferación de este tipo de artículos, destacando los que se derivan de grandes volúmenes de información en abierto. Este fenómeno se redondearía con la aportación de información derivada de una investigación periodística.
- Infografía, visualizaciones y artículos multimedia. En el marco de la prensa digital, y con la irrupción del *big data* y los *open data*, se ha conseguido que muchos de estos gráficos integren o se basen en grandes cantidades de datos, ofreciendo la información de forma visual, atractiva e interactiva.
- Conjuntos de datos en abierto. Se trata de la publicación, en modo abierto, de los datos con vistas a que los usuarios-lectores puedan no sólo visualizarlos sino gestionarlos. En cualquier caso, debe respetar su accesibilidad y la capacidad de reutilización. En algunas ocasiones se ofrece el conjunto de los datos que ha servido de base para las noticias publicadas, en otros momentos se hacen accesible de forma independiente. Otra opción es la elaboración de bases de datos con capacidad de consulta por parte de los lectores. Esto supone generalmente el trabajo colaborativo entre periodistas, documentalistas, programadores e infografistas.

En este contexto, también se deben tomar en consideración las dinámicas que se proponen dinamizar los contenidos que emergen globalmente. La rapidez exigida por los lectores se ve favorecida por lenguajes informáticos y por las páginas de datos en bruto, que permiten nuevas vías de distribución de los contenidos. Así, el agregador de información ciudadana y noticias *EveryBlock 13* se organiza geográficamente y explota el concepto de *mashup*, esto es, una aplicación que utiliza contenidos digitales.

#### LA VISIÓN DESDE EL *OPEN DATA* Y LA WEB SEMÁNTICA

Los datos abiertos se denominan así porque son de libre acceso y pueden ser utilizados, reutilizados y distribuidos por cualquier persona con la única obligación de presentar la atribución de autoría y sabiendo que pueden compartirse sin restricciones tecnológicas en la línea de la Open Knowledge Definition. En el fondo, subyace la exigencia del mayor nivel de autodescripción

posible, siempre pensando en los más diversos contextos de uso y de aplicaciones. Las iniciativas, en este sentido, se vinculan al movimiento de *Open Data* (datos abiertos) y *Open Gov Data* (datos gubernamentales abiertos o datos públicos), en el marco del *Open Government* (gobierno abierto) (Chiaretti, 2013; Hernández Pérez y García Moreno, 2013).

Partiendo de una escala del propio Berners-Lee (2006), se hablaría de diversas opciones: a) datos en abierto en cualquier formato, b) datos abiertos estructurados, c) datos abiertos estructurados en un formato que no sea propietario, d) datos que implican el uso de estándares abiertos en el marco del W3C (Rdf y Sparql), que permiten la identificación de cosas y propiedades, e) la vinculación de los datos con información que le sirva de contexto.

En este sentido, la clave se sitúa en los procedimientos de representación-descripción de la información y el conocimiento, así como de las relaciones entre ellas, fundamentalmente con capacidad para que sean comprensibles desde un punto de vista informático. Esto nos retrotrae no sólo a los formalismos para la descripción de entidades, sino también a instrumentos como las ontologías.

Ferrer-Sapena y Sánchez-Pérez (2013), en el contexto del *open data* y *big data*, apuntan las relaciones entre periodismo de datos y web semántica. La apertura de datos cambia el modo de hacer periodismo, sabiendo que la accesibilidad de la información también modifica el rol del ciudadano y del propio periodista. A esto se suma que no toda la información está en la web, lo que supone múltiples modos de metainformación que requieren diversos procedimientos para su representación y reutilización.

No obstante, la posibilidad de compartir datos supone un nuevo reto en las tareas básicas de selección, análisis y difusión que realiza el documentalista. En este contexto, Peset, Ferrer-Sapena y Subirats-Coll (2011) plantean el *open data* como una eficaz vía para detectar duplicados documentales, eliminar las posibles ambigüedades terminológicas y suministrar la información y datos a usuarios de otras especialidades profesionales, como es el caso particular de los periodistas.

Un posible marco de trabajo viene establecido por Heravi y McGinnis (2013) al mencionar el social *semantic journalism*, con el ánimo de presentar soluciones ante el reto que supone para las instituciones mediáticas las grandes cantidades de los contenidos generados por los usuarios (UGC) frente al tiempo, limitado, que tienen los periodistas para extraer noticias derivadas de estos datos no estructurados, no filtrados y no verificados. De igual modo, se plantea para superar el obstáculo en el proceso de selección de la información, puesto que su realización manual puede ser ineficaz por la gran cantidad de datos.

Las tecnologías de la web semántica son un medio para proporcionar datos con estructura legible y facilitan también la integración de la información a partir de varias fuentes que son construidas usando las mismas tecnologías subyacentes. En este sentido, la web semántica sería la base sobre la cual construir la interoperabilidad de las plataformas de los medios sociales, proporcionando normas para “soportar” el intercambio de datos. El objetivo es generar una red de conocimiento interligado y semánticamente enriquecido, reuniendo las aplicaciones de la web social y los lenguajes y formatos de representación del conocimiento de la web semántica. Con este fin, López Borrull y Canals (2013) contextualizan estos lenguajes y formas cognoscitivas en la extraordinaria importancia que entraña para las ciencias experimentales la colaboración en la recogida de datos, probablemente más que para su análisis documental y su difusión que implica, respectivamente, el tratamiento de datos y la explotación de resultados.

En este caso, la clave está en ontologías como SIOC (Semantically Interlinked Online Communities) o FOAF (Friend Of a Friend). SIOC permite nuevos tipos de escenarios de uso para los datos *online* generados por usuarios (o en *community sites*) así como nuevas aplicaciones semánticas. Por su parte, FOAF es una ontología que describe a la gente y los vínculos entre ellas. Además, también interesan otras normas como rNews, que proporciona marcas semánticas para la anotación de noticias para los documentos web.

Con afán aclaratorio y organizador, traemos a colación la propuesta de Heravi y McGinnis (2013), que hace referencia a un marco de trabajo con cuatro fases:

1) Fase del descubrimiento del contenido. Aquí se trabaja con el contenido bruto de los medios sociales para enriquecerlos con metadatos semánticos, lo que permite el uso posterior.

- a) En primer lugar, se recoge una muestra representativa de los datos (contenidos de los mensajes, contestaciones, las menciones de otros usuarios, el URI del post, los enlaces que aparecen en los mensajes, la fecha y hora del mensaje y cualquier información de ubicación explícita (etiquetas geo) adjunta en el mensaje). En segundo lugar, se realiza la extracción de la entidad y su correspondiente anotación semántica (lo que implica la construcción de una herramienta) con vistas a relacionarla con metadatos semánticos relevantes, mediante coincidencias, a partir los *Linked Open data* para la posterior indexación y la búsqueda.

- b) Detección de eventos. Se identifican las entidades semánticas (“la reina abriendo el parlamento”, “coche bomba”, “Manchester United vs Liverpool”) a partir, por ejemplo, de la aceleración de frecuencia de algunos términos con el tiempo.
- c) Detección de ruido. Es la detección y filtrado de contenido no relevante (o con ruido) en los flujos de información donde un tema ha sido ya identificado. El enfoque empleado oscila el completamente automático, sin intervención del usuario, al semiautomático, donde se solicita una implicación limitada del usuario final.
- d) Detección Burst. Es el descubrimiento de incrementos súbitos en la frecuencia de un tema y/o en la ubicación de eventos específicos, con el fin de identificar cuando un hecho particular o noticia se convierte en *trending*.

2) Filtrado y contextualización. Refina los metadatos derivados de la fase de descubrimiento, con capacidad para contextualizar las noticias asociadas, incorporando un rastro de procedencia. Los procedimientos son:

- a) Contextualización. Descubre los antecedentes y la información contextual para una noticia mediante los metadatos creados durante la etapa de descubrimiento de contenido. Esto se puede lograr utilizando fuentes de *Linked Open data*, archivos y repositorios, y aplicando la desambiguación de temas y consultas basadas en la similitud para la recuperación de contexto.
- b) Procedencia de construcción. Se logra mediante la combinación de técnicas de difusión de información. Esto produce una traza de procedencia y el gráfico para ser utilizado para la fase de comprobación de confianza.
- c) Usuario y geolocalización del contenido. Conecta la posición del usuario y el contenido generado. Esta combinación se realiza con la explotación de coordenadas espaciales concretas, a través de desambiguación semántica, uso de la anotación y la toma de datos sociales del gráfico.
- d) Análisis de comunidad. Se centra en los usuarios que generan contenidos, determinando las comunidades que existen, sus características y la fuerza de conexión entre usuarios. La elaboración de gráficos es pertinente en este apartado.

3) Confianza y verificación. Utiliza los datos y los conceptos extraídos de etapas anteriores para delimitar la veracidad de la información. Por un lado, se

analiza la procedencia (por ejemplo, mediante la identificación de fuentes empleadas, de testigos oculares, etc.), el punto de vista (con indicadores sobre la perspectiva a través de la minería de opinión, o la detección de la comunidad de usuarios), y la veracidad (a través de análisis estadístico de contenido y los métodos cuantitativos digitales como filtros preliminares).

#### 4) Publicación, junto al archivo de las noticias producidas.

Como se puede observar, es una propuesta de marco de trabajo automatizado en el entorno periodístico, pero que se puede tomar como punto de partida interesante en la intersección entre web semántica y *big data*, también desde una perspectiva manual.

### LA VISIÓN DESDE LAS REDES SOCIALES

El crecimiento en el uso de Internet y la aparición de las redes sociales ha provocado que en diferentes casos la información llegue antes al ámbito del periodismo ciudadano que a los grandes medios de comunicación. Marcos Recio (2014) plantea una relegación de la primicia porque, cada vez con más frecuencia, la noticia es captada antes por el lector a través de redes sociales tan populares como Twitter y Facebook. Incluso esta última genera, a través de FB Newswire, su propio archivo de informaciones al que recurren frecuentemente los medios de comunicación, lo que ha convertido a esta red social en una fuente primaria de noticias (Mitchell, 2014).

Esta situación no ha surgido de forma abrupta. Desde el 2007, cuando comenzó a popularizarse la web 2.0, se ha ido produciendo la transformación, si bien vertiginosamente pero de forma paulatina y aún, hoy en día, se sigue desarrollando. Los factores de este cambio han sido diversos aunque la mayoría de ellos se deben a una modificación de las actitudes en la audiencia. La infoxicación, término acuñado por Cornellá (2000) para referirse a la sobrea-bundancia de información generada por medio de las nuevas tecnologías, no sólo ha traído consigo inconvenientes a la hora de seleccionar las noticias de acuerdo a su calidad y su veracidad, también ha conllevado una madurez en el criterio del lector con una mayor capacidad de discernimiento al poder cotejar informaciones enfocadas desde diversos puntos de vista, más globalizadas y, en ocasiones, con posibilidad de recurrir a la fuente directa de la noticia.

En este contexto, el periodista ya no sólo es redactor de las noticias, sino también su distribuidor y copartícipe en potenciales tertulias que se pueden generar a raíz de la información expuesta en las redes. El documentalista,

por su parte, se ha visto obligado a abrir su círculo relacional: ya no trabaja únicamente para y con el equipo de redacción de cada medio, sino que, como señala Guallar (2012), debe mantener un contacto directo con los usuarios; una novedosa faceta que es posible gracias a la proliferación de las redes. A todo ello se añade una profunda alteración de las rutinas laborales que supone la constante actualización de las informaciones; una tarea derivada de los cambiantes contenidos en la web y en las redes sociales, cada vez más acelerados y globalizadores.

Pero estos no son los únicos cambios en los que se han visto involucrados. De hecho, los periodistas se están convirtiendo en documentalistas y los documentalistas en periodistas. Se ha producido un proceso simbiótico –que no de fagocitación– por el cual se genera tal cruce entre ambas profesiones que, en ocasiones, resulta prácticamente imposible situar los límites. Para Marcos Recio (2014) esta cuasifusión de tareas es prácticamente obligada si se pretende seguir el ritmo que imprimen las nuevas tecnologías y su cada vez mayor adopción por parte de las audiencias. El periodista debe saber buscar a través de Internet y, en particular, de las redes sociales para convertirlas en una de sus fuentes de referencia; mientras, el documentalista debe aprender a redactar noticias que sirvan como apoyo en el medio digital.

En definitiva, ya queda en el pasado la etapa en la que el documentalista priorizaba controlar la veracidad de las informaciones a través de archivos (Paul, 2002) y relegaba una cierta predisposición a formar parte real de la elaboración informativa. No obstante, hay tareas que aún mantienen: Rubio-Lacoba (2005) asegura que siguen consultando sus propias fuentes y archivos aunque, de manera simultánea, también recurren a la web, tanto a páginas oficiales como a redes sociales. Esta fusión en las formas de acceso supone, en algunos casos, la recuperación de contenidos con una única base subjetiva, intuitiva y algo arbitraria: fundamentada en la experiencia y la profesionalidad pero también en las posibilidades casuales que aportan los distintos buscadores *online*. El estudio de Micó-Sanz, Masip-Masip y García-Avilés (2009) refuerza este modo de recuperación de información por parte de los periodistas, quienes reconocen basarse en la aleatoriedad aportada por el entorno digital. Este hecho induce a pensar en una precaria especialización en la búsqueda de datos *online*; una tarea que actualmente se solventa en distintos medios de comunicación por los documentalistas en su calidad de formadores.

En este contexto, las redes sociales, al mismo tiempo que alimentan los algoritmos del *big data*, nutren al periodismo en tanto que potencial fuente de información. De hecho, las posibilidades que entrañan Internet y las redes sociales aportan, en opinión de Rubio-Lacoba (2005), una función

verificadora o rectificadora en escasos minutos de la cual los periodistas, en su función de autodocumentalistas, han sabido aprovecharse y cada vez con mayor profusión.

Sin embargo, estas nuevas rutinas periodísticas no ponen en riesgo el futuro del documentalista en los medios de comunicación. Arranz y Caldera (2014) concluyen que se está produciendo el efecto contrario con la asunción de responsabilidades propias del periodista y con el mantenimiento de las funciones tradicionales de la documentación; tareas que se suman a las ya descritas tiempo atrás por Paul (2002) de consultoría, asesoramiento y formación, actualizadas en los últimos años sobre la base de un entorno digital, cada vez más recurrente en la búsqueda y recuperación de datos.

La encuesta realizada por Oriella PR Network (Oriella Digital Journalism, 2013) a 533 periodistas pertenecientes a 15 países de todo el mundo registra que, para confirmar informaciones, el 42 % de ellos confía en los microblogs, principalmente Twitter, Facebook y Weibo; el 37 % en blogs cuyos autores les son conocidos, una práctica que para Paulussen y Harder (2014) es recomendable. Estas cifras aumentan entre un 54 y un 68 % en países como Reino Unido, Canadá, China y Francia. Si bien la gran mayoría de los encuestados reconoce que sigue empleando los canales *offline* para verificar o desmentir las informaciones, se evidencia un gran salto hacia las nuevas tecnologías para la autodocumentación.

Pero los periodistas no utilizan las redes sociales sólo para contrastar las informaciones a las que tienen acceso; de ellas se sirven además para encontrar nuevas historias. Una práctica cada vez más extendida tal y como registra el mencionado estudio de Oriella: más de la mitad de los profesionales utilizan publicaciones de los principales microblogs para recrearlas en sus propios medios. Unas conclusiones similares alcanzaron Hermida (2010) y Diakopoulos, De Choudhury y Naaman (2012), quienes vieron en el uso de los medios sociales una práctica periodística para la captación de informaciones cada vez más extendida.

En este contexto, y aunque no forma parte del objetivo central de este trabajo, se observa también una búsqueda de fuentes en el periodismo ciudadano *online*, principalmente cuando al profesional le resulta difícil cubrir una historia por diferentes circunstancias, por ejemplo, cuando se producen atentados o catástrofes en lugares lejanos a los que es imposible llegar y ofrecer la información en cuestión de minutos como posibilitan las nuevas tecnologías. A ello se añade un cambio en las rutinas profesionales que desembocan, según Paulussen y Harder (2014), en la navegación diaria de los periodistas por las redes sociales en la búsqueda de informaciones o para no quedarse rezagados en la difusión de los hechos. De hecho, un estudio de

Pew Research (2012) concluye que el 39 % de los videos sobre acontecimientos inesperados que fueron difundidos por medios de comunicación habían sido grabados por los ciudadanos.

De la importancia de las redes sociales, ya entendidas como *social media*, hablan claramente ejemplos como el “Blogs & Columns” de *The Washington Post*, “Un País de Blogs” de *El País*, “From The Archive” de *The Guardian*, que demuestran la adaptación de los medios de comunicación a la web, ofreciendo al usuario una versión documental de los acontecimientos noticiosos que van marcando la historia desde los diferentes enfoques. Pero la incorporación de las nuevas tecnologías e Internet no se queda aquí; también es significativa la apertura de nuevos perfiles en las principales redes sociales, tanto la propia marca de los medios como periodistas a título individual. En el citado informe de Oriella PR Network, más de la mitad de los periodistas encuestados tienen, al menos, un perfil en una red social, preferentemente en Twitter, y un tercio de ellos ha abierto un blog.

#### ALGUNOS INSTRUMENTOS DOCUMENTALES

Son muchos los instrumentos que están surgiendo y que tienen implicaciones y vinculaciones documentales. Destacamos algunos ejemplos. Mena Muñoz (2014) apunta los *content curators*, que podríamos traducir como selectores de contenidos. Se trataría de instrumentos y procesos que llevan la información filtrada a aquella persona que lo necesite tras un proceso de búsqueda, agrupación y organización, con una presentación que sea también proactiva.

En niveles internos, también destacan la puesta en práctica de nuevos lenguajes, como el caso de *El País*, descrito por Rubio-Lacoba (2012). Un lenguaje colaborativo que se puede definir como una folksonomía controlada, cuya base es la indexación social y por objeto la indización/categorización de las noticias. Destaca que la acción clasificatoria sea colaborativa, realizada por periodistas, con filtros efectuados por expertos con vistas a evitar la sinonimia y la polisemia, y enriqueciendo esta herramienta mediante relaciones de “parentesco”.

Basándose en experiencias como la de *The New York Times*, *Delicious o Soitu*, el etiquetado se ha comprobado como una forma de acceso a la información por parte de los usuarios. Además, cuenta como fuentes la base de datos, el IPTC utilizado en la versión digital, diferentes tesauros especializados en información de actualidad, así como técnicas SEO para la selección de términos. Un lenguaje que está estructurado en temas, personajes, organizaciones, lugares y eventos y que en 2012 contaba con 74 000 etiquetas (63 % para los personajes y el 6 % para temas).

Su dinámica de trabajo es la siguiente: el etiquetado se realiza a partir de la acción de cada periodista antes de la publicación. A continuación, se genera una “portadilla dinámica”, vinculando la noticia a temas con los que mantenga algún tipo de conexión. Estos temas surgen de forma automática tomando en consideración la pertinencia de los contenidos, la actualidad y el nivel de “parentesco”. Por su parte, el editor automático ofrece las siguientes opciones: a) autocompletado de etiquetas, b) desambiguación fundamentalmente de siglas, cargos y fechas, c) la inferencia de etiquetas por la que el sistema ofrece algunas etiquetas a partir de la comparación con noticias que se entienden semejantes, d) la solicitud de etiquetas por parte de los periodistas, ya que tienen la opción de pedir la aceptación de una etiqueta y que o bien se añadirá o bien se derivará a sinónimos aceptados, e) el sistema de herencia, por el que las etiquetas quedan asociadas unas a otras de modo automático, y f) las etiquetas combinadas, ante las peticiones de “conceptos editoriales” de los periodistas, ya que no se pueden quedar en el lenguaje principal.

#### APLICACIONES EN LAS SALAS DE REDACCIÓN

El *big data* supone hablar de la gestión eficaz y automatizada de volúmenes ingentes de datos. Además, el tiempo de procesamiento y respuesta debe ser muy reducido; estaríamos hablando de tiempo real. A esto se suma la variedad de las fuentes y de los tipos de datos, también no estructurados (textos, mensajes en redes sociales), en constante actualización, y que deben ser organizados. En este sentido, el *big data* exige flexibilidad puesto que debe adaptarse a múltiples formatos. Desde un plano documental (Serrano-Cobos, 2014) nos encontraríamos con datos de navegación (análisis web) y datos que existen en Internet (cibernetría) alrededor de nichos de mercado, comportamientos de búsqueda, segmentación de clientes o potenciales clientes, relaciones entre marcas, la conexión entre servicios y productos y sus usuarios, las relaciones con competidores, la vigilancia tecnológica, etc.

Por otra parte, si nos fijamos en alguno de los cursos existentes (<http://periodismodatos.okfn.es/talleres-2014/#talleres-madrid>), se podría concluir que para introducirse en el periodismo de datos y en el *open data* hay que saber gestionar, siempre sabiendo que el propio periodista debe ser capaz de crear aplicaciones. Algunos aspectos a considerar se mencionan a continuación:

- *Web scraping*. Técnica por la que se extrae información de la web y que se vincula con fórmulas de tratamiento y reconocimiento de información semántica. En este sentido, interesan las herramientas ETL

(*extract, transform and load*) dedicadas a la extracción, la transformación y cargado de datos; se manejan en entornos de grandes volúmenes de datos a partir de diferentes fuentes y están destinadas al refinado y difusión de nuevos documentos.

- Herramientas de refinamiento (como Open Refine o Google Refine) que permiten entender la estructura y calidad de unos datos, corrigiendo los tipos de errores comunes en ellos.
- Hojas de cálculo, como opción para la gestión y organización de la información estructurada.
- Programación con lenguajes de programación como JavaScript y lenguajes de marcado como HTML o CSS (hoja de estilo en cascada).
- Prácticas y herramientas de infografía y visualización. Mejoran el acceso a los usuarios a los datos organizados. Uno de los ejemplos de esta clase de utilidades es el de la generación de mapas interactivos.
- Prácticas y herramientas para la comunicación visual; ya sea en forma de videos y animaciones.
- Elaboración de noticias teniendo en cuenta datos georeferenciados, que hacen posible la localización de personas, instituciones, objetos o eventos en mapas.

Otra forma de analizar esta faceta es delimitar el plano documental en las nuevas aplicaciones que surgen. Siguiendo a Marshall (2013), de las aplicaciones destinadas a los periodistas algunas tienen un claro componente documental:

- Herramientas narrativas: curación de contenidos, creación de visualizaciones y creación de gráficos interactivos, enriquecimiento de imágenes con enlaces contextuales y conversión de hojas de cálculo en *timelines* interactivas.
- Herramientas de búsqueda en medios de comunicación, en redes sociales, de personas, con capacidad de respuesta a partir de datos estructurados y refinados, detección de fotos falsas.
- Herramientas de productividad: servicio de alertas de palabras clave en Google.

Otras aplicaciones puede filtrar noticias (Neatly), se pueden emplear como bloc de notas (Evernote), o bien generan diversas publicaciones gracias a los contactos y los medios en las redes (Glipboard).

Este nuevo contexto permite entender el periodismo como una forma de transparencia, como forma de participación y como forma de experimentación (Lewis y Usher, 2013). Como se ha comentado con anterioridad, la investigación y reflexión sobre qué instrumentos y procedimientos no sólo periodísticos sino también documentales se están utilizando. En el trabajo de Parasio y Dagiral (2013) se habla de la aceptación de la implicación tecnológica y automatizada por parte de los periodistas, si bien no todos se enfrentan del mismo modo al movimiento *open data*. También se observa una clara controversia sobre el valor periodístico real que tienen los datos, frente al valor que supone filtrar la información, interpretarla; a lo que se suma el recelo que genera la capacidad de manipulación de los datos por parte de las autoridades políticas. Por otro lado, se pone en tela de juicio la capacidad de los sistemas automáticos y de *big data* para identificar los asuntos públicos a pesar de que las aplicaciones *data-driven* se orientan para ofrecer noticias concretas a lectores específicos.

En cualquier caso, sigue existiendo un fuerte componente de la búsqueda de información en el marco de la información digital (Kemman *et al.*, 2013). Junto a esto, aparece paulatinamente la necesidad de poner en funcionamiento aplicaciones de noticias como sistemas continuos, normalmente gracias al uso de API en su faceta de fuentes de datos. La idea clave es la igualdad de acceso a los datos entre los periodistas y el público. Asimismo, parte de la información que gestionan los periodistas no puede ser gestionada automáticamente. Pero algunas funciones, como la comprobación de hechos, muestran un camino interesante desde este punto de vista. También la exploración de nuevos escenarios y posibilidades se destaca como elemento relevante (Stavelin, 2013).

## CONCLUSIONES Y LIMITACIONES

La aceleración de los cambios informativo-computacionales que se producen no ya sólo en la sociedad, sino en la práctica periodística, hace difícil el análisis de la respuesta desde la documentación a los retos, incertidumbres y problemas que surgen. De hecho, a veces resulta una tarea ardua la distinción y clasificación entre los procedimientos, técnicas e instrumentos del dominio periodístico, informático y documental.

Esta investigación está en la línea de aquellas que invitan a una reflexión sobre el campo bibliotecológico. Desde hace años se viene observando, por lo menos en su vinculación comunicológica, un menor peso en las rutinas productivas periodísticas en favor de la perspectiva tecnoinformática. En

cualquier caso, se remarca el carácter profundamente transdisciplinar de la aproximación que requieren los fenómenos aquí estudiados. Este trabajo intenta comprender las prácticas documentales en el contexto del *big data* y del *open data*, una cuestión que se desarrolla siempre en el marco de una creciente automatización en diversos frentes.

En la selección los periodistas utilizan los datos, los blogs y las redes sociales no sólo para la verificación de información, sino también para la creación de noticias. Así, asistimos a la evolución del análisis y tratamiento de la información en sintonía con el surgimiento de nuevos instrumentos cuya utilización entraña importantes derivaciones documentales; entre ellas cabe reseñar cambios en la recopilación de datos para la elaboración de informaciones que implica, en la labor documental, una mejor contextualización de los temas tratados como noticias con el empleo de recursos que facilitan la interpretación. Como caso particular destacan los llamados *content-curators* y los lenguajes documentales a cuya construcción se incorporan técnicas SEO.

En la recuperación, el periodista cambia sus hábitos o refuerza sus tareas cotidianas en función de la valoración de los datos, como consecuencia del incremento de la demanda, por parte de las audiencias, a la hora de recibir noticias cada vez más transparentes y documentadas. Las redes sociales también han traído consigo un cambio de rutinas en los medios no sólo para el periodista, sino también para el documentalista. Es tan profundo el cambio de prácticas que, en la mayoría de las ocasiones, no se vislumbran límites entre las tareas periodística y documental, aunque ambas figuras siguen siendo operativas ya que en el uso del *big data* y del *open data* se exige una actualización permanente. Tarea desempeñada en mayor medida por el documentalista que, lejos de ser una figura en riesgo de desaparición dentro de los medios, puede ver incrementada sus funciones a partir de las necesidades del profesional del periodismo.

En la difusión, van en aumento los casos en los que las bases de datos públicas hacen posible un mejor y mayor acceso a la información periodística. Entre sus efectos se observa un crecimiento de la actividad investigadora y analítica desde diversos enfoques: el desarrollo de contenidos, el estudio de audiencias y el uso de nuevas tecnologías; a todo ello se añade la incorporación de nuevos modos de generación y almacenamiento de noticias. Todo ello implica un periodismo más participativo y experimental, además de una igualdad de acceso para el periodista y para el público a las diferentes bases de datos. La consecuencia directa es una más rápida y efectiva comprobación de los hechos susceptibles de ser noticia y la exploración de datos que generen nuevas informaciones.

Estas innovaciones no están exentas de obstáculos y controversias que actualmente no se han superado. Uno de los problemas fundamentales es el de transparencia, comprensión y evidencia aún existente en la relación que sería deseable establecer entre *open data* y web semántica. Es preciso, por otra parte, considerar también el valor periodístico real que entraña la difusión sin filtros de los datos frente a una interpretación de la información derivada del manejo de aquellos.

Igualmente es necesario tener en cuenta la interacción entre la web 3.0 y el *big data*. No es excepcional el hallazgo de informaciones sin estructura y de una ingente cantidad de datos almacenados arbitrariamente, hecho que niega la comunicación y comprensión que se plantea desde la web semántica, aunque no es menos cierto que la aplicación de *big data* está cada día más desarrollada y organizada en los medios de comunicación, con importantes ejemplos de aplicación que han trascendido no sólo al propio trabajo periodístico. En efecto, con este fenómeno se puede desplegar datos contextualizados en informaciones cada vez más completas y documentadas, al mismo tiempo que se satisface la necesidad informativa de unas audiencias cada vez más exigentes y a las que se responden de modo personalizado, constantemente actualizado y automatizado.

Este trabajo, de carácter teórico y descriptivo, debe ser continuado a partir de las incógnitas que hemos delimitado. De igual forma, tampoco se ha tratado la conexión con la analítica web, o con el periodismo ciudadano, y otro tema clave que son los propios riesgos que supone el fenómeno del *big data*.

## REFERENCIAS

- Arranz, P. y J. Caldera. 2014. "Funciones del documentalista en la televisión digital". *Em Questão* 19 (2): 145-158.
- Berners-Lee, T. 2006. *Linked Data-Design Issues*. <http://www.w3.org/DesignIssues/LinkedData.html>
- Chiaretti, A. 2013. "Datos abiertos (enlazados) y democratización del acceso a la información en Chile". *XIV Simposio de la Internacional del Conocimiento*, Chile. [http://eprints.rclis.org/18677/1/Chiaretti\\_Simposio14\\_2013.pdf](http://eprints.rclis.org/18677/1/Chiaretti_Simposio14_2013.pdf)
- Cornellá, A. 2000. "Cómo sobrevivir a la infoxicación". *Infonomía.com* 8.
- Diakopoulos, N., M. de Choudhury y M. Naaman. 2012, mayo. "Finding and assessing social media information sources in the context of journalism". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2451-2460. ACM. [http://www.munmund.net/pubs/chi\\_2012.pdf](http://www.munmund.net/pubs/chi_2012.pdf)
- Ferreras Rodríguez, E. M. 2012. "Nuevos perfiles profesionales: el Periodista de Datos". *Actas IV Congreso Internacional Latina de Comunicación Social*. [http://www.revistalatinacs.org/12SLCS/2012\\_actas/062\\_Ferreras.pdf](http://www.revistalatinacs.org/12SLCS/2012_actas/062_Ferreras.pdf)

- Ferrer-Sapena, A. y E. A. Sánchez-Pérez. 2013. “Open data, big data: ¿hacia dónde nos dirigimos?”. *Anuario ThinkEPI* 7, 150-156. <http://eprints.rclis.org/21006/1/150-156-Ferrer-Sanchez-Open-data-Big-data.pdf>
- Fondevila Gascón, J. F. 2013. “Periodismo ciudadano y *cloud journalism*: un flujo necesario en la Sociedad de la Banda Ancha”. *Comunicación y Hombre* 9: 25-41
- Flores, J. y C. Salinas. 2012. “Entornos y perfiles profesionales para redacciones con cultura de red: Mashups y Data Delivery Editor (Datajournalism)”. *Actas del III Congreso Internacional de la Asociación Española de Investigación de la comunicación*. [http://www.aeic2012tarragona.org/comunicacions\\_cd/ok/194.pdf](http://www.aeic2012tarragona.org/comunicacions_cd/ok/194.pdf)
- Guallar, J. 2012. “Documentalistas de medios y redes sociales”. *Anuario ThinkEPI* 6: 170-172. [http://eprints.rclis.org/17705/1/thinkepi2012\\_guallar\\_documentalistas%20medios.pdf](http://eprints.rclis.org/17705/1/thinkepi2012_guallar_documentalistas%20medios.pdf)
- Heravi, B. R. y J. McGinnis. 2013. “A Framework for Social Semantic Journalism”. *First International IFIP Working Conference on Value-Driven Social & Semantic Collective Intelligence (VaSCo), at ACM Web Science VaSCo@ WebSci*, 13-18. Francia, 1 de mayo. <https://pdfs.semanticscholar.org/94a6/4d-95976475ff278615f1d01b1e3f73393750.pdf>
- Hermida, A. 2010. “Twittering the News: The emergence of ambient journalism”. *Journalism Practice* 4 (3): 297-308. <http://www.tandfonline.com/doi/pdf/10.1080/17512781003640703>
- Hernández Pérez, T. y M. A. García Moreno. 2013. “Datos abiertos y repositorios de datos: nuevo reto para los bibliotecarios”. *El profesional de la información* 22 (3): 259-263.
- Kemman, M., M. Kleppe, B. Nieman y H. Beunders. 2013. “Dutch Journalism in the Digital Age”. *Icono14* 11 (2): 163-181.
- Lewis, S. C. 2014. “Journalism in an era of Big Data”. *Digital Journalism*, 1-10. <http://dx.doi.org/10.1080/21670811.2014.976399>
- Lewis S. C. y N. Usher. 2013. “Open source and journalism: toward new frameworks for imaginig news innovation”. *Media, Culture & Society* 35 (5): 602-619.
- López-Borrull, A. y A. Canals. 2013. “La colaboración científica en el marco de nuevas propuestas científicas: Open Science, e-Science y Big Data”. En *La colaboración científica: una aproximación multidisciplinar*, G. González Alcaide et al. (coords.), 91-101. Nau Llibres. Edicions Culturals Valencianes S.A. [http://eprints.rclis.org/20965/1/Lopez-Borrull\\_Canals\\_Col\\_Valencia.pdf](http://eprints.rclis.org/20965/1/Lopez-Borrull_Canals_Col_Valencia.pdf)
- Marcos Recio, J. C. 2014. “Otra forma de plantear la información y documentación en los medios digitales”. *Revista General de Información y Documentación* 24 (1). <http://revistas.ucm.es/index.php/RGID/article/viewFile/119-134/42751>
- Marshall, S. 2013. *20 tools and apps for digital journalists*. <https://www.journalism.co.uk/news/20-tools-and-apps-for-digital-journalists/s2/a554321/>
- Mena Muñoz, S. 2014. “Herramientas contra la infoxicación en los Social Media: los ‘Content Curators’”. *Ámbitos* 24. <http://ambitoscomunicacion.com/2014/herramientas-contra-la-infoxicacion-en-los-social-media-los-content-curators/>
- Micó-Sanz, J. L., P. Masip-Masip y J. A. García-Avilés. 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* 18 (2): 284-290. [http://eprints.rclis.org/13756/1/Masip\\_EPI.pdf](http://eprints.rclis.org/13756/1/Masip_EPI.pdf)

- Mitchell, A. 2014. *Announcing FB Newswire, Powered by Storyful*. <http://newsroom.fb.com/news/2014/04/announcing-fb-newswire-powered-bystoryful>
- Oriella Digital Journalism. 2013. *The New Normal for News. Have global media changes forever*. [http://www.oriellaprnetwork.com/sites/default/files/research/Brands2Life\\_ODJS\\_v4.pdf](http://www.oriellaprnetwork.com/sites/default/files/research/Brands2Life_ODJS_v4.pdf)
- Parasie, S. y E. Dagiral. 2013. "Data-driven journalism and the public good: 'Computer-assisted-reporters' and 'programmer-journalists' in Chicago". *New Media & Society* 15 (6): 853-871.
- Paul, N. 2002. "News libraries: forging a professional relationship with newsroom". *Poynter Institute Online*. <http://www.poynter.org/uncategorized/2157/news-libraries-forging-a-professional-relationship-with-the-newsroom/>
- Paulussen, S. y R. A. Harder. 2014. "Social Media References in Newspapers: Facebook, Twitter and YouTube as sources in newspaper journalism". *Journalism Practice* (ahead-of-print): 1-10. <http://www.tandfonline.com/doi/pdf/10.1080/17512786.2014.894327>
- Peiró, K. y J. Guallar. 2013. "Introducció al periodisme de dades. Característiques, panoràmica i exemples". *Item* 57: 23-36
- Peset, F., A. Ferrer-Sapena e I. Subirats-Coll. 2011. "Open data y linked open data: su impacto en el área de bibliotecas y documentación". *El profesional de la información* 20 (2): 165-173. <https://riunet.upv.es/bitstream/handle/10251/29837/Peset-Ferrer-Subirats2.pdf?sequence=1>
- Pew Research Center's Journalism Project Staff. 2012. "Youtube & News: A New Kind of Visual Journalism". *Pew Research Center's Journalism Project*. <http://www.journalism.org/2012/07/16/youtube-news/>
- Rodríguez Brito, A. y L. García Chico. 2013. "Cartografiando el periodismo de datos. Big Data y periodismo en el continente americano. Cinco casos de estudio". *Telos* 95: 57-67. <http://telos.fundaciontelefonica.com/url-direct/pdf-generator?-tipoContenido=articuloTelos&idContenido=2013062110110002&idioma=es>
- Rubio-Lacoba, M. 2012. "Nuevas destrezas documentales para periodistas: el vocabulario colaborativo del diario *El País*". *Trípodos* 31: 65-78.
- Rubio-Lacoba, M. 2005. "Nuevos tiempos para la documentación informativa en el periodismo digital: viejas y nuevas funciones del servicio de documentación digital". *Comunicación y sociedad* 28 (1): 153-168. <http://eprints.rclis.org/16327/3/articulocys.pdf>
- Stavelin, E. 2013. "Computational Journalism. When journalism meets programming". Thesis, University of Bergen.
- Stone, M. L. 2014. "Big data for Media". *Reuters Institute for the Study of Journalism*, 1-31. UK: University of Oxford. [http://www.bigdatamedia.org/wp-content/uploads/2017/03/Big-Data-For-Media\\_2014-Stone.pdf](http://www.bigdatamedia.org/wp-content/uploads/2017/03/Big-Data-For-Media_2014-Stone.pdf)
- Serrano-Cobos, Jorge. 2014. "Big data y analítica web. Estudiar las Corrientes en un océano de datos". *El profesional de la información* 23 (6): 561-565.
- Sugimoto, C. R., Y. Ding y M. Thelwall. 2012. "Library and Information Science in the Big Data Era: Funding, Projects, and Future [a panel proposal]". *ASIST 2012*, 26-31.

*Para citar este texto:*

García Jiménez, Antonio y Beatriz Catalina García. 2018. “Una perspectiva documental y bibliotecológica sobre el *big data* y el periodismo de datos”. *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 32 (74): 77-99.

<http://dx.doi.org/10.22201/iibi.24488321xe.2018.74.57910>

DOI: <http://dx.doi.org/10.22201/iibi.24488321xe.2018.74.57910>