

5. Periodismo de datos: diseños de la pieza informativa

Data journalism: design of the informative piece

Aída María de Vicente Domínguez
amdevicente@gmail.com
Universidad de Málaga

Resumen: La comunicación pretende mostrar ejemplos de piezas informativas que se han creado a partir de una ingente cantidad de datos. En concreto, se dan a conocer las muestras elaboradas por el Laboratorio de Innovación Audiovisual de RTVE. Tras definir que es el Periodismo de datos se explican las diversas piezas informativas que conforman esta sección en el Lab de RTVE. De cada pieza se expone: temática, procedencia de los datos y cómo los han extraído, analizado, procesado, interpretado y presentado de forma comprensible para el lector. El análisis muestra diversos modelos de producción y edición de esta especialización periodística que se considera de interés para profesionales de los *mass media* y para docentes que impartan este modelo comunicacional.

Palabras clave: Periodismo de datos, Lab de RTVE, Digital, Internet.

Abstract: The communication aims to show examples of news items that have been created with a lot of data. Specifically, have been selected some samples made by the RTVE.es Lab. After defining "Data Journalism", the study analyse of each news items grouped in this section: theme, source of the data, and how they have extracted, analyzed, processed, interpreted and displayed the information. The analysis shows various models of production and editing of this journalistic specialization interesting to journalist, as well as for teachers teach this communication model.

Key words: Data Journalism, RTVE Lab, Digital, Internet.

1. Introducción: Periodismo de datos y sociedad red

La plataforma digital ha brindado a los profesionales de la información innovadoras herramientas para la producción y edición de las piezas informativas. Así como, la posibilidad de ampliar sus fuentes. Por tanto, ha generado nuevas formas de hacer periodismo y una de ellas es el Periodismo de datos.

Un proceso de comunicación basado en usar una fuente específica: datos en bruto. Datos que deben ser interpretados, analizados y expuestos visualmente de forma comprensible para el lector. Una modalidad comunicacional para cuya elaboración Internet pone a disposición del profesional de los *mass media* no sólo una ingente cantidad de bases de datos creadas por instituciones públicas y privadas, también multitud de herramientas gratuitas para presentar la información como las destacadas por De Vicente (2011, p. 13):

Para la captura y análisis de los datos: Adobe Acrobat, Crowdsourcing, DocumentCloud, Google Refine, Firebug, Fusion Tables, Nokogiri, Overview, OutWit Hub, Panda Project, Ruby, ScraperWiki, y Tesseract. Herramientas digitales que han sido abordadas por Nguyen (2010) y Mazotte (2012).

Para mostrar visualmente la información: Axiis, Data-Driven Documents (d3), Many Eyes, NodeXL, Tableau Public, Visual.ly, o Weave. Recursos disponibles en la red que han sido tratados por Brashaw (2010) y Turrado (2012).

Aportaciones que son muy significativas para contar historias a través de esta especialidad periodística. Y es que el Periodismo de datos “consiste en usar herramientas estadísticas y de visualización” (Flores, 2012, p. 7) y se estructura en “tres elementos fundamentales: acceso a los datos, su tratamiento y su publicación; cuyo denominador común son las nuevas herramientas y tecnologías” (Ferreras, 2012, p. 6)

De acuerdo con Crucianelli (2012) el mundo digital ha incrementado “la capacidad para acercar grandes cantidades de datos (Big Data): nuevas herramientas de búsqueda, nuevas

herramientas de extracción, nuevas herramientas de análisis y nuevas herramientas de visualización”.

Asimismo, destacable es otra de sus particularidades. Los productos informativos pueden mostrarse mediante modalidades divergentes como infografías, cronologías, gráficos o visualizaciones interactivas, entre otras:

“Periodisme de dades és l’especialitat o pràctica del periodisme que utilitza com a matèria primera i fonamental dades (i no altres fonts d’informació), per tal de crear productes informatius diversos, com ara: articles amb dades, infografies o visualitzacions interactives amb dades i conjunts de dades en obert” (Peiró y Guallar, 2013, p. 26).

Por tanto, este artículo pretende mostrar cómo es el proceso de elaboración de las piezas informativas que pertenecen al Periodismo de datos. Y para ello se exploran las elaboradas por el Lab de RTVE, un laboratorio de innovación audiovisual dedicado íntegramente a investigar las posibilidades que ofrece la red para contar historias a través de formatos y narrativas divergentes. En concreto, se analizan las catorce piezas que conforman su sección “Periodismo datos” para abordar los siguientes objetivos:

- Mostrar como datos en bruto pueden convertirse en piezas informativas.
- Identificar el proceso de producción y edición.
- Conocer la procedencia de las fuentes de cada pieza informativa.
- Averiguar las técnicas de recolección de datos utilizadas.
- Indicar cómo estructuraron los datos en brutos.
- Identificar el tipo de visualización usada para cada información.

Un análisis significativo para comunicadores presentes y venideros porque:

“los profesionales de los cibermedios requieren una formación específica que les permita adquirir las competencias y habilidades propias del nuevo perfil de periodista que inaugura el periodismo on-line” (Tejedor, 2006, p. 237).

Asimismo, esta materia puede extrapolarse al campo académico: sus contenidos se consideran aptos para exponerlos en clases de comunicación donde impartan esta modalidad informativa. Y es que estas muestras son un excelente referente para abordar el Periodismo de datos a nivel nacional.

2. Metodología

La metodología se ha estructurado en los siguientes pasos:

- Revisión bibliográfica: se han consultado diversas bases de datos (Dialnet, Rebiun, B-on y Teseo) para acceder a las investigaciones precedentes realizadas sobre este campo de estudio.
- Selección del medio: se ha escogido el Lab de RTVE porque ha experimentado con esta especialización periodística desde sus inicios, y dichas experiencias se consideran demostrativas para abordar como se producen y editan estas piezas informativas. Un *Medialab* que ha sido galardonado con diversos premios nacionales e internacionales (bronce en los Premios Internacional *Malofiej* 2014, Mención de Honor en los *Webbie Awards* 2014 y en los *Lovie Awards*, o el Premio de Periodismo Digital José Manuel Porquet 2014). Innovaciones que se consideran significativas para mostrar una panorámica de las últimas tendencias en el sector periodístico a nivel nacional.

Selección de la muestra: desde la inauguración del Lab de RTVE hasta finales del 2015 su sección “Periodismo de datos” la conforman catorce piezas periodísticas que ayudan a entender de forma sencilla hechos o acontecimientos que generan un gran volumen de información estadística. Por tanto, se analiza la totalidad de las informaciones con el objeto de mostrar todas las experiencias realizadas en esta modalidad comunicativa.

Visionado individual de cada pieza para extraer el año, la fuente, la temática, la estructura de la información y el tipo de visualización (presentación).

Contacto con la periodista del Lab de RTVE Mírian Hernánz para conocer los procesos de extracción o recolección de los datos, así como, para confirmar los datos extraídos por la analista.

Indicar en una tabla Excel de cada información: año, temática, procedencia de los datos, técnica de recolección de los datos, estructura de la información, y tipo de presentación (visualización).

3. Resultados

3.1. Temática, fuente y extracción de los datos

El denominador común de todas estas piezas informativas es que sus fuentes ofrecen datos en bruto. Datos que, en la muestra analizada, proceden de fuentes oficiales nacionales (Ministerio del Interior, Organización Nacional de Transplantes, ACNUR, Instituto Nacional de Estadística, Boletín oficial del estado, etc), fuentes oficiales Internacionales (Population Reference Bureau, Eurostat, Organización de las Naciones Unidas para la alimentación y la agricultura, etc), plataformas (Verkami) y redes sociales (API- Twitter).

Otro denominador común es que los profesionales de la información han accedido a estas fuentes mediante Internet. En coincidencia con Ramonet (2011, p. 18):

“El Periodismo de [bases de] datos es otro tipo de periodismo que está prosperando entre los nuevos nichos de Internet. Se le considera un gran logro de la democracia moderna porque permite buscar y acceder, de forma inmediata, a bases de información creadas por instituciones públicas o privadas”.

Y si atendemos a las técnicas de recolección, los resultados indican que la compilación de datos se ha obtenido mediante procesos divergentes. En concreto, se han aplicado casi todas las posibles alternativas referenciadas por Steve y Nieto (2014) en *Nuevos retos del*

periodismo especializado, así como por Flores y Salinas (2013) en *El Periodismo de datos como especialización de las organizaciones de noticias en Internet*:

- Datos propuestos directamente por una organización.
- Datos encontrados usando técnicas de búsqueda avanzada para surcar en las profundidades de sitios web (por ejemplo, de gobiernos).
- Al hacer scraping (rascar) de bases de datos escondidas online usando herramientas como OutWit Hub y Scraperwiki.
- Convertir documentos en algo que pueda ser analizado usando herramientas como DocumentCloud.
- Tomando información de API.
- Recolectando los datos por el propio periodista a través de la observación, encuestas, formularios online o crowdsourcing.

En la tabla nº 1 se puede observar el año, temática, fuente y técnica de recolección usada para elaborar cada una de las informaciones que conforman la sección “Periodismo de datos” del Lab de RTVE:

Tabla nº 1: Año, temática, fuente y técnica de recolección de datos

| Año | Temática | Fuente | Técnicas de recolección de datos |
|------|-----------------------------|--------------------------------|--------------------------------------------------------------------------|
| 2011 | Elecciones generales 20-N | Web del Ministerio de Interior | Obtenidos en formato PDF y convertidos a Excel. |
| 2012 | La reforma Laboral a examen | Web de Eurostat, OIT y OCD | Manual. |
| 2012 | Los oscar 2012, en Twitter | Páginas webs oficiales | Scrapeando la Web de los distintos premios y descargando los resultados. |

| | | | |
|------|----------------------------------------------------------------|--------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| 2011 | Así está la mujer en el mundo | Web Population Reference Bureau. | Obtenidos directamente en Excel. |
| 2012 | Evolución de ingresos y gastos en los presupuestos del estado. | Ministerio de Hacienda y administraciones públicas. | Manual a través de la lectura del libro blanco. |
| 2011 | Transplantes y donaciones de órganos en España | Web de la Organización Nacional de Transplantes (ONT). | Obtenidos en formato PDF y convertidos al formato OCR para descargarlos en Excel. |
| 2011 | ¿Qué hace el Primer Mundo por los refugiados? | "Informe tendencias Mundiales" Fuente: ACNUR | Obtenidos directamente en Excel. |
| 2012 | El éxodo de españoles se dispara | Instituto Nacional de Estadística. | Descarga de los datos directamente en formato Excel. |
| 2012 | Día Mundial de la alimentación 2012 | Organización de las Naciones Unidas para la alimentación y la agricultura (FAO). Banco Mundial. | Descarga de los datos directamente en formato Excel. |
| 2012 | El patrimonio de nuestros dirigentes uno a uno | Web del Congreso de los Diputados de España. Boletín Oficial del Estado (BOE) | Datos obtenidos en PDF, y conversión en formato OCR para descargarlos en excel. |
| 2012 | ¿Funciona el "Crowdfunding" cultural en España? | Verkami | Datos descargados directamente en Excel. |
| 2013 | Radiografía de la Iglesia | Web del Vaticano | Manual. |
| 2013 | Derechos Humanos en el mundo | Amnistía Internacional. | Datos aportados en PDF, Conversión a OCR para descargarlos en Excel. También mediante extracción manual. |
| 2013 | Todo sobre Tokio 2020 | Dossier oficial de las candidaturas de Tokio, Madrid, Estambul expuestas en sus respectivas Web. | Datos aportados directamente en Excel. |

Fuente: elaboración propia

El tercer aspecto que se considera significativo señalar es la variedad temática de los hechos tratados, lo que muestra como esta especialización periodística permite informar sobre temáticas pertenecientes a sectores divergentes (economía, deporte, sociedad, política, cultura, etc). En concreto, como se puede observar en la tabla nº 1 se han abordado: las elecciones generales, la reforma laboral, los Oscar, los presupuestos del estado, el crowdfunding cultural en España, la iglesia, los juegos olímpicos, etc. Por tanto, puede usarse para tratar hechos y acontecimientos que competen a todas las secciones que conforman un medio de comunicación.

3.2. Tratamiento y visualización de los datos

El segundo paso en esta modalidad periodística es tratar los datos: decidir sobre qué se va informar. El periodista comprobará las “historias escondidas dentro de las bases de datos” para estructurar la información (Flores, 2013, p. 23), y es que “hasta que no estudias los datos no sabes cual es la historia a contar” (Miriam Hernánz, comunicación por e-mail, 18 de enero de 2016).

Recolectada, analizada y estructurada, es preciso mostrar visualmente la información de forma comprensible. En este contexto argumenta Flores (2013, p. 16):

“Nos encontramos con el periodista de datos como un perfil profesional emergente, cuyo entorno de trabajo se basa fundamentalmente en la construcción de visualizaciones cada vez más novedosas que combinan contenidos (de información) más tecnología y que tienen múltiples aplicaciones en las diversas formas del periodismo especializado”.

En la tabla nº 2 se muestra cómo desde el Lab de RTVE estructuraron cada información en función de los datos extraídos y el tipo de visualización seleccionada para cada temática abordada.

Tabla nº 2: Estructura y visualización de la información

| Temática | Estructura | Visualización |
|-----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|
| Elecciones generales 20-N | Reparto de escaños por comunidades, por provincias y por municipios. Evolución comparativa del voto en España desde 1977 hasta 2011. | Gráficos interactivos, mapas interactivos y tablas. |
| La reforma Laboral a examen | Análisis comparativo de la legislación española en materia laboral con la de otros países: Alemania, Francia, Italia, Austria, reino Unido, Suecia y EE.UU. Se comparan los siguientes asuntos: ¿Cuánto vale el despido? Requisitos de indemnización, periodo de prueba, contratos | Tablas que muestran los datos de cada |

| | | |
|----------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------|
| | temporales consecutivos, duración de los contratos temporales. Temporalidad, salario mínimo interprofesional, tasa de paro, paro juvenil, contrastos tiempo parcial y salario medio anual. | país analizado y gráfico interactivo. |
| Los oscar 2012, en Twitter | Análisis de la interacción social producida durante los Oscar. La información se estructura en: "Portada Oscar 2012". "Nominados" (a mejor película, director, actor protagonista, actriz protagonista, actor secundario, actriz secundaria, banda sonora, película de animación, película extranjera). y "Alfombra roja" (vestidos que han generado mayor numero de comentarios en Twitter: Emma Stone, Angelina Jolie, Natalie Portman, Bradley Cooper, Sandra Bullock, Gwyneth Paltrow, Jennifer Lopez, Penélope Cruz, Colin Firth, Christian Bale, Uggie, Robert Downey Jr., James Franco, Stacy Kiebler) | Muro de actividad social y gráficos interactivos. |
| Así está la mujer en el mundo | Análisis comparativo de situación de la mujer en los cinco continentes. Los asuntos tratados son: Población femenina (millones), mujeres menores de 15 años, mujeres entre 15 y 49 años, mujeres mayores de 49 años, alfabetización entre 15 y 29 años, activas económicas (+ de 15) años, nº de hijos por mujer, asistencia médica en partos, matrimonios en menores de 18 años, mujeres en el parlamento y mujeres con VIH sida. | Gráfico interactivo. |
| Evolución de ingresos y gastos en los presupuestos del estado. | Comparativa de la procedencia de los ingresos del estado y de los gastos de los presupuestos del estado entre 2010, 2011 y 2012. | Gráfico. |
| Transplantes y donaciones de órganos en España | Comparativa de los órganos más transplantados (renales, hepáticos, cardíacos, pulmonares, páncreas, y de Intestino). Y las "Donaciones": por año y por comunidad autónoma. | Gráfico interactivo. |
| ¿Qué hace el Primer Mundo por los refugiados? | Identificar cuales son los principales países industrializados receptores de refugiados y qué volumen de población refugiada acogen. La información se estructura en población refugiada y solicitudes de asilo. | Gráfico interactivo. |
| El éxodo de españoles se dispara | Comparativa de la cantidad de emigrantes españoles entre los años 2009, 2010, 2011. La información se estructura en: hombres, mujeres, y Total. | Gráfico interactivo. |
| Día Mundial de la alimentación 2012 | Comparativa continental de los siguientes asuntos vinculados con la alimentación: adecuación dietética, índice de subnutrición, Insuficiencia alimentaria, déficit de alimentos, Índice precio de alimentos, volatilidad de precios, política y violencia. | Mapa interactivo y gráfico. |

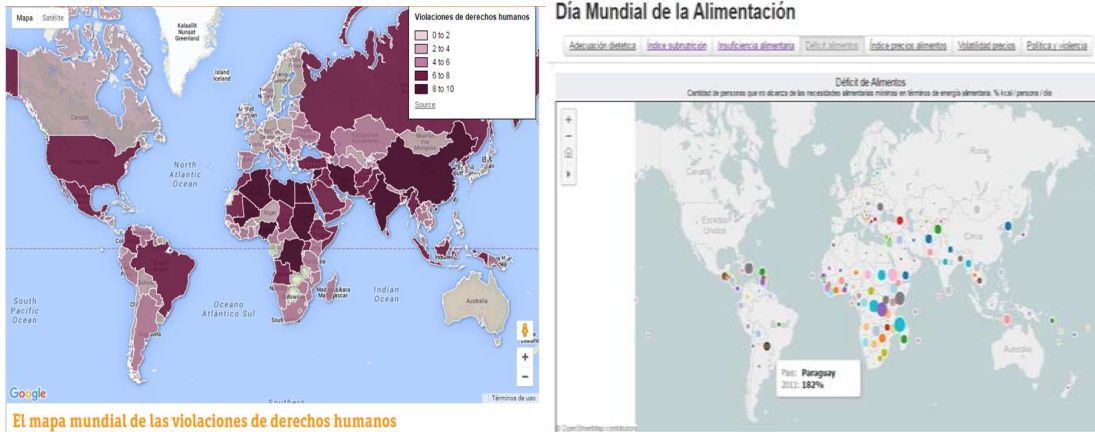
| | | |
|-------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|
| El patrimonio de nuestro dirigentes uno a uno | Identificación del patrimonio de los altos cargos: Ministros (PSOE-PP), Secretarios de Estado (PSOE- PP), Gobierno de Zapatero, Gobierno de Rajoy, Top 10. | Gráfico interactivo. |
| ¿Funciona el “Crowdfunding” cultural en España? | Identificación de los proyectos culturales que han fracasado y triunfado en España. Así como de las comunidades que más apuestan por estos proyectos. | Gráfico interactivo. |
| Radiografía de la Iglesia | Aportar información de los miembros del colegio cardenalicio. La información se divide en: tipos de cardenales, número de cardenales electores, número de cardenales del norte y del Sur en el Vaticano, media de edad de los cardenales, Potencias papales, países con menor y mayor número de católicos. | Gráficos y mapas interactivos. |
| Derechos Humanos en el mundo | Análisis de la violación de los derechos humanos en el mundo: violaciones de los derechos humanos, pena de muerte, homofobia, violencia de género, injusticia y cárcel. | Mapa y gráficos interactivos. |
| Todo sobre Tokio 2020 | Comparación de los presupuestos, datos básicos y distribución de las instalaciones deportivas de las tres sedes finalistas para organizar los juegos olímpicos (Estambul, Tokio y Madrid). | Mapas y gráficos interactivos. |

Fuente: elaboración propia

Por tanto, la estructura de los acontecimientos informativos indicados en la tabla anterior refleja como a partir de una base de datos se puede abordar un hecho desde diversas perspectivas.

Y si atendemos a los formatos narrativos usados para visualizar las piezas, en la muestra objeto de estudio predominan los gráficos interactivos seguido de los mapas. Pero aunque parezca que tan sólo han usado estas variantes es preciso destacar que el diseño de cada gráfico y cada mapa es divergente al resto. En el gráfico 1 se muestran dos modelos divergentes recurriendo un mismo tipo de presentación: los mapas. Y en el gráfico 2 otros dos modelos desemejantes recurriendo a un mismo tipo de visualización: los gráficos. Y es que las herramientas digitales permiten presentar visualmente la información a través de múltiples opciones: el periodista debe decidir cual es la más apropiada para cada caso.

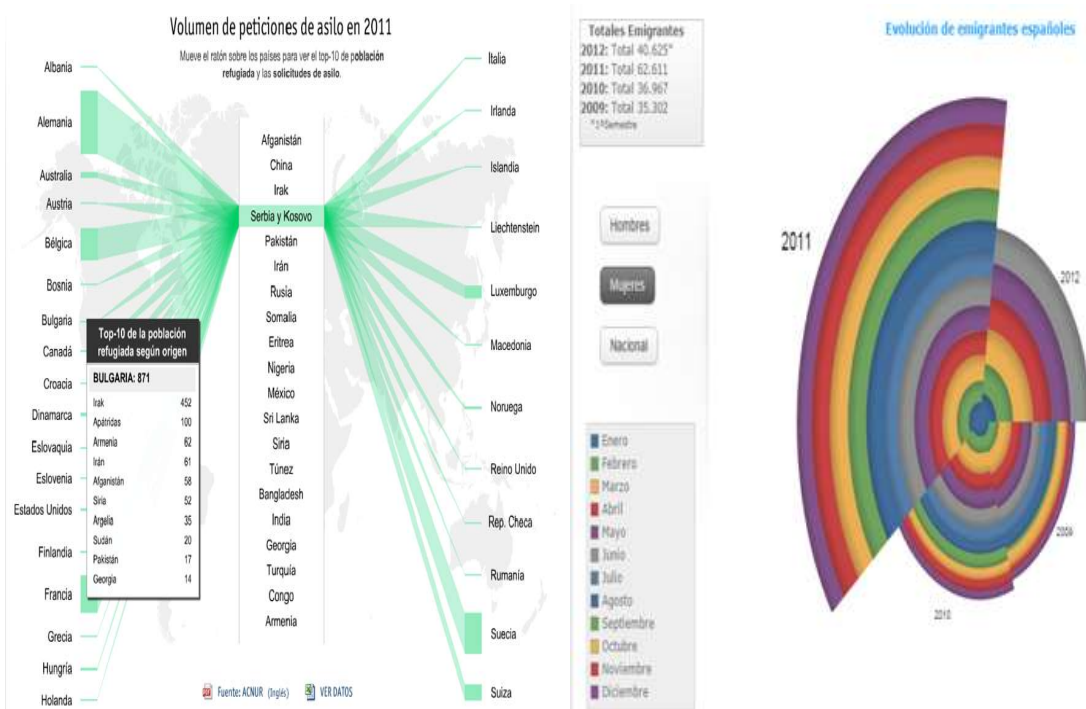
Gráfico nº 1: Mapas interactivos



El mapa mundial de las violaciones de derechos humanos

Fuente: Lab de RTVE

Gráfico nº 2: Gráficos interactivos



Fuente: Lab de RTVE

Unas visualizaciones interactivas que como muestran los gráficos no solo permiten presentar los datos, sino que el lector entienda lo que significa de un primer impacto o vistazo. Es por ello, que la visualización es un proceso clave para esta especialidad informativa.

4. Conclusiones

Las piezas informativas basadas en datos elaboradas por el Lab de RTVE son una herramienta útil para conocer cómo es el proceso de producción y edición de esta especialidad informativa. Un modelo comunicacional que puede aplicarse a temáticas divergentes (política, economía, deportes, economía, etc).

Los profesionales de la información pueden acceder a los datos desde la plataforma digital a través de diversos métodos: extraerlos directamente de la fuente seleccionada, realizando técnicas de búsquedas más avanzadas, e incluso usando técnicas manuales.

El tratamiento de los datos extraídos es fundamental para estructurar la información: de qué vamos a informar. Estructurada la información es preciso seleccionar qué tipo de visualización (gráfico, mapa, infografía, etc) y qué modalidad dentro de sus respectivas variantes es la más apropiada para informar al lector o espectador.

5. Referencias bibliográficas

P Brashaw (2010): “How to be a data journalist”, *The Guardian*, (1-10-10): <http://www.guardian.co.uk/news/datablog/2010/oct/01/data-journalism-how-to-guide>

S Crucianelli (2012): Webinario Introducción al periodismo de bases de datos [archivo de video]: <http://www.youtube.com/watch?v=WvHr0CshBxc>

AM De Vicente (2013): “Recursos Digitales para la investigación: Periodismo de datos”, en F Sabés y JJ Verón, *Comunicación y la Red*. España. Editorial Asociación Periodistas de Aragón.

EM Ferreras (2012): Nuevos perfiles profesionales: el Periodista de Datos. IV Congreso Internacional Latina de Comunicación Social “*Comunicación, control y resistencias*”: http://www.revistalatinacs.org/12SLCS/2012_actas/062_Ferreras.pdf

J Flores & C Salinas (2012). Sinergias en la construcción del Nuevo Periodismo derivadas del Data Journalism y el Transmedia Journalism. . III congreso Internacional Comunicación 3.0 “*Las media Enterprises y las industrias culturales, Investigar la Comunicación y los Nuevos Medios*”: http://www.2ip.es/wp-content/uploads/2013/01/sinergias_construccion_nuevoperiodismo.pdf

J Flores & C Salinas (2013). “El Periodismo de datos como especialización de las organizaciones de noticias en Internet”. *Correspondencias & Análisis*, nº 3, pp. 15- 34.

López García, et al. (2009). “Automatización de las bases de datos: potencialidades de herramientas básicas para otro periodismo posible”. *El profesional de la información*, 18, nº 3, pp. 308- 315.

N Mazotte (2012): Especialista en periodismo asistido por computadoras destaca tres herramientas innovadoras para reporteros [archivo de video], *Knight Center for Journalism in the Américas*: <http://knightcenter.utexas.edu/es/blog/especialista-en-periodismo-asistido-por-computadoras-destaca-tres-herramientas-innovadoras-para>

D Nguyen (2010): “Scraping for Journalism: A Guide for Collecting Data”, *Propublica* : <https://www.propublica.org/nerds/item/doc-dollars-guides-collecting-the-data>

K Peiró & J Guallar (2013): “Introducció al periodisme de dades. Característiques, panoràmica i exemples”, *Item: revista de biblioteconomia y documentación* 57, pp. 22-36.

F Steve & JC Nieto (2014): *Nuevos retos del periodismo especializado*. Madrid: SCHEDAS.

S Tejedor (2006): “La enseñanza del ciberperiodismo: hacia una transversalidad mixta”. *ZE. Revista de estudios de comunicación* 21, pp. 219- 239.

N Turrado (2012): “Las mejores herramientas de visualización de datos gratuitas”, [en línea], *Doctormetrics*: <http://www.doctormetrics.com/2012/08/29/herramientas-visualizacion-datos/>

Aida María de Vicente Domínguez es Doctora Cum laude por la Universidad de Granada. Licenciada en Periodismo por la Universidad Carlos III de Madrid. Licenciada en Humanidades por la Universidad de Navarra. Especialista universitaria en periodismo y comunicación científica. Miembro del Grupo de Investigación “Estudios sobre Comunicación y Sociedad de la Información” de la Universidad de Málaga (Plan Andaluz de Investigación de la Junta de Andalucía- HUM 664).