El periodismo de datos y la web semántica

Adolfo Antón Bravo*
adanton@ucm.es

(Abstracts y palabras clave al final del artículo)
Enviado: 9 de abril de 2013
Evaluado: 10 de abril de 2013
Aceptado: 11 de abril de 2013

INTRODUCCIÓN

La web siempre se pensó como un lugar donde las personas pudieran compartir y acceder a información, *un reto social más que técnico*, como expresaba su inventor Tim Berners-Lee[1]. La *gran telaraña mundial* (traducción de *World Wide Web*) puede ser una gran ventana para el conocimiento, una gran biblioteca abierta, una oportunidad de negocio para muchas empresas, un metamedio, un espacio donde relacionarnos... y también en el campo idóneo para la acción y el desarrollo del Periodismo. Cuando conocí la web estudiaba periodismo -ahora realizo la Tesis sobre las *Tecnologías de la web semántica*- y la web me reconcilió con las motivaciones que me hicieron comenzar a estudiar: la información como parte/motor indisoluble del cambio social, como herramienta epistemológica, como universo de potencialidades donde interactúan personas, tecnologías, contenidos, conocimiento... Años después y muchos términos, prácticas, tecnologías y usos entre medias, aparece el *periodismo de datos*, una especie de punto de inflexión donde se encuentra el periodismo, la *usabilidad* (facilidad de uso), el diseño para la interacción con el usuario, la infografía, la visualización, la accesibilidad, la web y otras tecnologías que posibilitan todo lo anterior.

El *periodismo de datos* se refiere al periodismo como ese lenguaje que *cuenta historias de la realidad*. Se utiliza mucho la expresión *los datos cuentan historias*, datos que se expresan en visualizaciones, por lo que también se habla de *narrativas visuales* o *visualización narrativa*. Se refieren a datos expresados de muchas formas: numéricos, alfanuméricos, contenidos textuales, bases de datos, tablas, etc. en archivos de distintos formatos de archivo. De entre todos ellos, me interesan sobre todo

---

* Doctorando y DEA del Departamento de Periodismo III:

[1] Cita de Tim Berners-Lee recogida en Wikiquote https://en.wikiquote.org/wiki/Tim_Berners-Lee procedente de su libro *Weaving the Web* http://www.w3.org/People/Berners-Lee/Weaving/Overview.html que se tradujo en español como *Tejiendo la red*. Fue el inventor de la web y dirige el consorcio que se ocupa de las tecnologías de la Web, *W3C* (*World Wide Web Consortium* o *Consorcio W3*)

ISSN: 1135-7991
http://dx.doi.org/10.5209/rev_CIYC.2013.v18.41718

los datos que cumplen con los estándares de la *web semántica* por varias cuestiones fundamentales: la estructura de estos datos, expresados en estándares abiertos de libre uso; su publicidad, disponibles normalmente en la web o a través de servicios web; la posibilidad de utilizar y reutilizar los datos para diversos fines; y la disponibilidad de herramientas de visualización de esos datos a disposición de cualquier persona, no solo de periodistas, investigadores, científicos o informáticos.

No es la primera vez que se utilizan los datos o análisis estadísticos en el periodismo, ya ocurrió con el *periodismo científico* o *de precisión*; tampoco es la única *propuesta* periodística surgida de la web, ya que el *periodismo ciudadano* o *participativo* se combinó con la web en el *periodismo 2.0*. Prestaremos atención también a estas cuestiones.

DEL PERIODISMO DE PRECISIÓN AL PERIODISMO DE DATOS

Para conocer del *periodismo de precisión* no hace falta irse muy lejos, encontramos en la Universidad Complutense a José Luis Dader[2]. El segundo capítulo del libro "Periodismo de precisión: la vía socioinformática de descubrir noticia"[3], lo titula *El periodismo de precisión como evolución y complemento del periodismo de investigación*.

Dader diferencia *periodismo de investigación* de *periodismo de precisión* ya que "responden a dos ejes axiológicos diferentes que, por eso mismo, les permite mantener su propia autonomía y producciones periodísticas en ocasiones bien distantes". También señala que si en el *periodismo de investigación* se utilizan métodos convencionales como entrevistas, lectura de documentos, etc., en el *periodismo de precisión* los denomina "métodos anticonvencionales" procedentes de las ciencias sociales, en concreto del campo estadístico y el análisis informático que "permiten practicar otro tipo de precisión periodística expositiva o aclarativa de cualquier otra información relevante aportada por fuentes voluntariamente identificadas sobre asuntos que impliquen un manejo de cifras o acumulaciones cuantitativas alfanuméricas". A este tipo de fuentes las califica de "información de declaraciones", ya que iguala el relato -narrativo- de un acontecimiento con "los resultados numéricos de una recopilación (...) expresada en gráficos evolutivos e índices estadísticos de significación", si bien, advierte que "la comprobación profesional de este último tipo de comunicados requerirá destrezas". He aquí un anticipo de lo que ocurre en la práctica del *periodismo de datos*.

El *periodismo de precisión* lo desarrolló el periodista norteamericano Philip Meyer en los años 1960' y se define por el uso de las matemáticas u otros métodos de las ciencias sociales para interpretar los datos. Famoso fue su estudio de las causas de las revueltas de Detroit (EE.UU. de Norteamérica) en 1967 que produjeron

---

[2] José Luis Dader, catedrático de Periodismo Universidad Complutense de Madrid http://pendientedemigracion.ucm.es/centros/webs/d163/index.php?tp=C.%20V.%20Profesores&a=profs&d=20588.php

[3] *Periodismo de precisión, la vía socioinformática de descubrir noticias*. José Luis Dader. Editorial Síntesis. 1997 (Reimpresión en 2002) (Edición electrónica en 2010).

más de 40 muertos, 467 heridos, 7.200 arrestos y 2.000 casas destruidas y se convirtió a la postre en la revuelta con más víctimas mortales en la historia de los EE.UU. Meyer obtuvo el *Pulitzer* y de su experiencia publicó en 1973 el libro "Precision Journalism: A Reporter's Introduction to Social Science Method", traducido al español como "Periodismo de precisión. Nuevas fronteras para la investigación periodística" [4] El diario británico *The Guardian* [5], uno de los máximos exponentes del *periodismo de datos*, con una sección propia[6], se fijó en el estudio de Meyer para investigar las revueltas juveniles de Inglaterra en 2011, a raíz del asesinato por parte de la policía de un joven de 29 años, buscando inspiración, puntos en común y formas de explicar lo ocurrido[7]. Conclusiones sobre las que no estuvo de acuerdo Darcus Howe, periodista y activista anglocaribeño, que hablaba claramente de *insurrección* y de políticas racistas y clasistas sufridas por la población inglesa[8], si bien nos pueden aportar distintas lecturas que nos ayuden a entender los acontecimientos.

Meyer explica[9] que en Detroit investigaron a través de una encuesta cuantitativa para identificar a quienes participaron en las revueltas y sonsacar las causas de la misma, tal como había hecho la Universidad de California con las revueltas de Watts (Los Angeles, EE.UU. de Norteamérica) en 1965, con la diferencia de que la universidad dedicó dos años a la misma y Meyer quería realizarla en tan solo tres semanas, para lo cual se ayudó de la psicóloga Nathan Caplan. *The Guardian*, por su parte, suele contar con la ayuda de *London School of Economics*[10], quienes han utilizado el enfoque *grounded theory* (teoría fundamentada)[11], un método de investigación que busca generar teoría de los datos y se utiliza en análisis cualitativos pero también en cuantitativos.

Con la investigación desmontaron varias ideas preconcebidas que "explicaban" la revuelta: la teoría *riff raff* y la de la asimilación. Según la primera, la revuelta se producía por ser la única forma posible de avance social que tienen las personas que se encuentran social y económicamente deprimidas; para la segunda, dado que la población afroamericana de Detroit provenía de áreas rurales del sur, tuvieron pro-

---

[4] Enlace de Amazon en la versión inglesa http://www.amazon.com/Precision-Journalism-Reporters-Introduction-Science/dp/0742510883; y la versión española, http://www.amazon.com/Periodismo-Precision-Spanish-Edition-Philip/dp/8476762380

[5] Web de *The Guardian* http://www.guardian.co.uk

[6] *Data*, sección de datos de *The Guardian* http://www.guardian.co.uk/data, donde encontramos *Data Blog* (blog de datos), donde se encuentra la frase *Facts are sacred* (los hechos son sagrados) http://www.guardian.co.uk/news/datablog

[7] *Reading the Riots. Investigating England's summer of disorder* (Leyendo las revueltas. Investigando el verano del desorden de Inglaterra, en traducción libre) http://www.guardian.co.uk/uk/series/reading-the-riots

[8] Reflexiones aparecidas en el artículo de Diego Sanz Paratcha en *Diagonal*: "Una línea que une 30 años de revueltas" https://www.diagonalperiodico.net/global/linea-une-30-anos-revueltas.html

[9] Artículo de Philip Meyer en *The Guardian*: "Riot theory is relative" (la teoría de la revuelta es relativa, en traducción libre) http://www.guardian.co.uk/commentisfree/2011/dec/09/riot-theory-relative-detroit-england

[10] Web de *The London School of Economics and Political Science*, http://www2.lse.ac.uk/home.aspx

[11] *Grounded theory*, *teoría fundamentada* o *muestreo teórico*, desarrollada por los sociólogos Barney Glaser y Anselm Strauss http://es.wikipedia.org/wiki/Muestreo_te%C3%B3rico

blemas para adaptarse al norte urbano e industrial, convirtiendo la frustración en revuelta. Con la investigación comprobaron que tanto los que participaron como los que no, compartían ingresos y nivel educativo y que entre los que participantes, los que habían nacido en el norte superaban tres veces a los que provenían del sur. Por lo que propusieron una tercera hipótesis: cuanto más cerca te encuentras del objetivo deseado mayor es la frustración de no alcanzarlo, lo que aumenta si además otros progresan mientras uno sigue estancado.

Para la investigación Meyer empleó un ordenador *IBM 7090* sobre el que aprendió a programar y desarrollar *CAR* (*Computer-Assisted Reporting*, investigación periodística asistida por ordenador), la aplicación del método científico al periodismo. Las tecnologías empleadas por *The Guardian* y *LSE* son las más modernas y tratan con volúmenes de datos mucho mayores, pero Meyer destaca que lo importante de ambas investigaciones no es el *hardware* sino el *software*, el método empleado y que puede servir a otros periodistas a hacer lo mismo.

Otra coincidencia entre el *periodismo de precisión* y el *periodismo de datos* la encontramos en la cita que Meyer destaca de Robert Maynard Hutchins [12] "the truth about the facts" ("la verdad de los datos", en traducción libre), similar al lema que reza en el *Data Blog* de *The Guardian* "Facts are sacred" (los hechos son sagrados, en traducción libre). En la actualidad, el *periodismo de datos* se aprovecha de un volumen de datos inmenso y que aumenta continuamente, lo que se conoce como *Big Data* o *grandes datos*, en referencia a los sistemas de información y comunicación que gestionan grandes conjuntos de datos o *data sets*. Estos datos son más o menos públicos y más o menos accesibles a través de la web -bien sea en el propio contenido, ayudado por el *HTML* o en los documentos enlazados-, en bases de datos o a través de otras herramientas informáticas de búsqueda. Luego hay que filtrar, seleccionar y visualizar de diversas formas posibles, en formato papel o audiovisual, vía web o TV.

PERIODISMO DE DATOS Y PERIODISMO 2.0

El *periodismo 2.0* toma el nombre de la *web 2.0* y guarda una estrecha relación con ella. En el ámbito de la web, *web 2.0* fue la denominación auspiciada porTim O'Reilly en 2005[13] para explicar el estado de la web, con todas las tecnologías que habían aparecido y que en apariencia distaba mucho de la web original. Sin embargo, la esencia era la misma y las tecnologías iban cumpliendo los propósitos originales de la web, por lo que Berners-Lee nunca le dio demasiada importancia al tér-

[12] Enlace de la Wikipedia http://es.wikipedia.org/wiki/Robert_Maynard_Hutchins

[13] Según la Wikipedia (http://es.wikipedia.org/wiki/Web_2.0), el término fue utilizado por primera vez por Darcy DiNucci en 1999 en su artículo *Fragmented Future* (http://www.darcyd.com/fragmented_future.pdf). En 2004, durante una taller de *O'Reilly Media*, Dale Dougherty, en una tormenta de ideas con Craig Cline de MediaLive, sugirió que la web estaba en un renacimiento, con reglas que cambiaban y modelos de negocio que evolucionaban. Finalmente, Dougherty, John Battelle, MediaLive y O'Reilly Media lanzaron la primera conferencia sobre la *web 2.0* en octubre de 2004.

mino. Por aquel entonces los sitios web dejaron de ofrecer únicamente información en un esquema unidireccional del *emisor* hacia el *receptor* sino que también facilitaban la interacción, retroalimentación e interoperabilidad gracias a diversas tecnologías y además mostraban un diseño más cuidado. Todo ello favorecía nuevos modelos de información periodística.

Se asocia a *sitios web 2.0* la posibilidad de permitir a los usuarios interactuar y colaborar entre sí como creadores de contenido (*User-generated content*, contenido generado por el usuario) en las primeras redes sociales de comunidades de usuarios agrupados por un interés común; los servicios de alojamiento de vídeos; los wikis[14]; y, sobre todo, los blogs [15]. Así, el periodista o el ciudadano no solo crea una página web -un documento *HTML*- de su *propiedad* para mostrar contenido sino que también puede compartirlo en otros sitios web que no le pertenecen. Esta comunicación dialógica no sería posible sin las innovaciones tecnológicas -*AJAX*, *SOAP*, *XML* y *JavaScript* [16]- que mejoran la experiencia de usuario (*UX*, *User eXperience*) de los navegantes.

Por tanto, una de las principales características de la *web 2.0* se produce con el cambio de lector pasivo a navegador activo y colaborador en la creación y producción de contenidos por parte de los navegantes, quienes no solo crean contenidos sino que en el uso de los distintos servicios web, cuando comparten sus gustos/experiencias/deseos con otros, se convierten en productores de contenido, una fuente de datos interminable que además enriquece los perfiles de usuario de las empresas [17] y que da cuerpo a la *web social*.

---

[14] *Wiki* (del hawaiano, significa *rápido*) es un tipo de software para sitios web que permite que las páginas sean editadas por una o más personas con una sintaxis propia que traduce en *HTML*.

[15] *Blog* (de *web log* o *registro web*, se puede traducir como *diario web* o *cuaderno de bitácora*) se trata de un tipo de software para sitios web que permite la publicación de artículos -*posts*- que se pueden organizar cronológicamente.

[16] Las tecnologías *AJAX*, *SOAP*, *XML* o *JavaScript* permitían la interacción usuario-navegador de tal forma que simulaba una aplicación de escritorio pero con los datos procedentes de algún lugar de la web o incluso de varios lugares de la web a la vez, enriqueciendo la experiencia del usuario.

AJAX, *Asynchronous JavaScript and XML* (*JavaScript* asíncrono y *XML*) es un método utilizado en el lado cliente -el navegador del usuario- para crear aplicaciones web asíncronas, interactivas o *RIA* (*Rich Internet Applications*). Las aplicaciones pueden enviar o recibir datos de un servidor sin interferir en la disposición de los elementos o en el comportamiento de una página web. Por ejemplo, la información de la página se puede ir actualizando sin necesidad de recargarla en el navegador. Los datos se pueden recuperar con *XMLHttpRequest* o con *JSON* (*JavaScript Object Notation*)y las solicitudes no tienen por que ser asíncronas.

SOAP, *simple object access protocol* o *protocolo de acceso a objeto simple* (traducción libre) es un protocolo utilizado en servicios web para que dos *objetos* se puedan comunicar por medio de intercambio de datos *XML*.

XML, *eXtensible Markup Language* o *lenguaje de marcas extensible* (traducción libre), desarrollado por *W3C* (*World Wide Web Consortium*), derivado de *SGML* (*Standar Generalized Markup Language*, *estándar de lenguaje de marcado generalizado*, utilizado para la organización y etiquetado de documentos), permite definir la gramática de lenguajes específicos para estructurar documentos o conjuntos de documentos. Además de la web, se puede utilizar como estándar de intercambio de información estructurada entre diferentes plataformas en bases de datos, editores de texto, hojas de cálculo, etc.

[17] Normalmente, el contenido generado por los usuarios en una página web propia está asociada a la persona por la propiedad del documento *HTML*, el dominio (gestionado por *DNS*) o la dirección *IP* (*Internet Protocol*) desde la que se conecta. Algunos datos son públicos, como los que incluyamos en los documentos *HTML* o en el *DNS*, mientras que la dirección *IP* y otros datos empleados por las tecnologías para el funcionamiento de las mismas, van a estar en poder, normalmente, de los proveedores de acceso *ISP* (*Internet Services Provider*, proveedor de servicios de Internet).

Podríamos añadir, además, los datos provenientes de dispositivos con los que interrelacionamos conscientemente o no, como los que provienen de nuestras operaciones con tarjetas de crédito u otras tarjetas de acceso a productos y/o servicios. Un paso más allá, en un sentido orwelliano, se encuentra *la Internet de las Cosas* (traducción de *Internet of Things*)[18], la conexión de cualquier objeto con el que tenemos relación en nuestra cotidianeidad, desde el frigorífico hasta un semáforo, que envían información sobre su funcionamiento.

En el mismo espacio temporal y físico -California, al igual que O'Reilly- Dan Gillmore[19] daba nombre al *periodismo 2.0* desde su práctica periodística en el *San Jose Mercury News*, un *periodismo ciudadano* o *participativo*, más personal, que maneja la interacción y la participación con el lector gracias a las tecnologías de la *web 2.0*.

Hemos de volver unos años atrás, alrededor de 1999 -época de la *web 1.0*- para constatar procesos similares poco conocidos, cuando movimientos sociales de distintos lugares y sin relación directa entre sí habían comenzado a explorar los límites del *periodismo participativo* y *ciudadano* con consignas muy clarificadoras -*dar voz a los sin voz* o *don't hate the media, become the media* (no odies los medios, conviértete tú en un medio)[20]- y propuestas tecnológicamente muy avanzadas y muy participativas: Londres, con las primeras retransmisiones de las manifestaciones denominadas *Reclaim The Streets* (*Reclama las calles*, realizadas luego también en Madrid y Barcelona) en tiempo real; Madrid y Barcelona, con la cobertura y relato de diversas manifestaciones de movimientos sociales a través de la *Agencia -de noticias- en Construcción Permanente*, *ACP*; y sobre todo, conocido en todo el mundo, Seattle, con las movilizaciones contra la cumbre de la *Organización Mundial del Comercio* (*WTO*, *World Trade Organization*) a través de *Indymedia.org*[21]. En los tres casos, la apuesta por un *periodismo participativo*, editado colaboratibamente (publicación *P2P* o *peer-to-peer*, *igual a igual*[22]), con licencias libres[23], abierto y dialógico propició el uso, creación o implementación de sendas herramientas informáticas -software de sistemas de gestión de contenidos (*CMS* o *Content management System*). Poco después, ambas tres iniciativas formarían parte de la red

---

[18] Internet de las Cosas en la Wikipedia: https://es.wikipedia.org/wiki/Internet_de_las_Cosas

[19] Periodista de *San Jose Mercury News* de 1994 a 2005, director del *Knight Center for Digital Media Entrepeneurship*, fundador del *Centre for Citizen Media*, autor del libro *We the Media: Grassroots Journalism by the People, for the People* (http://wethemedia.oreilly.com/) y *Mediactive* (http://mediactive.com/).

[20] *Don't hate the media become the media*, frase de Jello Biafra, cantante y líder del grupo punk *Dead Kennedys* recogida por Indymedia como lema. https://es.wikipedia.org/wiki/Jello_Biafra

[21] *Indymedia* sería el acrónimo para *Independent Media*, si bien el nombre completo es *Independent Media Center* o *centro de medios independientes*, por cuanto resultaba un referente informativo de las movilizaciones de colectivos sociales, a menudo asociada a movilizaciones antiglobalización com las que la fundaron http://es.wikipedia.org/wiki/Manifestaciones_contra_la_cumbre_de_la_OMC_en_Seattle

[22] *Peer to Peer*, entre iguales, se refiere a redes informáticas o de estructuras organizativas donde todos los nodos -ordenadores, personas- tienen las mismas posibilidades de trabajar en la red. En periodismo, se refiere a la escritura colaborativa o a proyectos que promueven estas prácticas. http://p2pfoundation.net

[23] Copyleft se refiere al conjunto de licencias libres, es decir, que cedes derechos otorgados por el copyright como el de uso, copia, modificación y distribución de la obra protegida, pudiendo combinar estas cuestiones. http://es.wikipedia.org/wiki/Copyleft

Indymedia, compuesta por más de cien nodos en más de treinta países y todavía en funcionamiento.

Quizás lo fundamental del *periodismo 2.0* consiste en la constatación de la incorporación a la práctica periodística el uso de Internet y la *web* como fuente de información y como vía de transmisión de las noticias, algo que ahora resulta muy extraño no contemplar pero que unos años más atrás no lo parecía tanto. El *periodismo de datos* hereda del *periodismo 2.0* el uso de las tecnologías, la colaboración que se establece entre los periodistas, la participación de personas que combinan aptitudes periodísticas o técnicas, las interacción con los lectores o la posibilidad de realizar otras narrativas.

Muchas redes sociales y empresas hacen de los contenidos su modelo de negocio por la cantidad de información que disponen de los usuarios. En *Facebook*, *Twitter*, *Blogspot*, *Tumblr*, *Yahoo* o muchas otras, podemos participar de los servicios que ofrecen con una identificación de usuario, aunque normalmente, hay un rango de servicios gratuitos y otros de pago, bien seamos un tipo de usuario u otro, una empresa u otra entidad. La gratuidad de estos servicios se compensa con la información que ofrecemos de nosotros mismos, que servirá para reunir perfiles de usuario-consumidor más avanzados. Los datos que generemos (*UGC*), según las políticas de acceso a los mismos, podrán ser públicos, pero dependerá de la propia empresa que controle el servicio la facilidad y valor de estos datos. Su acceso se realizará a través de una *API* (*Application Programming Interface*, Interfaz de acceso a la aplicación) pública y documentada.

Todo este caudal de información expresado en *RDF*[24] junto con otros, fue denominado por Berners-Lee como *Giant Global Graph* (*GGG*, *gráfico global gigante*)[25], -evolución de *WWW*- en vez de *web 3.0*, por cuanto estos datos podrían crear un *gráfico social* (*social graph*)[26] interminable, vinculados a través de *Linked Data*.[27] Siguiendo a Berners-Lee, más que de *periodismo 3.0* podríamos hablar de *periodismo de datos* para la práctica de *contar historias* que se encuentran en el gran volumen de información disponible en la web, en iniciativas como *Linked Data*, en contenido generado por el usuario (*UGC*) y/o en otros tipos de conjuntos de datos.

---

[24] De la Wikipedia, *RDF* (*Resource Description Framework*, marco de descripción de recursos) propone un modelo que transforma las declaraciones de los recursos en expresiones con la forma sujeto-predicado-objeto (tripletes). El sujeto es el recurso que se está describiendo; el predicado es la propiedad o relación que se desea establecer acerca del recurso; el objeto es el valor de la propiedad o bien otro recurso con el que se establece la relación. https://en.wikipedia.org/wiki/Resource_Description_Framework y http://www.w3.org/RDF/

[25] Artículo del blog de Berners-Lee "Gian Global Graph" http://dig.csail.mit.edu/breadcrumbs/node/215, 2007

[26] Berners-Lee enlaza el texto "Thoughts on the Social Graph" (Ideas sobre el gráfico social, en traducción libre) de Brad Fitzpatrick y David Recordon. http://bradfitz.com/social-graph-problem/, 2007

[27] Gráfico de Linked Data http://lod-cloud.net/versions/2011-09-19/lod-cloud_1000px.png

DATOS ABIERTOS VINCULADOS

Como he señalado anteriormente, la iniciativa más interesante para el *periodismo de datos*, si bien no la única, de la *web semántica* es *Linked Data*[28] (*Datos Vinculados*). *Linked Data*, término que se se utiliza en informática en general para relacionar unos datos con otros, lo adoptó Berners-Lee para exponer una de las iniciativas de la *web semántica*, denominada igualmente *Linked Open Data* (*Datos Vinculados Abiertos*), de cara a conectar el contenido de las páginas webs, las webs -documentos *HTML*- y los sitios web completos, con el objetivo de mejorar su búsqueda, acceso y reutilización. Para Berners-Lee, utilizar *Linked Data* y participar en la *web semántica* "no es solo poner datos en la web sino crear enlaces para que una persona o una máquina puedan explorar la web de datos (...), encontrar otros datos relacionados". En resumen, se enlaza la información de cara a que las personas, asistidas por ordenadores, exploremos esos datos relacionados escritos en *RDF*. El encuentro con el *periodismo de datos* es formidable ya que todos los *datos abiertos vinculados* pueden utilizarse como fuentes de datos para las diversas investigaciones.

La web en sí ya es una fuente de información, con multitud de datos disponibles. La importancia de la web en el *periodismo de datos* no se circunscribe únicamente a los *datos abiertos vinculados* de la *web semántica* sino que la web ya supone una organización de la información en cada documento y en cada sitio web. Si la web es un conjunto de documentos *HTML* -páginas web- reunidos en sitios web -los identificamos por su dominio normalmente- a los que puedes acceder desde un navegador, la *web semántica* pretende identificar el contenido con múltiples tecnologías, describiéndolo con anotaciones, con etiquetas, ontologías, taxonomías... metainformación. Para hacer esto, pasamos de un lenguaje de marcas limitado como es el *HTML* a un lenguaje *XML* donde las marcas son extensibles -podemos construir las nuestras propias o participar de unas que ya existan para contenidos similares- y con una estructura sintáctica basada en *RDF*. La web Semántica nos acerca las tecnologías de información con la práctica periodística, si bien no es la única tecnología posible.

Lo que diferencia la *web de hipertexto -web 1.0-* de la *web de datos* es la información que aportan las relaciones que se establecen. Si en la primera las relaciones son en *HTML*, en la segunda se expresa en *RDF*. Esto se consigue gracias a *URIs*[29] que identifican cualquier tipo de objeto o concepto, *HTTP*[30] *URIs*, para que esas denominaciones puedan llegar a otras personas o *UAs* (*User Agents*, agentes de usuario[31]; *RDF* para aportar información, metainformación y enriquecer el conteni-

---

[28] Boceto de la definición de *Linked Data* por parte de Tim Berners-Lee, director del *W3C* http://www.w3.org/DesignIssues/LinkedData.html

[29] *URI*, *Uniform Resource Indicator* o *identificador uniforme de recursos*, cadena de caracteres que identifica inequívocamente un recurso de red. https://es.wikipedia.org/wiki/Uniform_Resource_Identifier

[30] *HTTP* o *HyperText Transfer Protocol*, protocolo de transferencia de hipertexto utilizado en la web. https://es.wikipedia.org/wiki/HTTP

[31] *User agents* o *agentes de usuario* es un término que puede referirse a varias cosas, como por ejemplo a un servidor web intermedio que nos filtra una búsqueda o un dispositivo que nos muestra esa búsqueda, es decir, cualquier *agente* (ordenador que realiza algún servicio) intermedio entre el navegante y la localización de la web.

do que se denomina; *SPARQL*[32] para su búsqueda; y enlaces a otros sitios *URIs* para descubrir la información relacionada.

En 2010, de cara a animar al uso de *Datos Vinculados* en ámbitos generalistas y en especial por parte de los distintos gobiernos, anunciaron un sistema de cumplimiento con *Linked Data* basado en cinco estrellas[33] que corresponden a distintos cumplimientos:

Una estrella, los datos están disponibles en la web, en cualquier formato y con licencia abierta.

Dos, si los datos están disponibles como datos estructurados, legibles por ordenadores, como por ejemplo un archivo *PDF*; no sería el caso de una foto de un texto.

Tres, para datos estructurados legibles por ordenadores no tiene un formato propietario. Por ejemplo, cumpliría *csv*[34] y no valdría *xls* (formato de hojas de cálculo privativo y propietario de *Microsoft Office Excel*).

Cuatro, si además utiliza estándares de la web como *RDF* y *SPARQL* para identificar los datos, lo que permite que otras personas puedan enlazarlos.

Por último, cinco estrellas si además de todo lo anterior, los datos se enlazan con otros, se vinculan.

Lo que une el *periodismo de datos* con la *web semántica* también diferencia a la web de otras fuentes de datos, ya que no tienen por qué estar enlazados entre sí ni van a tener tanta metainformación como los datos de la *web semántica*, por lo que necesitará de otras tecnologías para la extracción, análisis y representación de los datos.

Desde otros sectores han surgido iniciativas que también abogan por la publicidad, accesibilidad y reutilización de los datos como es el movimiento *Open Data* (*Datos abiertos*) y *Open Gov Data* (*datos gubernamentales abiertos* o *datos públicos*) para los datos gubernamentales, que prevengan además de prácticas de corrupción y mejoren la democracia, relacionado con el movimiento *Open Government* (*gobierno abierto*). La definición de datos abiertos[35] ha sido realizada por la *Open Knowledge Foundation* (*OKFN*, *Fundación por el Conocimiento abierto*[36]) y resume los principios que definen lo abierto en relación a los datos y el contenido: son abiertos los datos o contenidos si cualquiera puede usarlos, reusarlos y redistribuirlos, con la única restricción permitida de que obligue a la atribución de su autoría y/o a compartir los datos con la misma licencia. Si estos datos cumplen con los prin-

---

[32] Acrónimo recursivo de *SPARQL Query Language* o lenguaje de consultas *SPARQL* para *RDF*. *Recomendación W3C* 15 enero 2008 http://www.w3.org/TR/rdf-sparql-query/, si bien ahora hay una nueva versión, *SPARQL 1.1* del 21 de marzo de 2013 http://www.w3.org/TR/sparql11-overview/

[33] Web de la iniciativa *5 estrellas* para datos vinculados: http://5stardata.info/; imagen de la campaña: http://5stardata.info/5star-steps.png

[34] *csv*, acrónimo de *comma separated values*, valores separados por comas se trata de un tipo de documento en formato abierto y muy sencillo que sirve para representar los datos en forma de tabla, donde las columnas están separadas por comas -o punto y coma si la coma es el separador decimal- y las filas por saltos de línea https://es.wikipedia.org/wiki/CSV

[35] Open Definition, iniciativa de OKFN http://opendefinition.org/

[36] Web de OKFN, http://www.okfn.org

cipios de *Linked Open Data*, se les denomina *Linked Open Government Data* (*datos gubernamentales abiertos vinculados*).

Los ocho principios del *Open Government Data*[37] hablan de la necesidad de datos públicos, completos, originales, fechados/datados, accesibles, electrónicos, sin discriminar el acceso a cualquier persona, sin discriminar su acceso por *software* propietario para su acceso, con licencia de copyright libre y, en último término, revisables. Los objetivos de los portales gubernamentales de *datos abiertos* los resume el portal de la *Iniciativa Datos Abiertos* del *Gobierno Vasco*[38]:

Generar valor y riqueza, obteniendo productos derivados por parte de empresas, infomediarios y ciudadanía en general. Por ejemplo, la reutilización de datos del tiempo.

Fomentar la transparencia. Al tener los datos públicos disponibles podremos evaluar la gestión pública.

Interoperabilidad entre administraciones. Si los datos son abiertos, distintas administraciones pueden trabajar con los mismos datos, enriqueciéndolos y mejorando nuestra relación con la administración.

Ordenación interna de los datos públicos, promoviendo la eficiencia en la documentación y clasificación de datos.

Helen Darbishire, de *Access Info*[39], señalaba en el taller de *Medialab*[40] sobre *Periodismo de datos* titulado "Derecho de acceso a la información pública"[41] que España es el país más grande de la Unión Europea que no cuenta todavía con una **Ley de Acceso a la Información**, que se concreta en promocionar y facilitar el acceso a los datos públicos. En la actualidad se encuentra tramitándose en el Congreso la **Ley de Transparencia**, si bien el proceso ha sido muy criticado por *Access Info* y *Civio* por llevar un proceso lento y poco transparente. Sin embargo, la información ambiental se encuentra regulada por la *Ley 27/2006*, que garantiza el acceso público a toda la información ambiental[42], como traslación de una directiva europea.

La *Open Knowledge Foundation* recoge en el manual de datos abiertos "Open Data Handbook Documentation"[43], las áreas en las que encuentra beneficioso el *Open Data*, sobre todo *Open Government Data*: transparencia y control democrático, participación ciudadana, innovación y nuevas formas de conocer de distintas

---

[37] Los 8 principios del *Open Government Data* aparecen recogidos en http://www.opengovdata.org

[38] *Iniciativa Datos Abiertos del Gobierno Vasco*, http://www.irekia.euskadi.net/es/news/11661-como-trabajar-con-datos-amigables-curso-periodismo-datos?t=1

[39] *Access Info Europe* se dedica a promover y proteger el derecho de acceso a la información en Europa y el mundo, como una herramienta para la defensa de las libertades civiles y los derechos humanos, para facilitar la participación pública en la toma de decisiones y la fiscalización de los gobiernos. http://www.access-info.org/

[40] Helen Darbishire de *Access Info Europe* en la primera sesión formativa de *periodismo de datos* titulada "La captura de datos", 12/01/2012 http://medialab-prado.es/article/derecho_acceso_informacion_espana

[41] Charla impartida el 12 de enero de 2012 en el marco de la primera sesión formativa de *periodismo de datos*: La captura de datos las diferentes formas de conseguir datos de las instituciones públicas y de otros portales de información. http://medialab-prado.es/article/derecho_acceso_informacion_espana

[42] Ley 27/2006, de 18 de julio, por la que se regulan los derechos de acceso a la información, de participación pública y de acceso a la justicia en materia de medio ambiente (incorpora las Directivas 2003/4/CE y 2003/35/CE). https://www.boe.es/buscar/doc.php?id=BOE-A-2006-13010

[43] Versión en línea en español: http://opendatahandbook.org/es/ versión 0.0.0 14 noviembre 2012

fuentes y volúmenes de datos. La misma *OKFN* ha lanzado *Open Data Commons* [44] de cara a conocer cómo facilitar y usar datos abiertos.

EL *PERIODISMO DE DATOS*: RECOPILACIÓN, ANÁLISIS, VISUALIZACIÓN

Según Elena Egawhary y Cynthia O'Murchu, autoras del libro "Data Journalism Book" (el libro del *periodismo de datos*)[45], editado por el *CIJ* (*Centre for Investigative Journalism* o *Centro por el Periodismo de Investigación*)[46], el *periodismo de datos* es "la capacidad de analizar y examinar números, de manejar conjuntos de datos y de leerlos correctamente" de cara a "encontrar y apoyar las historias" en las que se basa el periodismo, es decir, amplía las posibilidades de construcción de noticias y de los materiales en las que éstas se basan. Si bien su medio "natural" es la web, también se expresa en papel.

El *periodismo de datos* no solo *cuenta historias* a través de artículos, visualizaciones o infografías, también se refiere al uso de los datos para crear una historia, una visualización o representación de los datos que proponga esa u otras historias, o incluso una combinación de ambas técnicas, lo cual puede convertirse en una "aplicación informática" propia.

Va a ser difícil realizar *periodismo de datos* en la transmisión de una rueda de prensa o en la presentación de un informe que no admitan preguntas, o en la construcción de una noticia que adopte una nota de prensa de una entidad o empresa sin cuestionarse esos datos, sin una posterior evaluación, análisis y crítica. Hay que hacerse preguntas, y en este estadio del proceso periodístico también actúa el *periodismo de datos* ya que permite hacerse diversas preguntas al contar con diversos e innumerables datos. Siguiendo los ejemplos anteriores, quizás se pueda hacer una noticia, contar una historia, de las veces que tales personas o instituciones no han admitido preguntas en una rueda de prensa, o de las palabras empleadas en esas ruedas de prensa que no admiten preguntas, o de los datos del informe contratastados con otros datos de otras fuentes.

En un mundo donde el discurso periodístico valora la objetividad, bien sea de las declaraciones o de los hechos basados en datos, en la actualidad podemos acceder a un volumen de datos excepcional que nos sitúa ante un escenario nuevo. Si antes necesitábamos una hemeroteca, acceso a archivos diversos, etc., ahora contamos con más datos de los que la comprensión humana es capaz de analizar. Necesitamos apoyarnos en tecnologías de extracción, depuración, análisis, visualización y representación.

---

[44] Iniciativa de *OKFN* http://opendatacommons.org/

[45] "Data Journalism Book", Elena Egawary y Cynthia O'Murchu. Disponible en *PDF* en http://www.tcij.org/resources/handbooks/data-journalism

[46] Web de *CIJ*, http://www.tcij.org

Hay quienes utilizan la acepción *Data-Driven Journalism* para nombrar al *periodismo de datos* basado en datos abiertos o públicos[47], incluso Berners-Lee lo ha considerado el futuro del periodismo[48] pero normalmente son términos que se confunden y se ha popularizado *Data Journalism* más que *Data Driven Journalism*.

Mar Cabra y David Cabo son dos abanderados del *periodismo de datos* en España y representan la confluencia de saberes que entran en juego: periodismo e informática. Hace unos días comentaban en *Twitter* una frase de Richard Gingras, jefe de productos de noticias de Google, sobre el futuro del periodismo en la web de *El País*: *El reportaje de investigación del mañana no será la narración de una historia de 15.000 palabras sino que será un persistente reportaje de investigación con minería de datos y cruces de información*[49]. Gingras advería a los medios tradicionales sobre su funcionamiento, que no podía ser el mismo después de un siglo haciéndolo de la misma manera ya que estamos en un mundo de información en tiempo real, si bien reconocía que seguirán siendo necesarios "periodistas de calidad, la base para mantener la confianza de los usuarios, más incluso que las marcas tradicionales".

Paul Bradshaw,[50] periodista y profesor en las universidades *Birmingham City University* y en *City University London*, explicó el proceso del *periodismo de datos* en *la pirámide invertida*[51]: recopilación de datos, análisis y limpieza, contextualización, combinación/comparación/fusión/mezcla/remezcla y finalmente narración de los datos. Bradshaw cree que la forma más obvia de comunicar los datos es con narraciones visuales, visualizaciones, si bien caben otras formas de comunicar que resume en la pirámide de la comunicación[52]: visualización de datos a través de infografías, diagramas o aplicaciones interactivas; textos explicativos que acompañan a la visualización; uso de las redes sociales o de aplicaciones móviles; *humanización* de los datos, "tratarlos con respeto"; personalización, a través de los perfiles de usuario de las redes sociales o de la geolocalización, por ejemplo; y finalmente, uti-

---

47 José Luis Rodríguez escribe sobre *Data Driven Journalism* en "El nuevo periodismo se llama... OPEN DATA", disponible en su web http://www.territoriocreativo.es/etc/2011/02/el-nuevo-periodismo-se-llama-open-data.html

48 Declaración de Tim Berners-Lee en el lanzamiento de la iniciativa gubernamental de datos públicos en Londres, recogido por *The Guardian* en "Analysing data is the future for journalists, says Tim Berners-Lee" ("analizar los datos es el futuro para los periodistas, según Tim Berners-Lee", en traducción libre) http://www.guardian.co.uk/media/2010/nov/22/data-analysis-tim-berners-lee

49 Richard Gingras, jefe de productos de noticias de Google, entrevistado para *El País* en el foro global de los medios de *The Paley Center*: http://elpais.com/elpais/2012/05/08/videos/1336500464_908951.html

50 Paul Bradshaw, periodista y *blogger* inglés, especialista en *periodismo de datos*, autor del blog sobre periodismo http://www.onlinejournalismblog.com y de la investigación participativa http://www.helpmeinvestigate.com. Información en inglés en *Wikipedia*: https://en.wikipedia.org/wiki/Paul_Bradshaw_%28journalist%29

51 Artículo de Paul Bradshaw "The inverted pyramid of data journalism", en inglés: http://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/ y traducción al español: "La pirámide invertida del periodismo de datos" http://onlinejournalismblog.com/2011/07/08/the-inverted-pyramid-of-data-journalism-in-spanish/

52 Artículo de Paul Bradshaw "Six ways of communicating data journalism" http://onlinejournalismblog.com/2011/07/13/the-inverted-pyramid-of-data-journalism-part-2-6-ways-of-communicating-data-journalism/ y traducción en español, "Seis formas de comunicar Periodismo de Datos" http://onlinejournalismblog.com/2011/07/14/in-spanish-the-inverted-pyramid-of-data-journalism-part-2/

lidad, vía los perfiles de usuario de redes sociales o de aplicaciones específicas que trabajen con ciertos datos en evolución.

El proceso de recopilación de datos empieza por portales que ofrezcan los datos, de sitios que ofrezcan documentos, de la información disponible en las páginas web o bien de un archivo de datos que alguien haya recopilado haciendo, por ejemplo, *web scrapping* (rascando, escarbando de la web) de una más webs. Para poner esos datos en un documento, antes de limpiarlos, la forma más sencilla consiste en utilizar un programa que trabaje con hojas de cálculo (*Microsoft Office Excel*, *OpenOffice Calc*, *LibreOffice Calc* o *Google Drive*). Egawhary y O'Murchu reconocen que si bien muchos portales de datos ofrecen conjuntos de datos en *CSV* o *XLS*, con grandes volúmenes conviene utilizar gestores de bases de datos más potentes como *MySQL* o *PostgreSQL*[53]. En este sentido, Alex Howard[54], periodista del blog *Strata*[55] de *O'Reilly Media*, observa en el *periodismo de datos* un continuum en la narrativa que se ayuda de las innovaciones tecnológicas que comienza con *CAR*.

El uso de una hoja de cálculo es más común porque, por un lado, los datos que le interesan a un periodista no siempre están disponibles en *RDF*, donde se pueden analizar con herramientas semánticas, y por otro, porque las herramientas de análisis de datos *RDF* parecen más complejas de utilizar que una hoja de cálculo, reservadas a personas de perfil más técnico. Además, la misma herramienta con la que trabajamos en una hoja de cálculo nos puede servir para visualizar esos datos de diversas formas, y se comparten tutoriales para compartir sus experiencias con estas herramientas.

La herramienta más común para *web scrapping PDF scrapping* se trata de *Google Drive* (antiguo *Google Docs*), donde podemos realizar hojas de datos (*spreadsheet* o *datasheet*) que importen datos de la *Web* con una sencilla fórmula, tal como explica Bradshaw en "Scrapping for journalists"[56]. Por ejemplo, vamos a importar a la hoja de cálculo el contenido web que se encuentra en el interior de una tabla de datos, recogida por el elemento de *HTML table* de la tabla de *Campeones y subcampeones de los campeonatos de la Liga BBVA* que aparecen en esta página de la Wikipedia (http://es.wikipedia.org/wiki/Liga_BBVA). Para ello, tendremos que escribir:

=ImportHtml("http://es.wikipedia.org/wiki/Liga_BBVA";
"table"; 3)

En esta fórmula, *importHtml* es la función que importa los datos de una lista o tabla de una página web, por lo que irá a esa página web, buscará por una tabla y la

---

[53] *MySQL* y *PostgreSQL* son sistemas de gestión de bases de datos *opensource*. http://www.mysql.org y http://www.postgresql.org/

[54] Artículo de Alex Howard "The growing importance of data journalism" en el blog *Strata* de *O'Reilly Media* http://strata.oreilly.com/2010/12/data-journalism.html

[55] *Strata* es el blog de O'Reilly sobre datos. http://strata.oreilly.com/

[56] "Scrapping for journalism", de Paul Bradshaw. En https://leanpub.com/scrapingforjournalists está disponible un capítulo en *PDF*

primera que encuentre la importará en la hoja de cálculo; *table* es el elemento que ha de buscar; y 3, es el número de tabla del que debe extraer los datos.

He aquí un ejemplo de por qué resulta fundamental que los datos que se escriban en las páginas webs aparezcan bien estructurados, en los *elementos HTML* correspondientes y creados para tal fin.

APRENDIENDO JUNTOS

Las propuestas para aprender conjuntamente de *periodismo de datos* se encuentran en la esencia misma del proceso periodístico, ya que normalmente los periodistas trabajan colaborativamente en el proceso de análisis, limpieza y selección de datos y documentan el trabajo en webs, *wikis*, *blogs* u otras herramientas de escritura en red. Pero además, aparecen propuestas de encuentro y aprendizaje colectivo que van desde las clásicas conferencias y talleres a los innovadores hackatones [57] o barcamps[58].

Una iniciativa en este sentido se trata de *Hacks y hackers*[59], creado por Burt Herman, Aron Pilhofer (fundador de *DocumentCloud*[60]) y Richard Gordon. *Hacks y Hackers* propone un encuentro de periodistas (*hacks*, periodista en jerga callejera), informáticos (*hackers*, apasionados de la informática) y personas con conocimientos en ambos campos, para que hagan propuestas informativas y/o periodísticas. El encuentro se realiza en cada ciudad donde un grupo de personas quiere reunirse y lo publican en el sitio común, pudiendo realizar desde una simple charla a un taller sobre los distintos entornos de desarrollo de información dinámica creados con tecnología *JavaScript* (*js*). *Hacks y Hackers* reúne a personas que exploran las tecnologías disponibles para filtrar, visualizar y distribuir la infromación, y también aquellas que usan las tecnologías para encontrar y contar historias. Se produce una colaboración estrecha, una comunidad digital de personas que buscan inspirar y contagiar al otro, compartir información y código y participar en el futuro de los medios y del periodismo.

Si en todo el mundo son varias las universidades que se fijan en esta disciplina, en España es un fenómeno relativamente nuevo. Tanto que lo encontramos en noviembre de 2011 en el *V Congreso de Periodismo en Red*[61] celbrado en la *Universidad Complutense de Madrid* (*UCM*) o en la pasada edición de los *Cursos*

---

[57] *Hackaton*, neologismo compuesto por las palabras *hack* y maratón, encuentro donde se realizan una serie de tareas complejas en poco tiempo por un número determinado de personas.

[58] *Barcamp* es un neologismo que viene de *Foocamp*, un evento abierto y participativo organizado anualmente por O'Reilly Media para hablar informalmente de temas informáticos. En este caso, *BarCamp* se propone hablar de aplicaciones web, tecnologías de código abierto, protocolos de redes sociales o *periodismo de datos*.

[59] Web de *Hacks and Hackers*: http://www.hackshackers.com

[60] Web de *Document Clud*: http://www.documentcloud.com

[61] El *V Congreso de Periodismo en red* se celebró en la Facultad de Ciencias de la Información de la Universidad Complutense de Madrid. En su comité científico estuvo el Dr. Wenceslao Castañares http://congresoperiodismoenred.es/

de Verano del Escorial de la UCM[62]. También contemplan su estudio en el *Máster de Periodismo ABC UCM*[63] y en la primera edición del *Master en Periodismo de Investigación, Datos y Visualización*[64] impulsado por *El Mundo* junto con la *Universidad Rey Juan Carlos* (*URJC*) y *Google*. Por su parte, la *Universidad de Navarra* destaca por la formación de periodistas o infografistas[65].

De entre de todas las iniciativas de formación, destacaría, sin duda, por sus ponentes, sus temas tratados, su gratuidad y su calidad, la serie de conferencias y talleres organizadas por *Medialab Prado*[66]. Lo promueven un grupo heterogéneo de personas, entre las que se encuentran Cabra y Cabo, y se celebran periodicamente desde octubre de 2011 -continúan en la actualidad. Han participado algunas de las más importantes autoridades en la materia con el objetivo de formarse colectivamente en *periodismo de datos*, realizar propuestas de investigación y facilitar su práctica periodística. Reconocen, sin embargo, que se encuentran con dos obstáculos importantes: el primero, como ya hemos comentado, la ausencia de una ley de acceso a la información pública que apueste además por los principios del *Open Governemnt Data*, sin lo cual resulta complicado investigar pues los datos públicos son una fuente importante de datos; y el segundo, la capacitación que el ejercicio requiere, para lo cual ya se han puesto manos a la obra.

UNA NUEVA NARRATIVA CONTEMPORÁNEA

El *periodismo de datos* también supone una nueva forma de contar historias, una nueva narrativa visual que combina la visualización de los datos pero que también deja abierta la construcción de una narración propia del lector, lo que remite de nuevo a lo hipertextual, la web y la web semántica. Alberto Cairo expone en su libro "Infographics and visualization"[67] los dos principios que marcan que explican el *periodismo de datos*. El primero es la importancia de las visualizaciones [68]

---

[62] Información del *Curso de Verano de El Escorial* de *Periodismo de Datos*, información del blog *233 grados* de lainformacion.com http://233grados.lainformacion.com/blog/2012/05/la-universidad-complutense-ofrece-un-curso-verano-de-periodismo-de-datos.html

[63] *Master de Periodismo* del diario *ABC* http://www.abc.es/servicios/master

[64] *Master de Periodismo de Investigación* del diario *El Mundo* http://www.escuelaunidadeditorial.es/master-periodismo-de-investigacion.html

[65] Noticia del *XIII Congreso de Periodismo Digital de Huesca* por Idoia de Carlos, http://huesclick.wordpress.com/2012/03/26/huesclick-se-despide-del-xiii-congreso-de-periodismo-digital-de-huesca-tras-recibir-un-accesit-como-reconocimiento-a-la-cobertura-realizada/. La web del congreso, cuya XIV edición acaba de celebrarse: http://www.congresoperiodismo.com/

[66] El *grupo de trabajo de Periodismo de Datos Medialab* opera desde el 20 de octubre de 2011, tras un seminario organizado por *Medialab Prado*, *Access Info Europe* y la *Fundación ciudadana Civio* para promover el ejercicio de esta disciplina en España. http://medialab-prado.es/article/periodismo_de_datos_grupo_de_trabajo

[67] "The functional art", Alberto Cairo. http://www.thefunctionalart.com/

[68] ibídem, pág. 16

> (...) el cerebro no solo procesa la información que llega por los ojos. También crea imágenes visuales mentales que permiten razonar y planear acciones que permitan la supervivencia.

En segundo lugar, explica las diferencias entre infografías y visualizaciones de una manera muy clara, ya que si las primeras "cuentan historias diseñadas por periodistas", las visualizaciones ayudan a los lectores "a descubrir las historias por si mismos".

Es decir, el *periodismo de datos* no solo se apoya en datos para contar una historia sino que es consciente de que esa historia puede aportar lecturas distintas a la/s que el periodista quería transmitir en un primer momento. En realidad, todos los gráficos que muestran datos permiten cierto grado de exploración y de interpretación de esos datos, pero según los datos que presenten, el nivel de exploración es más o menos limitado. En este sentido, Ben Shneiderman apuntó que "el propósito de la visualización es la percepción, no las imágenes"[69]

Según Cairo, el propósito de los *diseñadores de información* es modelar todos esos datos en bruto con un sentido determinado, con el gráfico como herramienta para aumentar nuestras habilidades y capacidades adquiridas que nos permita ver, leer e interpretar más allá de lo que solemos hacer. Por ello es muy importante elegir el tipo de gráfico para representar y visualizar los datos, que disponga de las funcionalidades que nos permitan contar nuestra historia del conjunto de datos que hemos utilizado en la investigación, donde no se priorice la belleza o lo agradable sino el entendimiento del mensaje.

Cairo propone tres reglas de oro para realizar una visualización de datos. La primera consiste en utilizar los datos de forma diversa ya que en la mayoría de los casos una sola forma gráfica no es suficiente para contar la historia por completo. La segunda pone el acento en los datos que pueden resultar inesperados, los que salen de la norma, los relevantes -que relevan algo nuevo-, verdades que no aparecerían de otra manera. Y la tercera e igualmente importante, acompañar información textual a los gráficos, desde el titular a destacados, o bien distintas capas de contenido que ofrezcan un contexto para la información.

Edward Segel y Jeffrey Heer[70] identifican las dimensiones relevantes de la historia visual, incluyendo que técnicas gráficas e interactivas pueden cumplir varios niveles de flujo estructural y narrativo. Describen asimismo siete géneros de la visualización narrativa: estilo revista, diagrama con anotaciones, cartel dividido, diagrama de flujo, historieta, presentación de diapositivas y vídeo.

---

[69] "The purpose of visualization is insight, not pictures", Stuart Card, Jock Mackinlay y Ben Shneiderman, "Readings in Information Visualization: Using Vision to Think", Londres, Academic Press, 1999.

[70] "Narrative Visualization: Telling Stories with Data", Edward Segel and Jeffrey Heer, 2010. http://vis.stanford.edu/papers/narrative

CONCLUSIONES

He tratado una aproximación al *periodismo de datos*, poniendo el énfasis en la *Web Semántica* y *Datos Vinculados*, sin olvidar los precedentes que desde el campo del periodismo y de la organización de la información se han producido, si bien he realizado una visión panorámica que merece la pena explorar con mayor atención. No obstante, he dejado de lado el impacto de los gráficos, de la infografía, su evolución y contaminación con el campo estadístico y con la ingeniería de sistemas; no he incluido más que un ejemplo de cómo prácticar *periodismo de datos* y no he citado los muchos ejemplos que se producen, si bien siguiendo los enlaces que propongo en las notas se pueden descubrir herramientas, investigaciones y proyectos de gran interés; no he trasladado el debate alrededor de la evolución de la web y HTML5; no he tenido en cuenta la evolución tecnológica en los dispositivos, fijos o móviles... Quizás, también, habría sido más interesante analizar el impacto que actualmente está consiguiendo el *periodismo de datos* en su conjunto.

Simon Rogers, editor de *The Guardian*, escribió que el *periodismo de datos* era el nuevo fenómeno *punk*[71], siguiendo la filosofía del *hazlo tú mismo*, ya que al igual que las bandas de música *punk*, donde cualquiera podía utilizar tres acordes que servían y bastaban para componer una canción y todo un álbum, en el *periodismo de datos* tienes unos datos por aquí, otros datos por allá y un conjunto de herramientas gratuitas con los que trabajar, solo has de contar con ideas sobre las que trabajar las investigaciones periodísticas. Pongámonos manos a la obra, practiquemos, empapémonos y abarquemos esas cuestiones en el futuro.

[71] "Anyone can do it. Data journalism is the new punk", artículo de Simon Rogers en *Data Blog* de *The Guardian*, 2012 http://www.guardian.co.uk/news/datablog/2012/may/24/data-journalism-punk

RESUMEN

Uno de los fenómenos más interesantes del periodismo contemporáneo es el denominado *periodismo de datos*, donde la evolución del periodismo asistido por ordenador y las representaciones de datos han puesto la atención en la usabilidad, la interacción, la visualización y la participación de los usuarios. El trabajo periodístico se ve alterado desde el inicio, produciéndose una extraordinaria colaboración entre periodistas y también la cooperación con diseñadores e informáticos, produciendo nuevas narrativas visuales para artículos o reportajes a partir de la utilización de un gran volumen de datos, muchos de ellos provenientes de la web semántica, una revolución cultural sobre la propiedad y uso de los datos que afecta a los procesos de producción de información y conocimiento.

Palabras clave: Periodismo, periodismo de datos, web, WWW, Web Semántica, Linked Data, Datos Vinculados, Datos Abiertos Vinculados, Datos Abiertos, Datos Públicos, Open Data, Open Government, Open Government Data, periodismo de precisión, periodismo asistido por ordenadores, periodismo de investigación, investigación periodística, colaboración, cooperación, talleres, cursos, narrativa, narrativa visual, visualización, infografía, interacción, usabilidad, barcamp, hackaton, HTML, RDF, XML, JS, CSV

ABSTRACT

One of the most interesting phenomena of contemporary journalism is called *Data Journalism*, where the evolution of computer-assisted reporting and data representation have put attention on usability, user interaction, visualization and user participation. have focused attention to usability, user interaction and user participation. The journalism is altered from the begining, producing an extraordinary collaboration between journalists and also cooperation with designers and hackers, producing new visual narratives for the stories that come from big datasets, many of them from Semantic Web, a cultural revolution about ownership and data use affecting to production processes of information and knowledge.

Keywords: Journalism, data journalism, web, WWW, Semantic Web, Linked Data, Open Linked Data, Open Data, Open Government Data, precision journalism, data driven journalism, computer-assisted journalism, investigative journalism, investigative reporting, collaboration, cooperation, courses, workshops, narrative, visual narrative, visualization, infographics, interaction, usability, barcamp, hackaton, HTML, RDF, XML, JS, CSV

RÉSUMÉ

L'un des phénomènes les plus intéressants du journalisme contemporain est appelé le journalisme de données, où l'évolution du journalisme assisté par ordinateur et de représentations de données ont attiré l'attention sur la convivialité, l'interaction, la visualisation et la participation des usagers. Le journalisme est altérée dès le début, en produisant une extraordinaire collaboration entre les journalistes et la coopération avec des designers et informatiques, la production de nouveaux récits visuels pour les fonctionnalités de l'utilisation d'un grand volume de données, beaucoup d'entre eux de la Web sémantique, une révolution culturelle sur la propriété et l'utilisation des données qui affecte les processus de production des informations et des connaissances

Mots clé: Journalisme, le journalisme de données, web, WWW, Web sémantique, Linked Data, Linked Data, Linked Open Data, Open Data, journalisme de données publiques, Open Data, gouvernement ouvert, données gouvernementales ouvertes, journalisme de précision, les rapports assistés par ordinateur, recherche, journalisme d'investigation, collaboration, la coopération, des ateliers, cours, contes, narration visuelle, visualisation, infographie, l'interaction, convivialité, HTML, RDF, XML, JS, CSV.