

Big Data: una perspectiva desde la comunicación ciudadana

Han sido muchas las metáforas que han ido marcando el desarrollo de una sociedad de la información desigual y tan inclusiva como la sociedad posindustrial que la viera nacer. Big Data es hoy la metáfora ganadora. Para entender su alcance debemos poner en contexto su nacimiento y evolución, así como su efecto específico sobre los nuevos procesos de participación, comunicación y acción social en la Red.

Eso pasa por entender que estamos ante un cambio de escala, en lo grande y en lo complejo, del escenario sociotécnico que constituye nuestro hábitat. La especificidad de las tendencias asociadas, como el Small Data, da cuenta de algunas de las características clave de ese escenario que han apalancado la evolución pareja del Periodismo Ciudadano. Es en esa especificidad donde radica la complejidad de tal fenómeno y, por tanto, su potencialidad.

“Traed madera ¡Es la guerra!”

Introducción

Si bien se trata de una pregunta trampa, hay pensadores que afirman, no sin razón, que somos “doblemente digitales”: por la codificación universal que utilizamos en la Red y por el uso cada vez más intensivo de nuestros dedos para acceder a la misma (1).

Nuestra huella digital nos delata: un reguero de datos ofrece pistas imborrables de nuestra actividad en esa infraestructura ubicua, invisible y universal caracterizada, entre otras cosas, por la rastreabilidad de las acciones sociotécnicas que realizamos sobre ella.

Quiero reflexionar en este artículo sobre lo grande y lo complejo. El fenómeno del ‘Big Data’ es la metáfora ganadora en la batalla retórica que domina el desarrollo de nuestra sociedad de la información.

En cualquier informe de tendencias, en cualquier ámbito y a todos los niveles, ‘Big Data’ ocupa ya un lugar privilegiado. En cierta forma, viene a ofrecer una vía para “decanter” el sedimento más sustancioso del proceso evolutivo de los medios sociales para la información, la relación y la comunicación en la Web 2.0 (IRC 2.0).

Este tipo de medios, específicamente representados por la explosión de las redes sociales, capitalizadas –hacia delante y hacia atrás–, a partir de sus contenidos primero y su conectividad ubicua después, han catalizado nuestra característica avidez por la información que a su vez ha dado lugar a la proliferación de su representación técnica más inocua, el dato.

Desde el punto de vista que nos interesa aquí, el de la comunicación ciudadana y la participación activa de esa ciudadanía en los procesos informativos, es importante entender cómo se consolida un escenario como el actual y, sobre todo, cómo se comienzan a dimensionar los

(1)
“Nativos digitales, inteligencia digital ¿Homo digitalis?”. Fernando Sáez Vacas, TELOS 86, ene.-mar. 2011. Disponible en el Archivo Digital de la Universidad Politécnica de Madrid, <http://oa.upm.es/23261/>

diferentes atributos de este fenómeno: cómo se diluye rápidamente el interés por el volumen de este fenómeno y se comienza a hablar de '*Small Data*' buscando la potencialidad de la hiperlocalidad; y cómo, frente a las aplicaciones comerciales más evidentes, crece el interés por explorar ámbitos de aplicación más directamente relacionados con el desarrollo humano que con el interés comercial más evidente.

Todo conectado, en su contexto

Big Data no es un fenómeno aislado, autónomo, ni siquiera desde el punto de vista de la retórica. Hemos seguido una evolución de términos vacuos y discursos sesgados y desinformados que apenas se sostienen sobre unos datos estadísticos endebles; que encontramos asociados a otros lugares comunes, como la informática en la nube o la Internet de las cosas. Todos esos elementos suman y contribuyen a nuestro entendimiento de un fenómeno complejo.

La informática en la nube *Cloud Computing* ha venido para quedarse y ha conseguido que el volumen de información que los usuarios "confían" a la Red crezca más rápido que nunca. Ejemplos evidentes como *Dropbox* han acelerado la proliferación de servicios de almacenamiento, como *Google Drive*.

La era de los datos se sostiene sobre una infraestructura nebulosa, la nube informática. Podemos hablar de la infraestructura en sí misma (*IaaS, Infrastructure-as-a-Service*), de una plataforma (*PaaS, Platform-as-a-Service*) en sentido amplio, como puede ser el propio entorno de desarrollo o del software (*SaaS, Software-as-a-Service*) como modelo para la distribución de servicios. La virtualización ha alcanzado a la gestión de todo tipo de recursos en la Red: comunicación, potencia de cálculo (computación) y almacenamiento.

El hecho de que la gestión de esos recursos haya llegado a la Red ha convertido a las nubes públicas, privadas o híbridas en un nivel más de abstracción a la hora de analizar el ecosistema digital que hemos convenido en llamar sociedad de la información.

Los responsables de la gestión de las tecnologías de la información en las organizaciones empresariales de todo tipo deben balancear la flexibilidad de este tipo de servicios con los requisitos legales de los servicios que sus compañías comercializan en un entorno cada vez más competitivo.

El emprendimiento de base tecnológica ha conocido un nuevo renacer en un escenario capaz de proporcionar a los innovadores un "tejido" técnico mucho más fiable para lanzarse a la aventura de hacer realidad sus ideas. La realización de productos "mínimamente" viables, así como pruebas de concepto y de mercado está hoy al alcance de cualquier emprendedor gracias, en gran parte, a la consolidación de esa nube informática.

La Internet de las cosas ha conseguido hacerse un nombre como tendencia de futuro. Generalmente asociada a iniciativas diversas provenientes de ámbitos como el de las *Smart Cities*, han descubierto un filón en forma de un universo "sensorizado" que es capaz de cuantificar prácticamente todo, apalancando directamente el fenómeno del '*Big Data*'.

Tal y como podemos leer en la ilustración de tendencias reproducida en la figura 1 y extraída del informe "eEspaña 2014" publicado recientemente por Fundación Orange, "*relojes, gafas, zapatillas y cada vez más elementos de nuestra indumentaria se conectan, como tantas otras "cosas" (vehículos, aéreos y terrestres, edificios, etc.), a una Internet que les confiere el atributo de la inteligencia que los humanos ponemos en gran parte de nuestra tecnología. Personas y cosas que interactúan con diferente grado de simbiosis, en un hogar más digital, en una ciudad más "inteligente", ...*

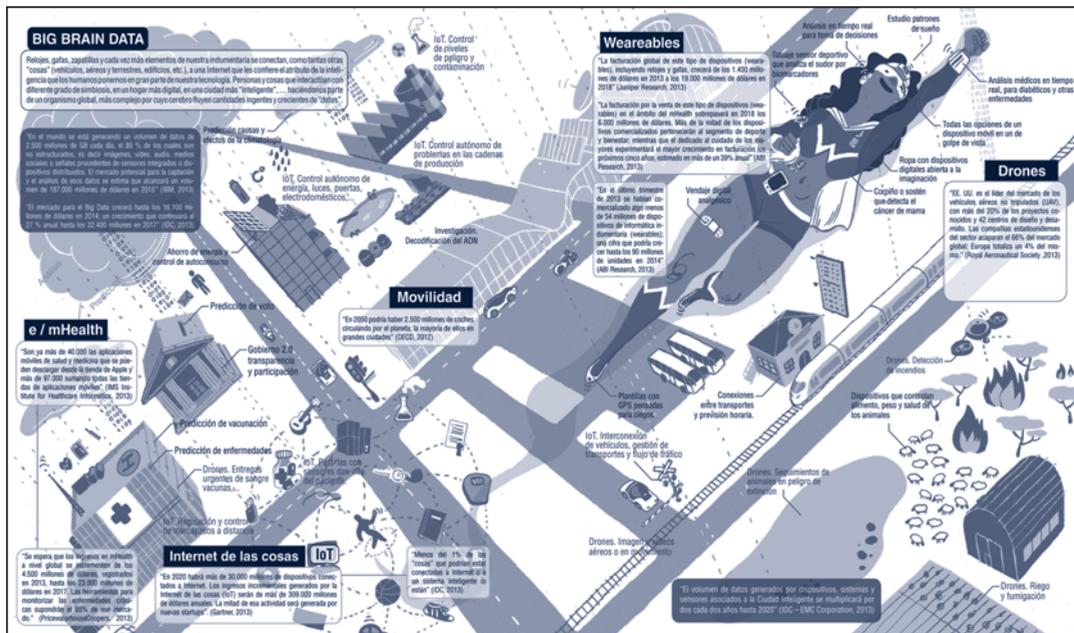


Figura 1. Ilustración de tendencias. Tomada del informe eEspaña 2014 (Fundación Orange).

haciéndonos parte de un organismo global, más complejo por cuyo cerebro fluyen cantidades ingentes y crecientes de datos".

La inclusión de más capacidades y funcionalidad en cada vez más objetos ha corrido pareja a la evolución de la característica "protésica" de los dispositivos móviles inteligentes como interfaz por defecto con la Red.

En un escenario de ese tipo, en una ciudad inteligente, todo está conectado: los edificios, los coches, nuestro hogar, nuestra ropa, ¿nosotros? La proliferación, por tanto, de sensores empieza por nuestra propia interacción con la Red, con otras personas, con otros objetos a través, por ejemplo, de nuestros *smartphones*.

La informática indumentaria y los dispositivos "vestibles" (*wearable computing*) han cobrado un protagonismo muy relevante en este escenario que apenas comienza a realizarse. Las proyecciones de los analistas de mercado afirman que solo un escaso 1% de los objetos que podrían estar conectados a Internet lo están realmente; fundamentalmente alimentando el ascenso de las expectativas de las comunicaciones M2M (*Machine-to-Machine*).

(2) "Informe 50 mejores Apps de Salud en español", disponible en formato PDF en la siguiente dirección <http://madrid.theappdate.com/wp-content/uploads/2014/03/Informe-TAD-50-Mejores-Apps-de-Salud.pdf>

(3) "V Estudio Anual IAB Spain Mobile Marketing", septiembre 2013. Disponible en la siguiente dirección http://www.iabspain.net/wp-content/uploads/downloads/2013/09/V_Estudio_Mobile_Marketing_vision_corta.pdf

Con una población envejecida y cada vez más preocupada por el bienestar físico, las aplicaciones desarrolladas en el ámbito de *mHealth* cobran especial importancia, tal como atestiguan los datos del reciente informe de *The App Date* (2).

El despliegue de infraestructuras de banda ancha móvil y la popularización de los dispositivos móviles inteligentes (*Smartphones*) han conseguido que proliferen los servicios para la captura, edición y publicación de contenidos audiovisuales en la Red.

En 2014, por primera vez, se venderán en el mundo más *smartphones* que teléfonos móviles tradicionales, según estimaciones de los analistas de *Gartner Group*. Los datos en España (3) hablan de una penetración creciente -que se sitúa ya en un nada despreciable 80%- de los cada

vez menos teléfonos móviles inteligentes. Un dispositivo, convertido en segunda pantalla en el cara a cara con otros medios tradicionales u *offline*, comienza, además, a compartir protagonismo con las tabletas.

Con un mercado creciente que se concentra alrededor de los dos jugadores más fuertes, *Google* –que gana en número de descargas y dispositivos– y *Apple* –que lidera la facturación a nivel mundial–, la *App Economy* se ha convertido en algo más que una metáfora, moviendo flujos financieros a nivel mundial, además de crear yacimientos de empleos considerablemente productivos.

La industria del hardware se acelera para producir dispositivos más funcionales, potentes y versátiles: las tecnologías de integración *System-on-Chip* (SoC) y las arquitecturas '*multicore*' se estandarizan para nutrir a un ecosistema más competitivo en el que comienzan a consolidarse las primeras asociaciones industriales, como la *Open Handset Alliance*, que apuestan por una estandarización más amplia.

La consolidación de la arquitectura web, por su parte, ha dinamizado el escenario inicialmente polarizado por los ecosistemas nativos de dos gigantes de Internet como son *Google* y *Apple*. Las *WebApps* hacen sombra a las aplicaciones nativas dentro de una batalla por la atención del usuario que se libra cada vez más en el ámbito móvil; y los sistemas operativos abiertos se abren su hueco al calor de la iniciativa de *FirefoxOS*, de *Mozilla*.

Por otro lado, la popularización de las redes sociales, como la realización práctica más relevante de los medios sociales para la información, la relación y la comunicación en la *Web 2.0*, ha provocado que la cantidad de información personal de los usuarios registrada en múltiples bases de datos se convierta en un filón nada despreciable para las empresas.

El ejemplo paradigmático más reciente es *Facebook*, que una vez cotizada en bolsa, ha identificado el *Big Data* como uno de sus grandes desafíos, junto con su expansión internacional donde la competencia o la prohibición no le permiten crecer con la rapidez que demandan sus accionistas.

Los patrones de comportamiento de los usuarios de esas plataformas han convertido cada una de sus interacciones, con otras personas, con piezas de contenido, con objetos, con organizaciones de todo tipo, con marcas y productos en fuente potencial de datos, el nuevo oro de la Red.

Las tecnologías relacionadas con el tratamiento masivo de datos (4) se han convertido en un lugar común para los analistas de tendencias. Las denominaciones de algoritmos (*MapReduce*), herramientas e implementaciones diversas (*Hadoop*, *Hive*) han saltado a los titulares de los medios sectoriales y generalistas reactivando sectores olvidados como los tradicionales *Data Mining/Warehouse* o *Business Intelligence*.

Los perfiles profesionales formados en la "Ciencia de los Datos" son objeto de deseo para los gigantes de la Red. Proliferan los cursos masivos online –los populares MOOC–, sobre esta disciplina, de la misma manera que los acuerdos entre la Industria y la Academia para darle forma a un programa de desarrollo curricular necesariamente híbrido, inter y multidisciplinar.

(4)

En la siguiente dirección se puede acceder a un resumen muy sintético de las principales tecnologías que se utilizan hoy en este ámbito poco conocido, <http://centeropenmiddleware.blogspot.com.es/2014/01/pildoras-com-tratamiento-masivo-de-datos.html>

Evolución

En algo más de una década de supuesta conversación hemos visto pasar por las estanterías de las librerías de puente aéreo todo tipo de evangelios "dos-punto-cero", generalistas y sectoriales, que pontifican sobre la

buena nueva, de los blogs primero, de las redes sociales después, de los dispositivos móviles cada vez más inteligentes o de cómo la metafórica nube informática nos devolvía un tsunami de datos y metadatos que nosotros mismos estábamos generando consciente o inconscientemente.

Cada palabra clave, cada metáfora, cada nuevo giro de una superficial retórica, convertida en mal entendido paradigma, se ha convertido en un arma de destrucción masiva en manos de gurús, consultores y mercenarios de todo tipo que han medrado en una primigenia sociedad del conocimiento. Sembrando a su paso la destrucción y la confusión más absolutas; tras la tormenta, la calma y la reflexión tenaz que ha conseguido sedimentar algunas enseñanzas, unos pocos aprendizajes.

Tal como señalaba en (Fumero, 2011), el fenómeno “dos-punto-cero”, que se popularizaba a partir de los años 2002 a 2004 tras el colapso provocado por la locura financiera asociada a los valores de las empresas de base tecnológica que centraban sus actividades en Internet, propugnaba cierto “cambio de versión”, adoptando la retórica informática, con un profundo impacto transformador que se irían traduciendo, lentamente, en cambios drásticos en diferentes industrias, con ritmos también muy distintos.

En España, este fenómeno se consolidaba hace apenas un lustro alrededor de los blogs, investidos a principios de siglo de cierto carácter paradigmático como quintaesencia del supuesto impacto “democratizador” de la Red. En un contexto extremadamente complejo, crecía la disponibilidad masiva de herramientas -generalmente construidas alrededor del formato blog y/o wiki y algunas tecnologías asociadas, como los gestores de contenidos (*CMS, Content Management Systems*), los formatos estándares para la sindicación de contenidos, o los microformatos semánticos, tagsonomías o folksonomías.

Esas tecnologías se veían amplificadas funcionalmente por un escenario en el que se disparaba el número, la variedad y accesibilidad de dispositivos móviles e inalámbricos, además de las conexiones de banda ancha, fija y móvil, que se abarataban fruto, en gran parte, de la competencia salvaje en precios a la que el marco regulatorio obligaba a las tradicionales *Telcos*.

Aprendimos con el fenómeno *blog* el impacto que podían tener los efectos de red al “liberalizar” los medios para la producción informativa, editorial e incluso audiovisual. El cambio de escala ha transformado por completo industrias enteras, como es el caso de los medios de comunicación de masas.

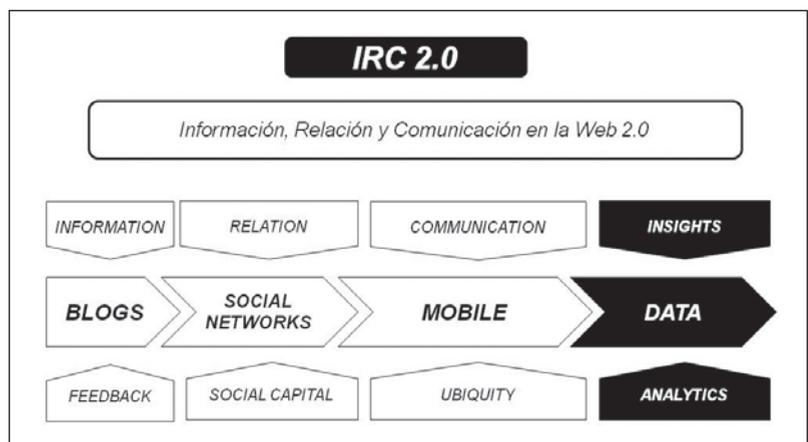


Figura 2. Evolución de los medios para la información, la relación y la comunicación en la Web 2.0 (elaboración propia).

Las redes sociales nos enseñaron que la oportunidad (creativa, innovadora, de negocio en definitiva) se desplazaba realmente de los nodos a las conexiones. En un escenario plagado de dispositivos móviles cada vez más “inteligentes” y de sensores integrados en todos sus ámbitos, el poder de nuestra humanidad (léase tecnología) se desplazaba hasta la punta de nuestros dedos. Se extendía así la característica “protesicidad” de una red que no puede ser sino universal y digital.

La evolución que se muestra esquemáticamente en la figura 2 quiere ilustrar cómo ha cobrado relevancia el fenómeno de los datos, el metafórico ‘Big Data’ en el desarrollo de nuestra sociedad de la información a raíz de la consolidación de otros fenómenos relacionados en el ecosistema Internet.

Lo que trataré de hacer a continuación es desgranar las distintas vertientes de este fenómeno desde el punto de vista del periodismo ciudadano en particular y de la comunicación ciudadana y la acción participativa en general.

Un modelo

La reflexión que propongo en este breve artículo se sustenta sobre una base sistémica que pretende servir como marco para abordar una situación de complejidad concreta, o más específicamente un tipo de situaciones que conforman ese ámbito de la comunicación ciudadana en el que ha crecido el fenómeno del periodismo ciudadano.

Esa base sistémica y compleja forma parte de un marco conceptual más amplio que ha sido bautizado como STIC (SocioTecnología de la Información y Cultura) y descrito a partir de una aplicación particular dentro del área de la tecnología educativa en el libro “Argumentación Didáctica de una Innovadora Cultura Socioinfotecnológica” (5).

El modelo (figura 3) se ha descrito esquemáticamente a partir los tres procesos básicos que contempla, información, relación y comunicación.

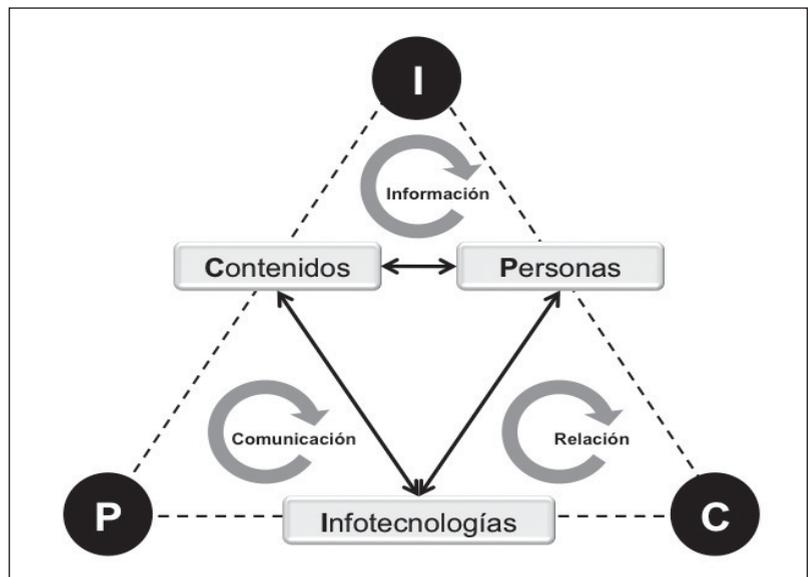


Figura 3. Información - Relación - Comunicación (IRC) (elaboración propia).

(5) Disponible en formato PDF para su descarga gratuita en el sitio web de Publixed Editorial, en la dirección <http://www.libreriapublixed.com/digital/details/1215/12/productos-digitales/digital-argumentacion-didactica-de-una-innovadora-cultura-soci.html>

Esos procesos se representan en el plano formado por otro trío de elementos, Infotecnologías, Personas y Contenidos (IPC) extraídos de una interpretación interesada de la teoría de la actividad más tradicional.

Este modelo –con sus evidentes limitaciones–, quiere servir aquí como herramienta conceptual para una aproximación crítica a las dimensiones de lo grande y lo complejo habitualmente asociadas al fenómeno del ‘Big Data’.

Big Data. Sobre lo grande y lo complejo

Evidentemente, ya lo decía más arriba, la retórica va muy por delante de la realidad vigente en nuestra era de Internet y este término, así como el interés de los divulgadores y “expertos” de todo tipo por su rentabilización, no son nuevos.

New York Times hablaba de la era del ‘Big Data’ en febrero de 2012 con un artículo de análisis que atraía numerosas reacciones (6); *The Economist* publicaba un especial sobre este tema en febrero de 2010 (7); *Nature* lo hacía en septiembre de 2008 con un número especial (8), lo mismo que la revista *Wired* (9). Siendo así, este tema no ha pasado desapercibido para los cazadores de tendencias -apareciendo pronto en el radar de *O’Reilly* (10)–, ni para las consultoras multinacionales, como *PWC* (11).

La propia definición de ‘Big Data’ resulta intencionadamente elusiva y ambigua (12), mostrándose lo suficientemente flexible y maleable como para que siga manteniendo la misma vigencia que mostraba en los primeros informes de los analistas de *Gartner* cuando aun eran *META Group*, allá por 2001, y que siguen manteniendo como modelo de análisis hoy (13), que en las portadas más recientes de cabeceras como *NYT*.

Más allá de lo evidente, el término hace referencia no solo al volumen y/o a la velocidad con la que crece, sino también a la complejidad de las

(6) “*The Age of Big Data*”, disponible en la dirección <http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html>

(7) “*Data, data everywhere*”, disponible en la dirección <http://www.economist.com/node/15557443>

(8) “*Big Data*”, disponible en la dirección <http://www.nature.com/news/specials/bigdata/index.html>

(9) “*The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*”, disponible en la dirección http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

(10) “*Release 2.0*”, febrero 2009, disponible en la dirección <http://cdn.oreilly.com/radar/r2/r2.0.11excerpt.pdf>

(11) “*Technology Forecast, 2010 Issue 3. Making Sense of Big Data*”, disponible en la dirección <http://www.pwc.com/us/en/technology-forecast/2010/issue3/index.jhtml>

(12) En la propia Wikipedia podemos ver las diferentes vertientes de la misma en la historia reciente del término, http://en.wikipedia.org/wiki/Big_data#Definition. También vale la pena consultar el artículo de IBM sobre el tema en <http://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>

(13) “*3D Data Management. Controlling Data Volume Velocity and Variety*”, disponible en la dirección <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

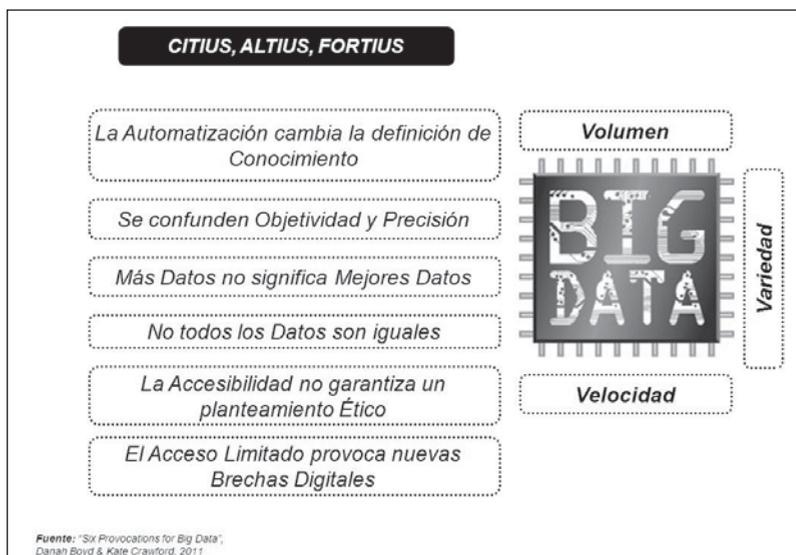


Figura 4. Las tres dimensiones del Big Data como un problema de gestión de TI frente a seis “provocaciones” propuestas por D. Boyd y K. Crawford (elaboración propia).

relaciones establecidas entre la variedad de datos, recogidos en diferentes colecciones, bases de datos, etc.; cobrando una especial relevancia socioeconómica gracias a la consolidación de una serie de fenómenos y tendencias ampliamente analizadas en los últimos años en términos del desarrollo de nuestra sociedad de la información.

Danah Boyd y Kate Crawford llamaban la atención en 2011 sobre este fenómeno lanzando seis “provocaciones” en forma de reflexiones que tienen mucho que ver con la naturaleza sociotécnica del mismo. Las recuperaré aquí brevemente con la intención de que sirvan para introducir los aspectos a los que me refería más arriba.

La automatización cambia la definición de conocimiento. Las propias autoras citaban en su artículo a Chris Anderson (2008) que ironizaba -bajo el título de “*The End of Theory*”-, sobre la relevancia que habían cobrado los datos, sugiriendo que nos olvidáramos -en referencia a las ciencias sociales-, de la “la taxonomía, la ontología, la psicología”: “los datos hablarán por sí mismos”, decía.

Se confunde objetividad con precisión. La tradicional crítica a la excesiva obsesión por cuantificar las Ciencias Sociales adquiere con el fenómeno del *Big Data* una nueva dimensión. La objetividad definida desde el punto de vista de la Filosofía de la Ciencia se confunde, de hecho, con la potencial precisión cuantitativa que ofrecen los datos; y lo cierto es que ni siquiera los datos pueden representar, desde el punto de vista metodológico, una “verdad objetiva”.

Disponer de más datos no significa tener mejores datos. Desde la óptica del investigador social, más allá de la metodología, se podría caer en la tentación de considerar que los datos que provienen de la API de cualquiera de las plataformas de redes sociales más populares (léase *Facebook* o *Twitter*) se pueden considerar como representativos de una amplia muestra de la sociedad. Sin embargo, cualquiera de nosotros es perfectamente consciente de que, por un lado estas herramientas se usan intensivamente por parte de una legión de ‘*early-adopters*’ que no resultan en absoluto representativos socialmente y, por otro lado, la cantidad y calidad de los datos que aquellas API permiten extraer es muy limitada con respecto al volumen real de las bases de datos que gestionan.

No todos los datos son iguales. Los datos, despojados de su “contexto”, no valen nada. No podemos pensar que el análisis de “redes sociales” ha nacido con la popularización de los medios para la información, la relación y la comunicación que comúnmente conocemos con esa denominación en Internet. Lo que otrora se investigaba con un enfoque de ‘*small data*’ no va a producir necesariamente mejores resultados con solo adoptar un enfoque basado en ‘*Big Data*’.

La accesibilidad (de los datos) no garantiza un comportamiento ético (desde el punto de vista del análisis que se realiza sobre ellos). La evolución que nos ha llevado a vivir la proliferación de *blogs* y publicaciones personales en la Web primero y de perfiles y datos personales después, que se han convertido en aún más ubicuos, si cabe, gracias al móvil, ha hecho que sea relativamente sencillo acceder a grandes conjuntos de datos. La disponibilidad de los mismos para su tratamiento no garantiza que éste siga un criterio ético. Es fácil caer en el error de que si un conjunto de datos es accesible da igual el tipo de proceso, algoritmo o tratamiento que hagamos con esos datos.

El acceso limitado provoca nuevas brechas digitales. La acostumbrada escasez que validaba las aproximaciones economicistas a la investigación social parece superada gracias, precisamente, a la abundancia de datos.

Pero esa abundancia no debe llevarnos al engaño: es el acceso a los conjuntos de datos lo que podemos limitar. Podemos hacer también accesibles unos datos u otros, creando -intencionadamente o no-, nuevas brechas digitales.

Periodismo y datos se encuentran

Saber documentar las historias que comunicamos siempre ha sido importante en el proceso periodístico e informativo. El necesario acceso a bases de datos diversas y las habilidades y herramientas que requiere su adecuado tratamiento no le son ajenas a profesional alguno. Pero ha sido el fenómeno sustantivo del *Big Data* el que ha venido a validar la pujanza de una ocupación que cobra naturaleza de disciplina periodística: el periodismo de datos (*Data Journalism*).

Data can be the source of data journalism, or it can be the tool with which the story is told — or it can be both. Like any source, it should be treated with scepticism; and like any tool, we should be conscious of how it can shape and restrict the stories that are created with it (14).

Esa reflexión, de Paul Bradshaw, la recoge la primera versión, en inglés, del Manual de Periodismo de Datos; y me servirá aquí para recuperar algunas consideraciones acerca de esta línea de desarrollo profesional que se impone en el periodismo actual.

Efectivamente, hay que relativizar este tipo de actividad. Periodismo y Datos tenían que encontrarse, con más motivo en plena era del *'Big Data'*. Pero debemos entender que ese encuentro es una circunstancia de contexto, no específica.

Como bien dice el profesor Bradshaw, los datos pueden ser “simplemente” la fuente para la historia que quieres contar, haciendo, por ejemplo, necesaria una labor de explotación de bases de datos en la fase de investigación y documentación; o bien pueden ser la herramienta para contar la historia... o ambas cosas, como puede observarse en la creciente tendencia a la elaboración de piezas infográficas en los medios actuales de todo tipo.

Traigo este argumento aquí con la intención de recuperar la aproximación que llevaba a periodismociudadano.com en 2011 (15) a raíz de un artículo que introducía el concepto de “*Computational Journalism*” en los siguientes términos.

“Researchers and journalists are exploring new methods, sources, and ways of linking communities to the information they need to govern themselves. A new field is emerging to promote the process: computational journalism. Broadly defined, it can involve changing how stories are discovered, presented, aggregated, monetized, and archived. Computation can advance journalism by drawing on innovations in topic detection, video analysis, personalization, aggregation, visualization, and sensemaking.”

Habiendo transcurrido más de un lustro desde que se elaboraran las líneas de desarrollo básicas de esa disciplina así definida, creo que resulta igualmente útil su planteamiento original como aproximación a una definición más amplia -y posibilista-, como catalizador para lo que debe ser entendido como un espacio virtual de oportunidades para la innovación social.

El oficio de periodista -tal y como afirmaba en el mismo sitio de periodismociudadano.com hace poco más de un año (16)- y su relación especial con la información y la sociedad red en que se sustancia hoy, así

(14) Definición debida a Paul Bradshaw, (*Birmingham City University*) que aparece en el *'Data Journalism Handbook'*, accesible en la dirección http://datajournalismhandbook.org/1.0/en/introduction_0.html

(15) “*Computational Journalism ¿periodismo o informática?*”, disponible en la dirección <http://www.periodismociudadano.com/2011/10/03/computational-journalism-%C2%BFperiodismo-o-informatica/>

(16) “*Periodista y desarrollador ¿escribes o programas?*”, disponible en la dirección <http://www.periodismociudadano.com/2013/03/04/periodista-y-desarrollador-escribes-o-programas/>

como la ciudadanía activa que debería caracterizar tal construcción social, se acercan más si cabe a las habilidades necesarias para construir con las herramientas cada vez más accesibles que nos proporciona la ingeniería del software.

Retórica

Detrás de afirmaciones como que “los datos son el nuevo petróleo” o la asimilación de este fenómeno con una nueva “revolución industrial” de naturaleza digital, además de la consolidación de cierta tendencia más o menos sistémica, hay un notable ejercicio de retórica.

Tal y como afirmaba *Danah Boyd* (17) hace apenas un par de años (18), “*We’ve entered an era where data is cheap, but making sense of it is not*”. En esa búsqueda de sentido encontramos diferentes ejercicios de retórica a la caza de otra metáfora ganadora.

Hace casi un año, en agosto de 2013, los analistas de *Gartner Group* – considerado como uno de los oráculos de la desinformación corporativa a más alto nivel-, dedicaban su conocido mapa de tendencias o ‘*Hype Cycle*’ a la relación Humano-Máquina (19), ilustrando cómo algunos tópicos convertidos en lugares comunes (la ludificación, la impresión 3D para el gran consumo o los dispositivos de electrónica indumentaria), entre los que se encontraba el fenómeno del ‘*Big Data*’, se encontraban en lo que ellos consideran el “pico” de las expectativas generadas a su alrededor.

Precisamente, porque estamos llegando a ese pico en el que la inflación de expectativas se encuentra de bruces con la realidad del negocio, que se rige por el valor aportado o el retorno conseguido.

A lo largo de la segunda mitad de 2013 ha crecido considerablemente el número de artículos llamando la atención sobre la relevancia de empezar a trabajar con una aproximación basada en lo que se ha dado en llamar ‘*small data*’.

Desde el punto de vista de la retórica, esa aproximación se ha apropiado de metáforas ya existentes para afirmar ser “la última milla” (*last mile*)

- (17) Investigadora Principal en Microsoft Research, profesora ayudante e investigadora en la Universidad de Nueva York y miembro del Berckman Center en Harvard, <http://www.danah.org/>
- (18) “*Privacy and Publicity in the Context of Big Data*”, borrador disponible en la dirección <http://www.danah.org/papers/talks/2010/WWW2010.html>
- (19) *Gartner’s 2013 Hype Cycle for Emerging Technologies Maps Out Evolving Relationship Between Humans and Machines*. Nota de prensa disponible en la dirección <http://www.gartner.com/newsroom/id/2575515>

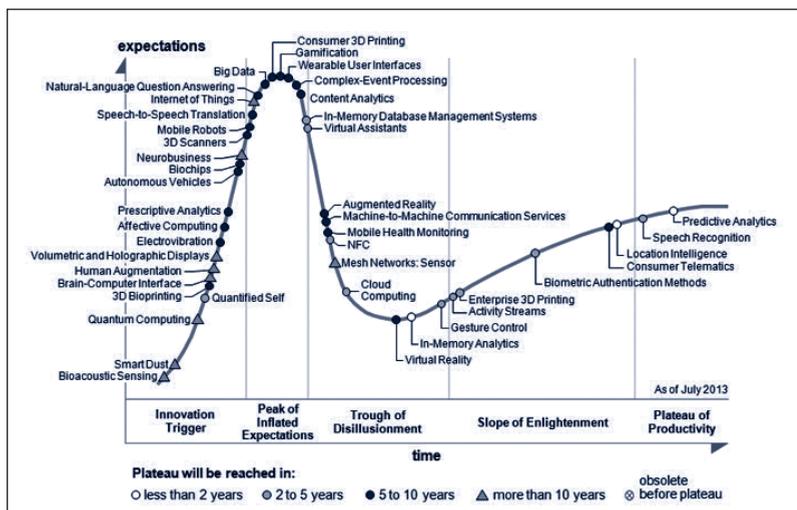


Figura 5. Hype Cycle, agosto 2013 (*Gartner Group*).

del *'Big Data'*, para enfatizar que, más allá de todo el *hype* que rodea ese fenómeno, hay pequeñas iniciativas que son las que consiguen aportar valor.

Desde *Small Data Group* establecen la siguiente definición, que reproducimos en inglés: *"small data connects people with timely, meaningful insights (derived from big data and/or "local" sources), organized and packaged -often visually-, to be accessible, understandable, and actionable for everyday tasks"*.

Estamos hablando de alguna manera de la dimensión "hiperlocal" de un fenómeno característico de lo grande y lo complejo. Se trata, en definitiva, tal como podemos leer en la definición propuesta por este grupo de profesionales, de conectar a las personas con información procesada y "empaquetada" para que resulte accesible, entendible y útil para nuestra vida cotidiana.

El matiz (el foco) se pone aquí en la fuente de los datos que vamos a procesar y organizar para convertir en información útil; y los instrumentos que vamos a utilizar para ello. Hay que entender, además, que esa característica hiperlocalidad incluye la dimensión temporal: Internet es "ahora" y a nadie se le escapa que el adecuado dimensionamiento de los conjuntos de datos a tratar condiciona de una manera determinante nuestra capacidad para tomar decisiones útiles.

El mensaje que se traslada al entorno empresarial –y que es relevante desde mi punto de vista–, es que no necesitamos una gran infraestructura para la gestión de bases de datos masivas, ni una legión de expertos profesionales (*Data Scientists*) para sacar partido a este fenómeno.

Esa determinación es la que, en la práctica, puede convertir en "periodista de datos" a cualquier persona que tenga el interés, el acceso a los conjuntos de datos que necesita y una pericia mínima en el manejo de lo que algunos investigadores hemos convenido en llamar **Tecnologías para la Vida Cotidiana** (TVIC) en los términos que utilizara Fernando Sáez Vacas (2007) para analizar el escenario sociotécnico en el que nos desenvolvemos.

Las iniciativas de *Open Data*, ligadas a la acción social y la transformación de ciertas políticas gubernamentales (*oGov*) han conseguido que la accesibilidad a ciertos conjuntos de datos de interés público dispare los proyectos destinados a mejorar su "tratabilidad" y facilitar su visualización y compartición en la Red.

La accesibilidad a un número y una diversidad crecientes de conjuntos de datos, de todo tipo, así como herramientas para su procesado y visualización, ha hecho que el debate acerca de la privacidad y la seguridad de esos datos en la nube se generalice.

Un debate al que contribuyen, de un lado, la explotación interesada con fines comerciales de datos potencialmente lesivos para ciertos derechos fundamentales por parte de algunas compañías; y, por otro lado los prometedores resultados de ciertas iniciativas que pretenden sustanciar las bondades del fenómeno ganándose la denominación genérica de *'Big Data for Development'*.

Más allá de la necesidad retórica, la consolidación del fenómeno ha provocado la multiplicación de las iniciativas que dentro de las instituciones dedicadas a la cooperación para el desarrollo intentan sacar partido del mismo. El *Big Data* para el Desarrollo (Humano) o *BD4D* (*Big Data for Development*) ha eclosionado en multitud de proyectos.

Un caso ilustrativo es *Global Pulse* (20), un programa de Naciones Unidas que se lanzaba en 2009 con la intención de suplir la carencia de datos precisos y de calidad, en tiempo real, que manifestaban las instituciones responsables de tomar decisiones e instaurar medidas preventivas ante catástrofes socioeconómicas de dimensiones globales o regionales, como puede ser la actual crisis económico-financiera.

Este programa se apoya en la propia Red, no solo como fuente inagotable de datos acerca del comportamiento humano en relación con su contexto, el planeta, sino también como “plataforma” técnica y social para la búsqueda de soluciones innovadoras, ya sea a partir de la red de laboratorios (21) que se van uniendo al programa o mediante el lanzamiento de desafíos (22) en formato ‘*crowdsourcing*’.

En la Agenda Post-2015 de Naciones Unidas se habla de la necesidad de una “revolución de datos” (23) para conseguir cumplimentar los objetivos del milenio. Hay muchos ejemplos de proyectos concretos que dan cuenta de esa revolución:

- plataformas como *Ushahidi* (24), *Micromappers* (25), o *CrisisMappers* (26) han tenido un papel destacado en la organización de la ayuda humanitaria en situaciones de crisis o desastres naturales;
- las conocidas aplicaciones de *Google Flu Trends* (27) o *Google Dengue Trends* (28) han servido para predecir potenciales brotes epidémicos de estas enfermedades;
- el servicio *PriceStats* (29) proporciona estimaciones diarias de la inflación a nivel mundial a partir de la recolección de datos acerca de las ventas minoristas en la web;
- las propias redes sociales, como *Twitter* (30) o *Facebook*, se han convertido en fuentes muy precisas de datos para, por ejemplo, definir políticas de Salud Pública (31) o localizar terremotos con mayor exactitud (32).

(20)

Se puede encontrar un briefing en la siguiente dirección http://www.un.org/millenniumgoals/pdf/GP%20Backgrounder-General2013_Sept2013.pdf

(21)

<http://www.unglobalpulse.org/pulse-labs>

(22)

Un ejemplo reciente es el desafío contra el cambio climático <http://www.unglobalpulse.org/big-data-climate>

(23)

<http://post2015.org/tag/data-revolution/>

(24)

<http://www.ushahidi.com/>

(25)

<http://micromappers.com/>

(26)

<http://crisismappers.net/>

(27)

https://www.google.org/flutrends/intl/es_419/about/how.html

(28)

https://www.google.org/denguetrends/intl/es_419/

(29)

<http://www.pricestats.com/>

(30)

<http://www.technologyreview.com/view/524871/can-twitter-predict-major-events-such-as-mass-protests/>

(31)

http://www.cs.jhu.edu/~mpaul/files/2011.icwsm.twitter_health.pdf

(32)

<http://recovery.doi.gov/press/us-geological-survey-twitter-earthquake-detector-ted/>

Algunas conclusiones

La descripción que he desgranado más arriba, referida específicamente a un fenómeno sociotécnico de amplio espectro, como es el del ‘*Big Data*’ apunta varias implicaciones, más o menos profundas, en distintos ámbitos de los que conocemos como periodismo ciudadano, comunicación ciudadana o acción participativa en la Red.

A partir de la contextualización de ese fenómeno y sobre la base de un modelo sistémico genérico, he querido desarrollar en este artículo una serie de extremos que considero especialmente útiles tanto desde la óptica de la profesión periodística, como desde el punto de vista de cualquier ciudadano preocupado por la manera en que nos informamos en sociedad.

Más allá del proceso crítico de construcción que he seguido y del propio modelo de complejidad que he presentado como herramienta conceptual, es interesante observar al menos los siguientes extremos:

- La evolución experimentada por el fenómeno del *Big Data*, en lo que se refiere a su penetración en los ámbitos empresarial o institucional, va asociada, en gran parte, a un ejercicio renovado de retórica. En el mismo, tiene un papel muy relevante la cuantificación sesgada de una realidad

compleja cuyo análisis requiere una aproximación igualmente basada en la complejidad.

- La contextualización de este fenómeno requiere situarlo en un escenario sociotécnico de mucho más alcance que debe partir de una base sociotecnológica, multi y transdisciplinar por definición.
- La dimensión más relevante del *Big Data* trasciende el evidente volumen de datos que lo sustantiva y de la velocidad creciente con que se generan. Se trata de la “variedad”, entendida como complejidad de los mismos; es decir las interrelaciones que se dan entre esos datos.
- La asociación evidente del proceso periodístico con la “ciencia” de los datos va, también, más allá del desarrollo de una disciplina profesional relacionada con el “periodismo de datos”. Se trata de entender cómo cambia nuestra relación –como ciudadanos o como profesionales–, con las infotecnologías y los contenidos que intercambiamos a través de la Red en procesos informativos con un nuevo salto cualitativo en la complejidad de los datos que tenemos accesibles.

Evidentemente, estas conclusiones están muy lejos de estar cerradas; y lo único que pretenden es servir de base para alimentar un debate crítico –es decir constructivo, desde el escepticismo–, sobre esos procesos de información, relación y comunicación.

REFERENCIAS, BIBLIOGRAFÍA Y WEBGRAFÍA

- **Boyd, D.** (2010). Privacy and Publicity in the Context of Big Data. Consultado el 23/06/2014, de <http://www.danah.org/papers/talks/2010/WWW2010.html>
- **Boyd, D., Crawford, K.** (2011). Six Provocations for Big Data. Consultado el 23/06/2014, de http://softwarestudies.com/cultural_analytics/Six_Provocations_for_Big_Data.pdf
- **Fumero, A.** (2005). Un tutorial sobre *blogs*. El abecé del universo *blog*.TELOS, núm 65, pp. 46-59, oct.-dic. 2005. Consultado el 23/06/2014, de <http://sociedadinformacion.fundacion.telefonica.com/telos/articulocuaderno.asp@idarticulo=1&rev=65.htm>
- **Fumero, A.** (Coord.) (2008). Redes Sociales. TELOS, n. 76, pp. 49-131, jul.-sep. 2008. Consultado el 23/06/2014, de <http://sociedadinformacion.fundacion.telefonica.com/telos/cuaderno.asp@rev=76.htm>
- **Fumero, A.** (Coord.) (2010a). Una Sociedad en Movilidad: Nuevas Fronteras. TELOS, n.83, pp. 43-119 abr.-jun. 2010. Consultado el 23/06/2014, de http://telos.fundaciontelefonica.com/DYC/TELOS/NMEROSANTERIORES/Nmeros8095/DYC/TELOSonline/SOBRETELOS/Nmerosanteriores/Nmero83/seccion=1272&idioma=es_ES.do
- **Fumero, A.** (2010b). La Red en el Móvil. TELOS n. 83, pp. 43-49, abr.-jun. 2010. Consultado el 23/06/2014, de <http://telos.fundaciontelefonica.com/url-direct/pdf-generator?tipoContenido=articuloTelos&idContenido=2010051116020001&idioma=es>
- **Fumero, A.** (2011). IRC 2.0. Medios para la Información, la Relación y la Comunicación en la Web 2.0. El profesional de la información, 2011, nov.-dic., v. 20, n. 6, pp. 605-609. Consultado el 23/06/2014, de <http://www.elprofesionaldelainformacion.com/contenidos/2011/noviembre/01.pdf>
- **Fumero, A.** (2013). La Red en el Móvil. TELOS n. 95, pp. 91-93, abr.-jun. 2013. Consultado el 23/06/2014, de <http://telos.fundaciontelefonica.com/url-direct/pdf-generator?tipoContenido=articuloTelos&idContenido=2013062110120003&idioma=es>
- **Fumero, A.; Espiritusanto, O.** (2011). Jóvenes e infotecnologías: Entre nativ@ y digitales. INJUVE (Instituto de la Juventud), Madrid. ISBN: 978-8496028999. Consultado el 23/06/2014, de <http://infojove.caib.es/joveneseinfotecnologia.pdf>
- **Sáez Vacas, F.** (2007). TVIC: Tecnologías para la Vida Cotidiana. TELOS, n. 73, pp. 4-6, oct.-dic. 2007. Consultado el 23/06/2014, de <http://sociedadinformacion.fundacion.telefonica.com/telos/editorial.asp@rev=73.htm>
- **Sáez Vacas, F.** (2011). Nativos digitales, inteligencia digital. ¿Homo digitalis?. TELOS, n. 86, pp.6-8, ene.-mar. 2011. Consultado el 23/06/2014, de <http://telos.fundaciontelefonica.com/url-direct/pdf-generator?tipoContenido=articuloTelos&idContenido=2011012711540001&idioma=es>