

How journalists can use data to improve the news



The Data Journalism Handbook

Edited by Jonathan Gray, Liliana Bounegru
and Lucy Chambers

¿Qué es el periodismo de datos?

¿Qué es el periodismo de datos? Podría contestar, simplemente, que es periodismo que se hace con datos. Pero eso no es de gran ayuda.

Tanto “datos” como “periodismo” son términos problemáticos. Algunos creen que “datos” es cualquier colección de cifras, por lo general reunidas en una hoja de cálculo. Hace 20 años, esos eran prácticamente los únicos datos que manejaban los periodistas. Pero ahora vivimos en un mundo digital, un mundo en el que casi cualquier hecho puede ser (y casi todo es) descripto con números.

Su carrera profesional, 300.000 documentos confidenciales, las personas que componen su círculo de amigos; todo esto puede ser (y es) descripto con solo dos números: ceros y unos. Fotos, video, y audio; asesinatos, enfermedades, votos políticos, corrupción y mentiras, también descriptos con ceros y unos.

¿Qué es lo que hace que el periodismo de datos sea diferente del resto del periodismo? Quizás sean las nuevas posibilidades que aparecen, cuando se combina el tradicional “olfato para las noticias” y la capacidad de narrar una historia convincente, con la escala y alcance de la información digital disponible en la actualidad.

Y esas posibilidades pueden aparecer en cualquier momento del proceso periodístico: cuando contamos con la programación necesaria para automatizar el proceso de recoger y combinar información proveniente del gobierno municipal, la policía y otras fuentes civiles, como hizo Adrian Holovaty con [ChicagoCrime](#) y luego [EveryBlock](#).

O usar software para encontrar relaciones entre cientos y miles de documentos, tal como hizo The Telegraph con [los gastos de los parlamentarios](#).

Investigate your MP's expenses

Join us in digging through the documents of MPs' expenses to identify individual claims, or documents that you think merit further investigation. You can work through your own MP's expenses, or just hit the button below to start reviewing. (Update, Fri pm: we now have a virtually complete set of expenses documents so you should be able to find your MP's) Already created an account? [Log in here](#).

We have **458,832** pages of documents. **32,755** of you have reviewed **225,443** of them. Only **233,389** to go...

[Start reviewing](#)

Please read our [privacy policy](#) to find out how we use your data. You must also read our [terms of service](#). By reviewing pages, you are agreeing that you have read the terms of service, and that you agree to them.

Figure 3. Investigue los gastos de su representante parlamentario (The Guardian)

El periodismo de datos puede ayudar a un periodista a contar una historia convincente por medio de infografías atractivas. Por ejemplo, las conversaciones espectaculares de Hans Roslign sobre la visualización de la pobreza mundial con [Gapminder](#) (que se puede traducir como Recuerdabrecha, n. del t.) han atraído millones de visitas en todo el mundo. Y la obra popular de David McCandless al destilar grandes cifras –tales como poner en contexto el gasto público, o la polución generada por el volcán islandés- muestra la importancia de un diseño claro en [Information is Beautiful](#).

O puede ayudar a explicar cómo se relaciona una historia con un individuo, como hacen ahora la BBC y el Financial Times habitualmente con sus interactivos sobre el presupuesto (donde usted puede averiguar cómo el presupuesto lo afecta en particular a usted en vez de a un genérico “Juan Pueblo”). Y puede abrir el proceso mismo de búsqueda de información, como hace The Guardian de modo tan exitoso al compartir datos, contexto y preguntas en su [Datablog](#).

Los datos pueden ser la fuente del periodismo de datos, o pueden ser la herramienta con la que se narra la historia o ambas cosas. Como cualquier fuente, debe tratarse con escepticismo; y como cualquier herramienta, debemos ser conscientes de cómo puede modelar y limitar las historias que se crean con la misma.

— *Paul Bradshaw, Birmingham City University*

Por qué debieran usar datos los periodistas

El periodismo está sitiado. En el pasado, como sector, nos basábamos en ser los únicos que operábamos una tecnología para multiplicar y distribuir lo que había pasado de un día al otro. La imprenta servía como puerta de entrada. Cualquiera que quisiera llegar a la gente de una ciudad o una región a la mañana siguiente, recurría a los diarios. Esa era se acabó.

Hoy las noticias fluyen al mismo tiempo que suceden, a través de múltiples fuentes, testigos presenciales y blogs, y lo que ha sucedido es filtrado a través de una vasta red de conexiones sociales, se jerarquiza, se comenta y muy a menudo se ignora.

Por eso el periodismo de datos es tan importante. Reunir, filtrar y visualizar lo que sucede más allá de lo que nos muestran nuestros ojos tiene creciente valor. En la economía global de hoy el jugo de naranja que toma por la mañana, el café que prepara... hay relaciones invisibles entre estos productos, otra gente y usted. El lenguaje de esta red es el de los datos: pequeños puntos de información que a menudo son irrelevantes como instancia individual, pero enormemente importantes cuando se los ve desde el ángulo correcto.

En este momento, unos cuantos periodistas pioneros ya están demostrando cómo se puede usar datos para crear una visión más profunda de lo que sucede a nuestro alrededor y cómo puede afectarnos.

El análisis de datos puede revelar “la forma de una historia” (Sarah Cohen) o proveernos una “nueva cámara” (David McCandless). Usando datos, la tarea de los periodistas pasa de centrarse en ser los primeros en informar, a ser los que nos dicen lo que un proceso podría significar realmente. La gama de temas puede ser amplia. La próxima crisis financiera en ciernes. Los datos económicos detrás de los productos que usamos. El mal uso de fondos o errores políticos, presentados con una visualización convincente que deje poco margen para rebatirla.

Es por esto que los periodistas debieran ver los datos como una oportunidad. Es posible, por ejemplo, revelar cómo una amenaza abstracta (como el desempleo) afecta a la gente de acuerdo a su edad, su género o su nivel de educación. Usar datos transforma algo abstracto en algo que todos pueden entender y con lo que pueden relacionarse.

Pueden crear herramientas de cálculo personalizadas para ayudar a la gente a tomar decisiones, se trate de comprar un auto o una casa, decidir un rumbo educativo o profesional en su vida, o hacer un control de costos para no meterse en deudas.

Pueden analizar la dinámica de una situación compleja como disturbios o un debate político, mostrar falacias y ayudar a todos a encontrar posibles soluciones para problemas complejos.

Formarse en la búsqueda, depuración y visualización de datos es transformador para la profesión de reunir información también. Los periodistas que dominen esto descubrirán que apoyar sus artículos en datos y la visión que aportan es un alivio. Menos adivinar, menos buscar citas; en vez de ello, un periodista puede crear una posición fuerte apoyada en datos y esto puede afectar mucho el rol del periodismo.

Además, introducirse en el periodismo de datos ofrece una perspectiva para el futuro. Hoy, cuando las redacciones se reducen, la mayoría de los periodistas esperan cambiar el área de las relaciones públicas. Pero los periodistas de datos o los científicos de datos ya son un grupo de profesionales muy solicitados, no solo por los medios. Las empresas e instituciones de todo el mundo buscan “gente que encuentre sentido a las cosas”, y profesionales que sepan cómo revisar datos y convertirlos en algo tangible.

Los datos representan una promesa, y esto es lo que entusiasma a las redacciones, haciéndolas buscar un nuevo tipo de periodista. Para la gente que trabaja por su cuenta, manejar datos ofrece un camino para obtener nuevas oportunidades y un salario estable también. Véalo de este modo: en vez de contratar periodistas que llenen rápidamente páginas y sitios en la red con contenido de bajo valor, el uso de datos podría crear demanda para paquetes interactivos, que solo pueden crearse invirtiendo una semana entera en resolver una cuestión. Esto es un cambio positivo para muchos sectores de los medios.

Hay una barrera que impide a los periodistas usar este potencial: la necesidad de capacitarse para trabajar con datos en todos los pasos, desde una primera pregunta hasta un gran impacto periodístico basado en datos.

Trabajar con datos es como introducirse en un territorio vasto y desconocido. A primera vista los datos crudos resultan inteligibles para los ojos y la mente. Tales

datos son inmanejables. Es difícil ordenarlos correctamente para su visualización. Se necesita periodistas experimentados, que tengan la energía como para analizar datos crudos a menudo confusos o aburridos y “ver” las historias ocultas allí.

— *Mirko Lorenz, Deutsche Welle*

El estudio

El European Journalism Centre realizó una **encuesta** para saber más sobre las necesidades de capacitación de los periodistas. Descubrimos que hay una gran disposición de salir de la postura cómoda del periodismo tradicional, e invertir tiempo para dominar nuevas capacidades. Los resultados de la encuesta demuestran que los periodistas ven la oportunidad, pero necesitan un poco de apoyo para superar los problemas iniciales que les impiden trabajar con datos. Hay confianza de que si el periodismo de datos fuera adoptado de modo más universal, los flujos de trabajo, las herramientas y los resultados mejorarían rápidamente. Pioneros tales como The Guardian, The New York Times, The Texas Tribune, y Die Zeit siguen elevando el nivel con sus artículos basados en datos.

¿El periodismo de datos seguirá siendo el dominio de un pequeño puñado de pioneros o pronto toda organización de noticias tendrá su propio equipo de periodistas dedicados especialmente a los datos. Esperamos que este manual ayude a más periodistas y redacciones a aprovechar este campo emergente.

¿Por qué es importante el periodismo de datos?

Preguntamos a algunos de los principales practicantes y partidarios del periodismo de datos por qué piensan que el periodismo de datos es un desarrollo importante. Los datos crudos resultan inteligibles para los ojos y la mente. Esto es lo que dijeron.

Filtrar el flujo de datos.

Cuando había escasez de información, la mayor parte de nuestros esfuerzos estaban dedicados a buscarla y reunirla. Ahora que la información es abundante, es más importante el procesamiento. El procesamiento tiene dos niveles: 1) análisis para encontrar sentido y estructura en el flujo sin fin de datos y 2) presentación de

esa información para meter lo que es importante y relevante en la cabeza del consumidor. Al igual que la ciencia, el periodismo de datos da a conocer sus métodos y presenta sus descubrimientos de un modo que pueda ser verificado a través de su replicado.

— *Philip Meyer, Professor Emeritus, University of North Carolina at Chapel Hill*

Nuevos enfoques para narrar historias

El periodismo de datos es un término abarcativo que, para mí, incluye un conjunto de herramientas, técnicas y enfoques de la narrativa siempre crecientes. Puede incluir todo, desde el tradicional periodismo asistido por computadoras (usando datos como una “fuente”) hasta la visualización más avanzada de datos y aplicaciones de noticias. El objetivo unificador es periodístico: proveer información y análisis para ayudar a informarnos todos sobre asuntos importantes de actualidad.

— *Aron Pilhofer, New York Times*

Como periodismo fotográfico con una laptop

“El periodismo de datos” difiere del “periodismo escrito” solo en que usamos un equipo diferente. Todos nos ganamos la vida olfateando, reportando y relacionando historias. Es como el “periodismo fotográfico”; solo hay que cambiar la cámara por una laptop.

— *Brian Boyer, Chicago Tribune*

El periodismo de datos es el futuro

El periodismo de datos es el futuro. Los periodistas tienen que saber manejar datos. Hace un tiempo uno descubría cosas hablando con gente en bares, y puede ser que esto siga sucediendo a veces. Pero ahora también se trata de analizar datos, equiparse con herramientas, y analizarla y encontrar lo que es interesante. Tener todo en perspectiva, ayudando a la gente a ver cómo encajan las piezas (para no repetir todo), y qué pasa en el país.

— *Tim Berners-Lee, founder of the World Wide Web*

El procesamiento de cifras se une al pulido del lenguaje

El periodismo de datos es tender un puente para superar la brecha entre los técnicos estadísticos y los cinceladores de palabras. Ubicar cosas destacadas e identificar tendencias que no solo son significativas estadísticamente sino que también son relevantes para desentrañar el mundo de hoy, que es intrínsecamente complejo.

— *David Anderton, freelance journalist*

Actualizar sus capacidades

El periodismo de datos implica un nuevo conjunto de habilidades para buscar, comprender y visualizar fuentes digitales, en una época en que las capacidades básicas del periodismo tradicional ya no bastan. No lo reemplaza, le agrega cosas.

En un momento en que las fuentes se están volviendo digitales, los periodistas pueden y tienen que estar más en contacto con estas fuentes. Internet abrió posibilidades que van más allá de lo que podemos entender hoy. El periodismo de datos es solo el comienzo de la evolución de nuestras prácticas pasadas para adaptarse al online.

El periodismo de datos sirve a dos importantes propósitos para las organizaciones de noticias: encontrar historias únicas (no de los cables) y ejecutar la función de alerta. Especialmente en tiempos de crisis financieras, estos objetivos son importantes para los diarios.

Desde el punto de vista de un diario regional, el periodismo de datos es crucial. Existe el dicho: “una teja floja en su casa se considera más importante que disturbios en un país lejano”. A uno lo golpea en la cara e impacta en su vida de modo más directo. Al mismo tiempo, la digitalización está en todas partes. Debido a que los diarios locales tienen este impacto directo en su vecindario y las fuentes se vuelven digitalizadas, un periodista debe saber cómo encontrar, analizar y visualizar una historia a partir de datos.

— *Jerry Vermanen, NU.nl*

Un remedio para la asimetría de la información

La asimetría de la información –no la falta de información sino la incapacidad de absorberla y procesarla a la velocidad y con el volumen que nos llega- es uno de los

problemas más significativos que enfrentan los ciudadanos al elegir cómo vivir sus vidas. La información tomada de medios impresos, visuales y radiales influye en las opciones y las acciones de los ciudadanos. El buen periodismo de datos ayuda a combatir la asimetría de la información.

— *Tom Fries, Bertelsmann Foundation*

Una respuesta a las relaciones públicas basadas en datos

La disponibilidad de herramientas de medición y sus precios decrecientes –en una combinación auto-sustentada que se concentra en el desempeño y la eficiencia en todos los aspectos de la sociedad- han llevado a quienes toman las decisiones a cuantificar los avances de sus políticas, monitorear tendencias e identificar oportunidades.

Las compañías continuamente encuentran nuevas mediciones que muestran su buen desempeño. A los políticos les encanta alardear de las cifras sobre reducción de desempleo y crecimiento del PBI. La falta de conocimientos por parte de los periodistas respecto de los escándalos de Enron, Worldcom, Madoff o Solyndra es prueba de la incapacidad de muchos profesionales de ver más allá de las cifras. Hay una tendencia a aceptar las cifras más que otros datos, ya que tienen un aura de seriedad, aunque sean completamente inventadas.

El saber manejar datos ayudará a los periodistas a aguzar su sentido crítico al enfrentar cifras, y ojalá que les sirva para avanzar un poco en su relación con los departamentos de RRPP.

— *Nicolas Kayser-Bril, Journalism++*

Proveer interpretaciones independientes de información oficial

Luego del terremoto devastador y el subsecuente desastre de la planta nuclear de Fukushima en 2011, la importancia del periodismo de datos se ha hecho claro para la gente de medios en Japón, país que en general va a la zaga en materia de periodismo digital.

Quedamos a la deriva cuando el gobierno y los expertos no tuvieron datos creíbles acerca de los daños. Cuando los funcionarios ocultaron al público los datos SPEEDI (predicción de difusión de materiales radioactivos), no estábamos en condiciones de decodificarlos aunque se hubiesen filtrado. Voluntarios comenzaron a reunir

datos sobre radioactividad usando sus propios recursos, pero no estábamos armados con conocimientos estadísticos, de interpolación, de visualización y demás. Los periodistas tienen que tener acceso a los datos en crudo y aprender a no depender de las interpretaciones oficiales de los mismos.

— *Isao Matsunami, Tokyo Shimbun*

Manejar el diluvio de datos

Los desafíos y las oportunidades que presenta la revolución digital siguen complicando al periodismo. En una era de abundancia de información, los periodistas y los ciudadanos necesitan mejores herramientas, se trate de curar los samizdat del siglo XXI en Medio Oriente, procesar una avalancha de datos difundidos a medianoche, o encontrar la mejor manera de visualizar la calidad del agua en una nación. Al debatirnos con los desafíos del consumo que presenta este diluvio de datos, las nuevas plataformas de edición también están dando a todos el poder de reunir y compartir datos digitalmente, convirtiéndolos en información. Mientras los periodistas y editores han sido los vectores tradicionales de la colecta y diseminación de información, el ambiente horizontal de información ahora hace que las noticias se conozcan primero online y no en las redacciones.

En todo el planeta, de hecho, el vínculo entre los datos y el periodismo se está fortaleciendo. En una era de grandes cantidades de datos, la creciente importancia del periodismo de datos está en la capacidad de sus practicantes de dar contexto, claridad y –quizás lo más importante, encontrar la verdad en la cantidad en expansión de contenido digital en el mundo. Eso no significa que las organizaciones de medios integradas de hoy no tengan un rol crucial. Lejos de ello. En la era de la información, se necesita más que nunca a los periodistas para curar, verificar, analizar y sintetizar los datos. En ese contexto, el periodismo de datos tiene una profunda importancia para la sociedad.

Hoy, encontrarle sentido a los grandes volúmenes de datos, en particular los datos no estructurados, serán un objetivo central de los científicos de todo el mundo, trabajen en salas de redacción, Wall Street o Silicon Valley. Notoriamente esa meta se verá facilitada sustancialmente por un conjunto creciente de herramientas comunes, sean empleadas por tecnólogos del estado, tecnólogos de la salud o desarrolladores de las redacciones.

— *Alex Howard, O'Reilly Media*

Nuestras vidas son datos

El buen periodismo de datos es difícil, porque el buen periodismo es difícil. Significa cómo obtener los datos, cómo entenderlos, y cómo encontrar la historia. A veces hay callejones sin salida, y a veces no hay una gran historia. Al fin de cuentas, si solo fuera cuestión de apretar el botón indicado, no sería periodismo. Pero eso es lo que hace que valga la pena —en un mundo en el que nuestras vidas cada vez son más datos—, que sea esencial para una sociedad libre y justa.

— *Chris Taggart, OpenCorporates*

Una manera de ahorrar tiempo

Los periodistas no tienen tiempo para perder transcribiendo cosas a mano y complicarse tratando de obtener información de archivos PDF, por lo que aprender un poco de código (o saber dónde buscar gente que puede ayudar) es increíblemente valioso.

Un periodista de Folha do São Paulo estaba trabajando con el presupuesto local y me llamó para agradecernos por publicar online las cuentas de la municipalidad de São Paulo (2 días de trabajo para un solo hacker). Dijo que las había estado transcribiendo a mano los últimos 3 meses, tratando de encontrar una historia. También recuerdo haber resuelto un “problema de PDF” para *Contas Abertas*, una organización que monitorea noticias parlamentarias: 15 minutos y 15 líneas de código, en vez de un mes de trabajo.

— *Pedro Markun, Transparência Hacker*

Una parte esencial del herramental del periodista

Creo que es importante destacar el aspecto “periodístico” o de reportero del “periodismo de datos. El ejercicio no debiera ser analizar o visualizar datos por el gusto de hacerlo, sino utilizarlo como herramienta de modo de aproximarnos más a la verdad de lo que sucede en el mundo. Veo la capacidad de analizar e interpretar datos como parte esencial del set de herramientas actual de los periodistas, en vez de una disciplina por separado. Al fin de cuentas, todo tiene que ver con el buen periodismo y contar historias del modo más apropiado.

El periodismo de datos es otra manera de analizar el mundo y hacer que los poderes constituidos rindan cuentas. Con una creciente cantidad de datos disponible, ahora es más importante que nunca que los periodistas sean conscientes de las técnicas del periodismo de datos. Esta debe ser una herramienta que cualquier periodista debiera incorporar, se trate de aprender cómo trabajar directamente con datos, o a colaborar con alguien que lo pueda hacer.

Su verdadero potencial está en ayudarlo a obtener información que de otro modo sería muy difícil de encontrar o demostrar. Un buen ejemplo es la historia de Steve Doig que analizó patrones de daños del huracán Andrew. Unió dos conjuntos distintos de datos: uno que mapeaba el nivel de destrucción causado por el huracán, y otro que muestra las velocidades de los vientos. Esto le permitió señalar áreas en las cuales las malas prácticas en la construcción de edificios contribuyeron/intensificaron el impacto del desastre. Ganó por la historia un [Pulitzer Prize](#) en 1993 y sigue siendo un gran ejemplo de lo que es posible.

Idealmente se usan los datos para descubrir cosas destacadas, sorprendentes o áreas de interés. En este sentido, actúan como pistas. Si bien las cifras pueden ser interesantes, no basta escribir solamente sobre datos. Hay que hacer el trabajo de periodista para explicar qué significan.

— *Cynthia O'Murchu, Financial Times*

Adaptarse a cambios en nuestro ambiente de información

Las nuevas tecnologías digitales generan nuevas maneras de producir y diseminar el conocimiento en la sociedad. El periodismo de datos puede entenderse como el intento de los medios de adaptarse y responder a los cambios en el ambiente de la información, incluyendo maneras de contar historias más interactivas y multidimensionales, que permite a los lectores explorar las fuentes que subyacen a las noticias, alentándolos a participar en el proceso de crear y evaluar historias.

— *César Viana, University of Goiás*

Una manera de ver cosas que de otro modo podría no ver

Algunas historias sólo pueden entenderse y explicarse analizando –y a veces visualizando- datos. Las relaciones entre personas o entes poderosos quedarían sin revelar, las muertes causadas por políticas farmacéuticas permanecerían ocultas,

las políticas ambientales que dañan el medio continuarían sin límite. Pero cada una de estas situaciones han podido modificarse gracias a los datos obtenidos, analizados y aportados por los periodistas a los lectores. Los datos pueden ser simples como una planilla de cálculo, o un registro de llamadas telefónicas, o complejos como los resultados de pruebas escolares o datos de infecciones hospitalarias; como sea, allí hay historias que vale la pena contar.

— Cheryl Phillips, *The Seattle Times*

Una manera de enriquecer los artículos

Podemos pintar cuadros de nuestras vidas completas con nuestro rastro digital. Desde lo que consumimos y navegamos, hasta donde y cuando viajamos, nuestras preferencias musicales, nuestros primeros amores, los hitos de nuestros hijos, incluso nuestros últimos deseos, todo puede ser rastreado, digitalizado, almacenado en la nube y difundido.**Este universo de datos puede ser sacado a la superficie para narrar historias, responder preguntas e impartir una comprensión de la vida de maneras que actualmente superan incluso la más rigurosa y cuidadosa reconstrucción de anécdotas.

— Sarah Slobin, *Wall Street Journal*

No se necesitan nuevos datos para tener una primicia

A veces los datos ya son públicos y están disponibles, pero nadie los ha analizado atentamente. En el caso del informe de Associated Press sobre 4500 páginas de documentos desclasificados que describen las acciones de contratistas de seguridad privados durante la guerra de Irak, el material fue obtenido por un periodista independiente a lo largo de varios años, usando pedidos de Acceso a la Información dirigidos al departamento de Estado de EE.UU. Escanearon los resultados impresos y los subieron a DocumentCloud, lo que nos permitió hacer nuestro análisis general.

— Jonathan Stray, *The Overview Project*

Algunos ejemplos favoritos

Le preguntamos a algunos de nuestros colaboradores acerca de sus ejemplos favoritos de periodismo de datos y qué les gusta de los mismos. Sus respuestas, a continuación:

No causar daño, en el Las Vegas Sun

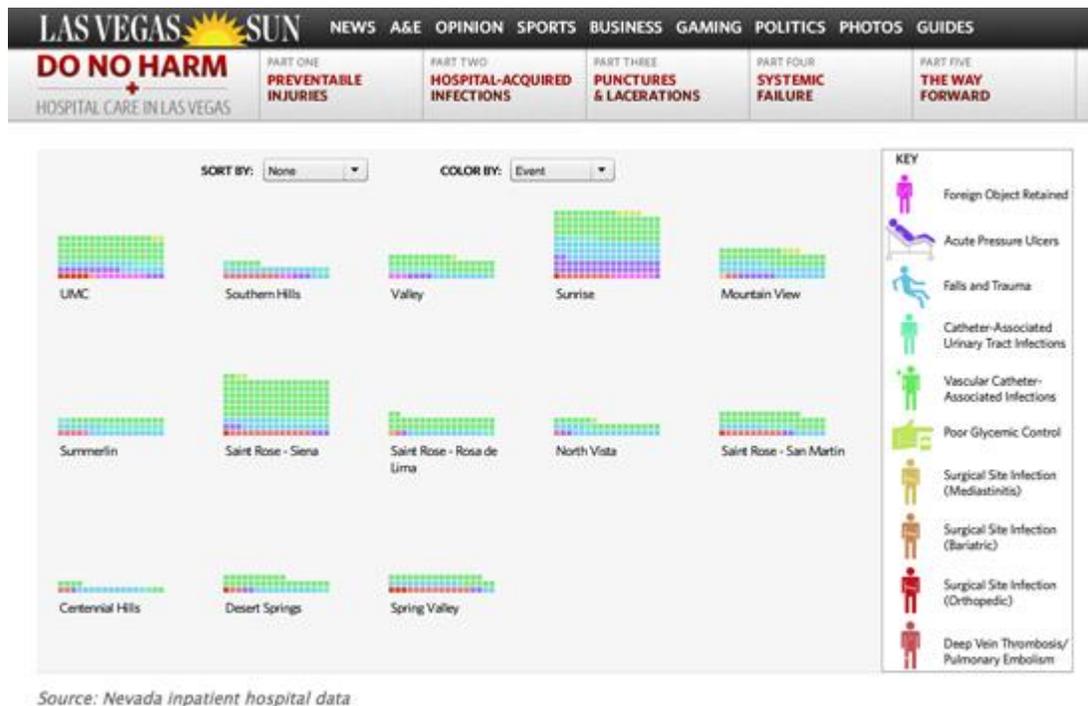


Figure 5. No causar daño (The Las Vegas Sun)

Mi ejemplo favorito es la serie [No causar daño](#) de 2010 en Las Vegas Sun, sobre la atención en los hospitales. El Sun analizó más de 2.900.000 de registros de aranceles hospitalarios, que revelaron más de 3600 lesiones, infecciones y errores quirúrgicos evitables. Obtuvieron datos a través de un pedido de acceso a archivos públicos e identificaron más de 300 casos en que los pacientes murieron por errores que pudieron haberse prevenido. Contiene distintos elementos, incluyendo un [gráfico interactivo](#) que permite al lector ver (por hospital) donde se dieron lesiones quirúrgicas más a menudo de lo esperado; un [mapa](#) con un cronograma que muestra cómo se extienden las infecciones hospital por hospital; y un [gráfico interactivo](#) que permite a los usuarios ordenar los datos por lesiones evitables o por

hospital, para ver dónde la gente se ve afectada. Me gusta porque es muy fácil de entender y navegar. Los usuarios pueden explorar los datos de manera muy intuitiva.

Además tuvo un impacto real: la legislatura de Nevada respondió con [6 legislaciones](#). Los periodistas involucrados trabajaron muy duro para obtener y desmenuzar los datos. Uno de los periodistas, Alex Richards, envió los datos a los hospitales y al Estado al menos una docena de veces para lograr que se corrigieran los errores.

— *Angélica Peralta Ramos, La Nación (Argentina)*

Base de datos de salarios de empleados del Estado



Figure 6. Salarios de Empleados del Estado (The Texas Tribune)

Me encanta el trabajo que pequeñas organizaciones independientes realizan todos los días tales como ProPublica o el Texas Tribune, que tiene a Ryan Murphy como gran periodista de datos. Si tuviera que elegir, optaría por el proyecto de base de datos de [Salarios de Empleados del Estado del Texas Tribune](#). Este proyecto reúne en una base de datos la información de los salarios de 660.000 empleados estatales

para que los usuarios busquen y ayuden a generar historias. Se puede buscar por ente estatal, nombre o salario. Es simple, significativo y pone a disposición del público información hasta ahora inaccesible. Es fácil de usar y genera historias de manera automática. Es un gran ejemplo que muestra por qué el Texas Tribune concentra la mayor parte de su tráfico en sus páginas de datos.

— *Simon Rogers, The Guardian*

Visualización de texto completo de los registros de la guerra de Irak, Associated Press



WikiLeaks Iraq SIGACTS (redacted) - Dec 2006

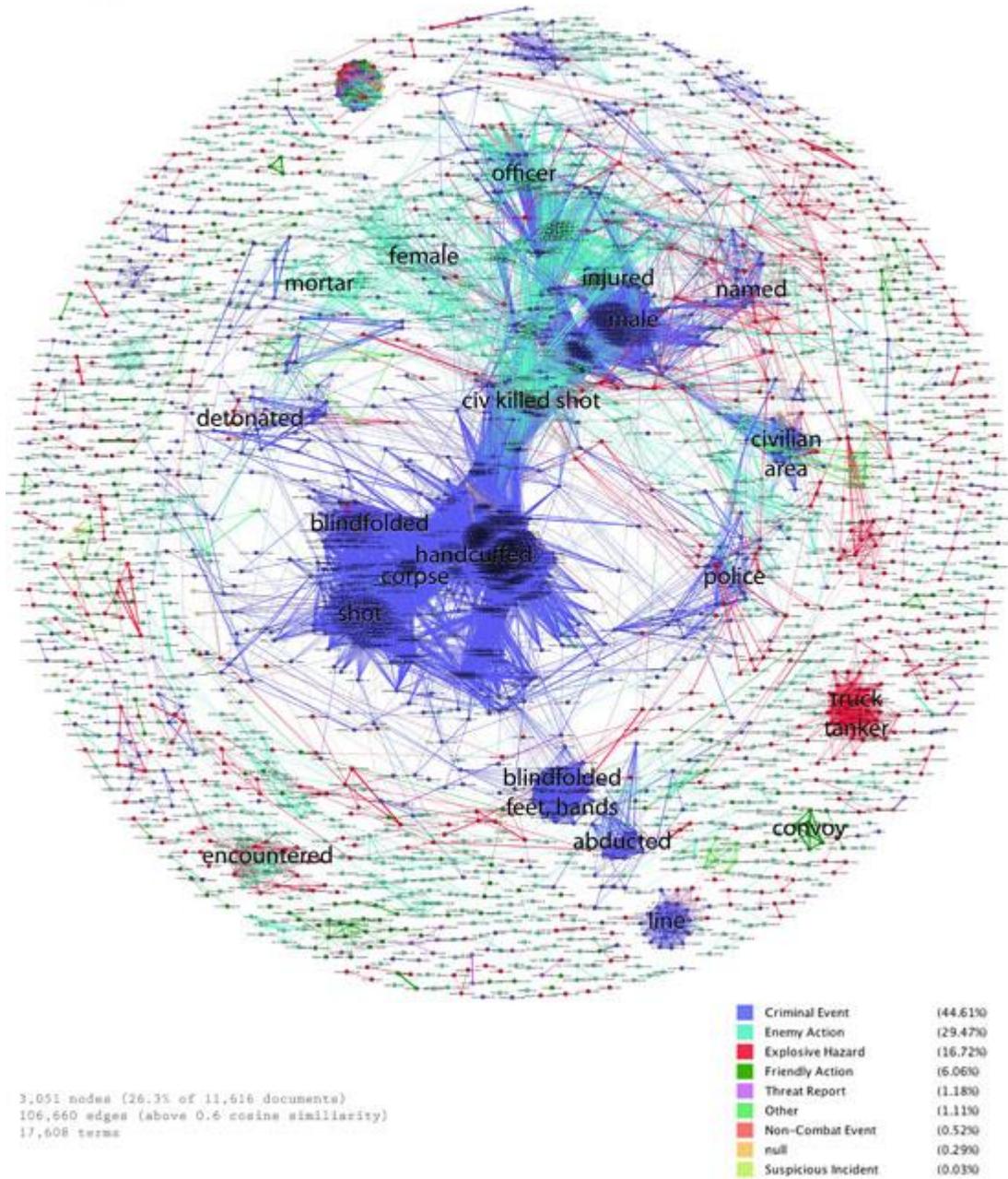


Figure 7. Análisis de los registros de guerra (Associated Press)

El trabajo de Jonathan Stray y Julian Burgess sobre los [registros \(logs\) de la Guerra de Irak](#) es una llamativa incursión en el análisis de texto y la visualización, utilizando técnicas experimentales para comprender temas que vale la pena explorar, dentro de un gran conjunto de datos en formato texto.

Por medio de técnicas y algoritmos de analítica de textos, Jonathan y Julian crearon un método que muestra concentraciones de palabras clave contenidas en miles de informes del gobierno de Estados Unidos sobre la guerra de Irak, difundido por WikiLeaks, en un formato visual.

Si bien este método tiene limitaciones y el trabajo es experimental, es un enfoque nuevo e innovador. En vez de tratar de leer todos los archivos o revisar los registros de guerra con una noción preconcebida de lo que puede encontrarse ingresando palabras claves y revisando el resultado, esta técnica calcula y visualiza temas/palabras clave de particular relevancia.

Con crecientes cantidades de datos en formato texto (emails, informes, etc.) y numérico llegando al dominio público, encontrar maneras de determinar áreas de interés clave se volverá cada vez más importante. Es un sub-campo interesante del periodismo de datos.

— *Cynthia O'Murchu, Financial Times*

Misterios de Asesinatos

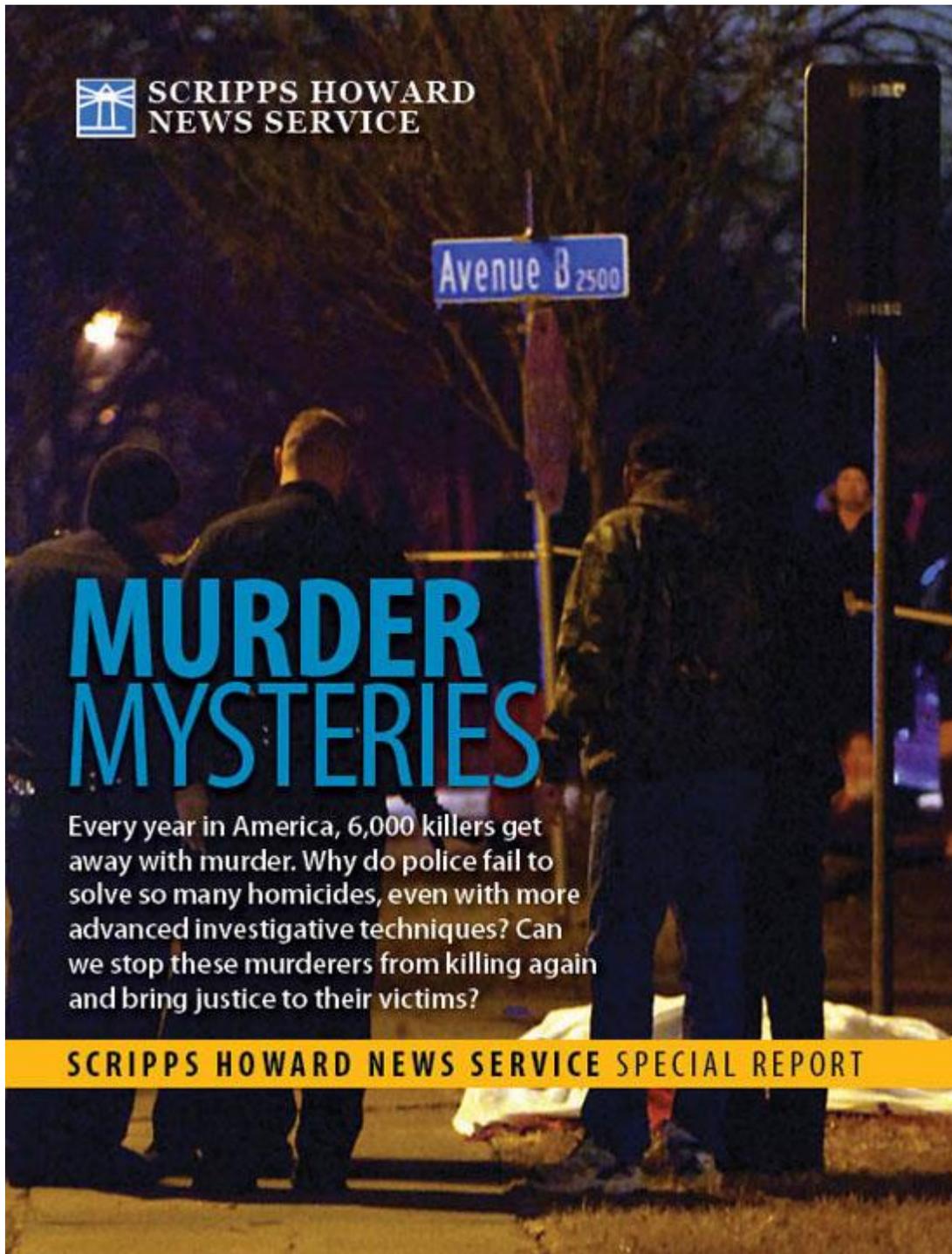


Figure 8. Misterios de asesinatos (Scripps Howard News Service)

Una de mis piezas favoritas de periodismo de datos es el proyecto de [Misterios de Asesinatos](#), por Tom Hargrove del Scripps Howard News Service. A partir de datos

oficiales y pedidos de acceso a registros públicos, creó una base de datos, que incluye el detalle demográfico de más de 185.000 asesinatos no resueltos, y luego diseñó un algoritmo para buscar patrones que sugieran la posible presencia de asesinos seriales.

Este proyecto tiene todo: un gran trabajo, una base de datos mejor que la del estado, análisis inteligente usando técnicas de ciencias sociales, y una presentación interactiva de datos online de modo que los lectores puedan explorar por su cuenta.

— Steve Doig, Walter Cronkite School of Journalism, Arizona State University

Máquina de Mensajes

Figure 9. Máquina de Mensajes (ProPublica)

Me encanta la historia de [Máquina de Mensajes](#) de ProPublica y su [blog nerd](#). Todo comenzó cuando un grupo de tuiteros expresó curiosidad por haber recibido correos electrónicos diferentes de la campaña de Barack Obama. La gente de ProPublica tomó nota y pidió a su público que reenviaran los correos que recibieran de la campaña. La presentación es elegante, un análisis diferencial visual de varios correos diferentes que fueron enviados esa noche. Es admirable porque recogieron sus propios datos (una pequeña muestra, pero lo suficiente como para contar la

historia). Pero es aún más admirable porque cuenta la historia de un fenómeno en curso: gran cantidad de datos utilizados en campañas políticas para dirigir mensajes a individuos específicos. Es sólo un anticipo de cosas por venir.

— Brian Boyer, *Chicago Tribune*

Chartball

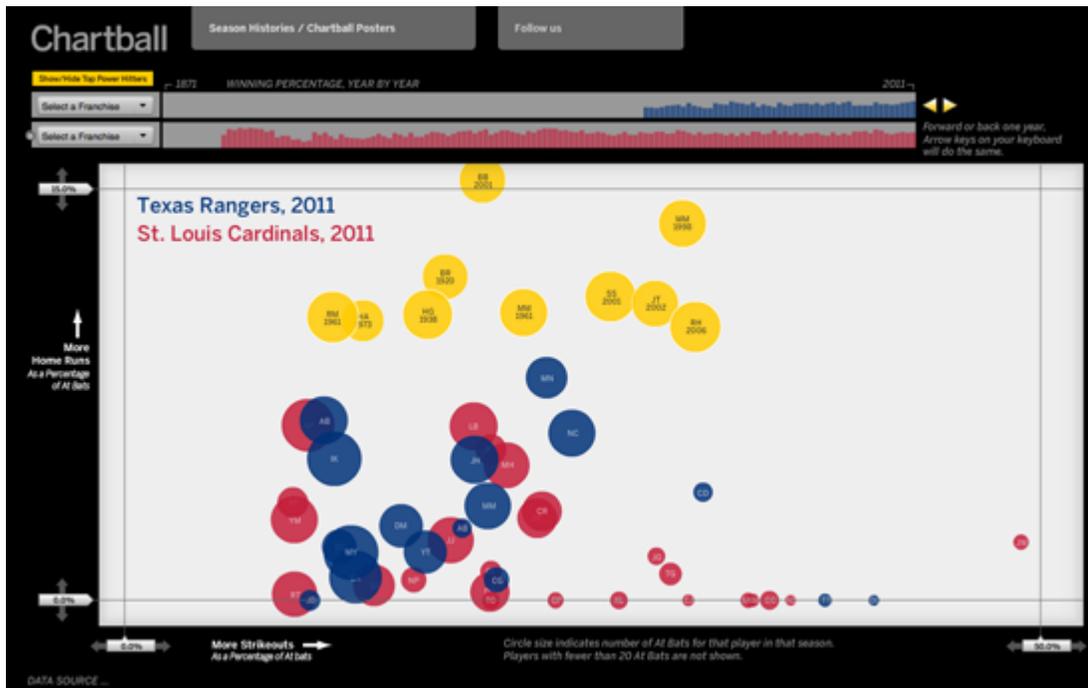


Figure 10. Gráfico de victorias y derrotas (Chartball)

Uno de mis proyectos de periodismo de datos favoritos es el trabajo de Andrew García Phillips sobre <http://www.chartball.com/Chartball>. Andrew es fanático de los deportes con un voraz apetito de datos, un ojo tremendo para el diseño y la capacidad de escribir código. En Chartball no solo visualiza el conjunto de la historia, sino que detalla los éxitos y fracasos de jugadores individuales y equipos. Ofrece contexto, un gráfico atractivo y su trabajo es profundo, divertido e interesante, y a mí ni siquiera me interesan demasiado los deportes.

— Sarah Slobin, *Wall Street Journal*

El periodismo de datos en perspectiva

En agosto de 2010 algunos colegas del European Journalism Centre y yo organizamos lo que creemos que fue una de las primeras **conferencias internacionales de periodismo de datos**, que se realizó en Ámsterdam, Holanda. En aquel momento no había mucha discusión respecto del tema, y solo había un par de organizaciones conocidas ampliamente por su labor en esta área.

La manera en que organizaciones de noticias como The Guardian y el New York Times manejaron las grandes cantidades de datos difundidos por WikiLeaks, es uno de los grandes casos que impulsaron el término. En aquel momento el concepto comenzó a tener un uso más amplio (junto con “el periodismo asistido por computadora”) para describir cómo los periodistas utilizaban datos para mejorar su cobertura y amplificar investigaciones profundas de un tema dado.

Hablando con periodistas de datos y estudiosos del periodismo **on Twitter**, parecería que una de las formulaciones más tempranas de lo que ahora reconocemos como periodismo de datos, en 2006 por Adrian Holovaty, fundador de EveryBlock, un servicio de información que permite a los usuarios saber lo que ha estado sucediendo en su área, en su manzana. En su breve ensayo “**Un modo fundamental en que los sitios de diarios tienen que cambiar**”, sostiene que los periodistas deben publicar datos estructurados procesables por la computadora, junto con el “gran bodeque de texto” tradicional:

Por ejemplo, digamos que un diario ha escrito una historia sobre un incendio local. Poder leer la historia en un celular está bien. Viva la tecnología. Pero lo que realmente quiero poder hacer es explorar los datos en crudo de esa historia, uno por uno, con capas de atribuciones, y una infraestructura para comparar detalles del incendio con incendios anteriores: fecha, momento, lugar, víctimas, número de la central de bomberos, distancia de la central de bomberos, nombres y años de experiencia de los bomberos que actuaron, el tiempo que les llevó a los bomberos llegar, e incendios posteriores, cuando sea que sucedan.

¿Pero qué es lo que distingue esto de otras formas de periodismo que usan bases de datos o computadoras? ¿Cómo y en qué medida el periodismo de datos es distinto de otras formas de periodismo del pasado?

Periodismo Asistido por Computadora y Periodismo de Precisión

Usar datos para mejorar los informes y presentar información estructurada (aunque no sea legible por la computadora) al público tiene una larga historia. Quizás lo más relevante en forma inmediata para lo que ahora llamamos periodismo de datos es el periodismo asistido por computadora, conocido por la sigla CAR, que fue el primer abordaje organizado y sistemático del uso de computadoras para recoger y analizar datos de modo de mejorar las noticias.

El CAR fue usado por primera vez en 1952 por CBS para predecir los resultados de la elección presidencial. Desde la década del '60 periodistas (en su mayoría de investigación y de Estados Unidos) han buscado controlar el poder de modo independiente analizando bases de datos de registros públicos con métodos científicos. También conocido como “periodismo de servicio público” los partidarios de estas técnicas con ayuda de computadoras han buscado revelar tendencias, demostrar la falsedad de creencias populares y revelar injusticias perpetradas por autoridades y corporaciones privadas. Por ejemplo, Philip Meyer trató de demostrar la falsedad de los informes de los disturbios en Detroit de 1967, para reflejar que no eran solo sureños poco educados los que participaban. Las historias de Bill Dedman sobre “El Color del Dinero” en la década del '80 reveló prejuicios raciales sistémicos en las políticas de crédito de las principales instituciones financieras. En su artículo “Lo Que Salió Mal” Steve Doig buscó analizar los patrones de daños del huracán Andrew a comienzos de la década del '90, para comprender el efecto de las políticas y prácticas de desarrollo urbanas fallidas. Los reportes basados en datos han generado valiosos servicios al público y permitido a los periodistas ganar importantes premios.

A comienzos de la década del '70 el término *periodismo de precisión* fue acuñado para describir este tipo de recolección de noticias: “la aplicación de métodos de investigación de las ciencias sociales y de la conducta a la práctica del periodismo” (de “**The New Precision Journalism**”, por Philip Meyer). Se creó el periodismo de precisión para que fuera practicado en las principales instituciones de medios por profesionales formados en periodismo y ciencias sociales. Nació en respuesta al “nuevo periodismo”, una forma de periodismo en el que las técnicas del periodismo se aplican a las noticias. Meyer sugiere que lo que se necesita son técnicas

científicas de recolección y análisis de datos, en vez de técnicas literarias, para que el periodismo pueda cumplir con su cometido de objetividad y verdad.

Se puede entender el periodismo de precisión como una reacción frente a algunas de las fallas y debilidades comúnmente citadas: la dependencia de informes de prensa (lo que se describió luego como “churnalismo”), el prejuicio en favor de fuentes autorizadas, etc. Meyer ve que estas debilidades derivan de la falta de aplicación de técnicas científicas de información y métodos científicos tales como encuestas y registros públicos. En los ‘60, el periodismo de precisión fue utilizado para representar a grupos marginales y sus historias. Según Meyer:

El periodismo de precisión era una manera de expandir el instrumental del periodista para hacer que temas antes inaccesibles o sólo accesibles de modo tosco, estuvieran abiertos a la investigación periodística. Fue especialmente útil para dar voz a grupos minoritarios y disidentes que luchaban por lograr representación.

Un artículo influyente publicado en la década del ‘80 respecto de la relación entre el periodismo y las ciencias sociales se hace eco del discurso sobre el periodismo de datos. Los autores, dos profesores de periodismo estadounidenses, sugieren que en las décadas de los años ‘70 y ‘80, la comprensión del público de lo que son las noticias se amplía, de una concepción más estrecha de “eventos noticiosos” al “reporte situacional” (o informes sobre tendencias sociales). Por ejemplo, al usar bases de datos de censos o encuestas, los periodistas logran “ir más allá de la información de eventos específicos, aislados, para proveer contexto que les da significado”.

Como era de esperar, la práctica de usar datos para mejorar el periodismo existe desde que hay datos. Como señala Simon Rogers, el primer ejemplo de periodismo de datos en The Guardian data de 1821. Es una tabla de escuelas en Manchester que da la cantidad de estudiantes que asisten a clases y los costos por escuela, Según Rogers, esto ayudó a mostrar el número real de estudiantes que recibían educación gratuita, que era mucho mayor de lo que mostraban las cifras oficiales.

DAY SCHOOLS.— <i>Establisments</i>	Boys	Girls	Total	Ann. Exp.	Remarks.
Grammar School	155	155	310	1800
Blue Coat ditto	50	50	100	2000	Taught, clothed and boarded.
Green Coat ditto	50	50	100	200	Taught and clothed.
Colligate Church ditto	10	10	20	80	And offertory money; <i>dit. do.</i>
St. Mary's ditto	12	12	24	100	(Supp.)—Taught and clothed.
St. John's ditto	9	9	18	60	Funds arising from Sacramental Offerings.
St. Paul's ditto	30	30	60	200	(Supp.)—Expenses raised by voluntary Subscription.
Ladies' Jubilee	30	30	60	250	Taught, clothed and boarded, by voluntary Subscription.
Black Kingstreet	21	21	42	80	(Supp.)—Taught and partly clothed. This School is supported by the benevolence of a single individual.
National Schools, Crosby-st.	194	119	313	600	Voluntary Subscription, and Collections at Churches.
..... <i>Edinboro-st., Salisbury</i>	300	170	470
Directors.	692	325	1017	400	Voluntary Subscription.
..... <i>Marshall-st.</i>	35	35	50	<i>Ditto ditto</i>
..... <i>Mosley-street</i>	198	121	319	104
SUNDAY SCHOOLS.	890	981	1871	£554
<i>Establisments.</i>
Collegiate Church, Shude Hill	201	205	406
St. Asch's, Black Kingstreet	50	56	106
St. Mary's, Back South Parade	120	110	230
St. Paul's, Green-street	150	163	313
..... <i>Turner-street</i>	68	71	139
..... <i>Jersey-street</i>	314	281	595
St. George's, St. George's	111	112	223
St. John's, St. John's-street	118	162	280
St. James's, St. James's-street	102	198	300
St. Michael's, Mill-street	254	352	606
St. Peter's, Jackson's-row	129	129
..... <i>Alport Town</i>	55	55	110
St. Clement's and St. Luke's	332	1071	1403
..... <i>Barnet-street</i>	181	207	388
..... <i>Oldfield-road</i>	139	201	340
Trinity, King's Head Yard	220	300	520
Holmes, Duke-street	185	189	374
All-Saints, Oxford road	196	191	387
Arwick	60	110	170
Total	3434	4219	7653	£2078

11, June 9, at 10, at Guildhall, London. N. Moter, Tomlinson and Bennett, Thompson street.

LEWIS Jonathan, of Sunderland, in the county of Durham, grocer, d. c. j. May 15, 19, June 9, 12, at Guildhall, London. Atts. Messrs. G. Hobbs and Gatty, Throgmorton-street.

PAYN Thomas, and John Daniel Paine, of Colchester, in the county of Essex, warehousemen d. c. j. May 5, 12, June 9, at 1, at Guildhall, London. Atts. Messrs. Boscawell-street.

SMITH John, now ex-lieut. of Fattington, in the county of York, grocer, in draper, d. c. j. May 11, 14, June 9, at 11, at Dog and Duck Tavern, King's-cross-Hall. Mr. Walsley, Hall.

TATE John, of Liverpool, in the county of Lancashire, provision-merchant, d. c. j. May 17, June 9, at 1, at the George Inn, Liverpool. Mr. Denton, Liverpool.

WARD Joseph, late of Basbury, in the county of Oxford, (but now a prisoner in the King's Prison), lawyer, d. c. j. May 3, 12, June 9, 12, at Guildhall, London. Atts. Messrs. F. and Munday, Holborn.

WHAETON Robert and Henry Wharton, of Lincoln, in the county of Lincoln, iron-works, late carpenters, late carpenters in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Atts. Mr. Hodgson, Liverpool.

WILLIAMS Daniel, of Princeton-street, Bath, in the county of Somerset, master-merchant, d. c. j. May 5, at 10, May 13, at June 9, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.

DECEASED.

May 22. Mansell & Atkinson, London, merchants. J. Davies, Silverbury, Essex-physician. W. Beeley, Manchester, tailor. W. & A. Copp, Exeter, drapers. W. Bunter, Exeter, draper and tailor. J. Dakin, Southampton, Kent, draper & tailor. J. Lee, London, warehouseman. T. Casaly, Liverpool, feather-merchant. J. Williams, London, draper. M. B. Schlessinger London, indigo-merchant.

June 5. Ryder & Nanseth, London, sugar-refiners. JOHN HARRISON and Brothers, Manchester, cut spicers. — Widow Welch and Sons, Manchester common carriers. — Geo. Ramsden and Co. M.

Figure 11. Periodismo de datos en The Guardian en 1821 (The Guardian)

Otro ejemplo temprano en Europa es de Florence Nightingale y su informe clave, "Mortalidad del Ejército Británico", publicado en 1858. En su informe al parlamento usó gráficos para promover mejoras en los servicios de salud para el ejército británico. El más famoso de ellos es su "coxcomb", una espiral de secciones que representan muertes por mes, en el que se destaca que la gran mayoría de las muertes eran por enfermedades prevenibles, en vez de balas.

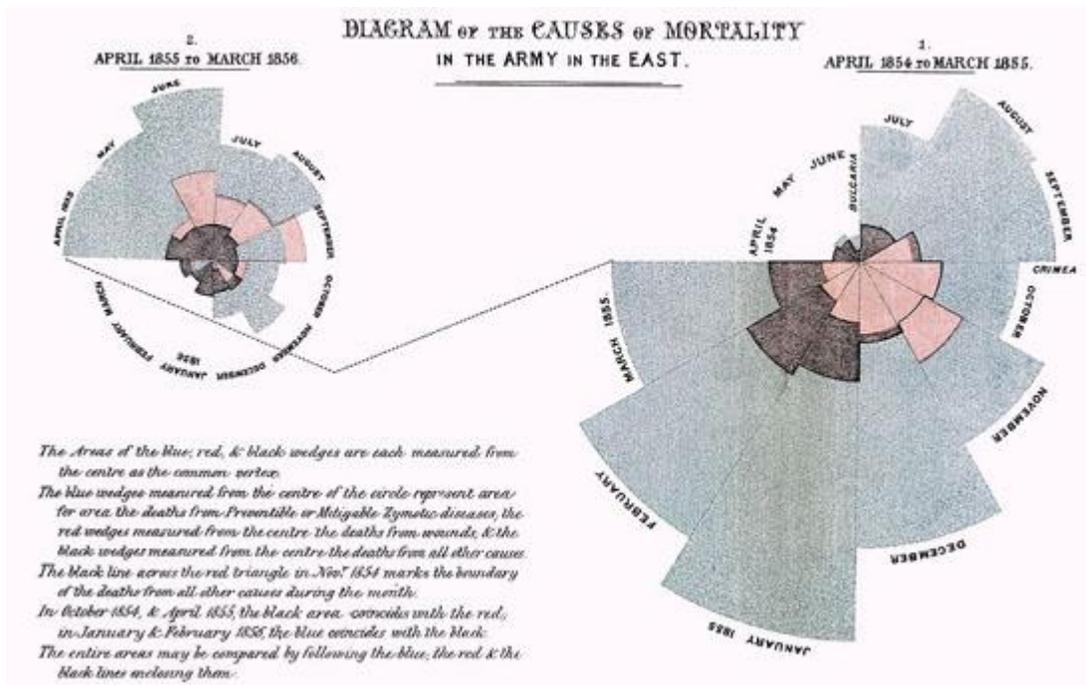


Figure 12. Mortalidad de la armada británica por Florence Nightingale (imagen de Wikipedia)

Periodismo de Datos y Periodismo Asistido por Computadora

En este momento hay un debate sobre “continuidad y cambio” en torno de la etiqueta de “periodismo de datos” y su relación con prácticas previas periodísticas que emplean técnicas computacionales para analizar conjuntos de datos.

Algunos sostienen que hay una diferencia entre CAR y el periodismo de datos. Dicen que CAR es una técnica para recoger y analizar datos como una manera de fortalecer el periodismo (generalmente de investigación), mientras que el periodismo de datos presta atención a la manera en que los datos se ubican en el conjunto del flujo de trabajo periodístico. En este sentido el periodismo de datos presta tanta –y a veces más- atención a los datos mismos, en vez de usarlos simplemente como un medio para encontrar o dar más fuerza a determinadas historias. De allí que encontremos el Datablog de The Guardian o que el Texas Tribune publica juegos de datos junto con los artículos –o incluso solo juegos de datos por sí mismos- para que la gente analice y explore.

Otra diferencia es que en el pasado los periodistas de investigación se encontraban faltos de información respecto de una pregunta que trataban de contestar, o una cuestión que trataban de abordar. Si bien esto sigue sucediendo, también existe

una abundancia abrumadora de información con la que los periodistas a veces no saben qué hacer. No saben cómo obtener valor de los datos. Un ejemplo reciente es el Sistema de Información Online Combinada, la mayor base de datos del Reino Unido de información sobre gasto público. Esta base de datos fue durante mucho tiempo un reclamo de los partidarios de la transparencia, pero dejó confundidos y sin respuesta a muchos periodistas cuando se publicó. Como me escribió recientemente Philip Meyer: “Cuando la información era escasa, la mayor parte de nuestros esfuerzos estaban dedicados a buscarla y recogerla. Ahora que hay información abundante, el procesamiento es más importante”.

Por otro lado, algunos sostienen que no hay ninguna diferencia significativa entre el periodismo de datos y el periodismo asistido por computadoras. A esta altura resulta claro que incluso las prácticas más recientes de los medios más novedosos combinan cosas conocidas desde hace tiempo con algo nuevo. Antes que debatir si el periodismo de datos es completamente nuevo, una postura más fructífera sería considerarlo como parte de una tradición más longeva, pero que responde a nuevas circunstancias y condiciones. Aunque no haya una diferencia en cuanto a metas y técnicas, el surgimiento de la etiqueta “periodismo de datos” al comienzo del siglo indica una nueva fase en la que el mero volumen de los datos libremente disponibles online –combinado con herramientas sofisticadas centradas en el usuario, la auto edición y las herramientas de colaboración abierta (crowdsourcing)- permite a más gente trabajar con más datos de modo más fácil que nunca.

El periodismo de datos tiene que ver con la alfabetización masiva en el manejo de datos.

Las tecnologías digitales y la red están cambiando de modo fundamental la manera en que se edita la información. El periodismo de datos es una parte del ecosistema de herramientas y prácticas que han surgido en torno a los sitios y servicios de datos. El citado y el compartir materiales de distintas fuentes es parte de la naturaleza de la estructura de hipervínculos de la red, y la manera en que estamos acostumbrados a navegar la información hoy. Yendo más hacia atrás, el principio que está en la base de la estructura de hipervínculos de la red es el principio de la cita usado en los trabajos académicos. Citar y compartir materiales y sus fuentes y los datos detrás de la historia es una de las maneras básicas en las que el

periodismo de datos puede mejorar el periodismo, lo que el fundador de WikiLeaks Julian Assange, llama el “periodismo científico”.

Al permitir a cualquier persona recurrir a fuentes de datos y encontrar información que es relevante, así como verificar afirmaciones y cuestionar los supuestos comunes, el periodismo de datos representa efectivamente la democratización masiva de recursos, herramientas, técnicas y metodologías que antes eran usadas por especialistas, fueran estos periodistas de investigación, científicos sociales, estadísticos, analistas u otros expertos. Si bien actualmente citar vínculos con fuentes de datos es algo específico del periodismo de datos, avanzamos hacia un mundo en el que los datos estarán integrados sin fisuras en el tejido de los medios. Los periodistas de datos tienen un rol importante en cuanto a ayudar a bajar las barreras a la comprensión y el manejo de datos, e incrementar la alfabetización en datos de sus lectores a escala masiva.

En este momento la comunidad creciente de personas que se llaman periodistas de datos es en gran medida diferente de la comunidad CAR más madura. Esperemos que en el futuro veamos vínculos más fuertes entre estas dos comunidades, del mismo modo que vemos a ONG y organizaciones de medios sociales como ProPublica y el Bureau of Investigative Journalism trabajando junto con medios tradicionales en investigaciones. Mientras la comunidad de periodismo de datos puede tener formas más innovadoras de difundir datos y presentar historias, el enfoque profundamente analítico y crítico de la comunidad CAR es algo de lo que el periodismo de datos podría aprender.

— *Liliana Bounegru, European Journalism Centre*

HomeCapítulo: IntroducciónSiguiente: En la redacción