



# LA NAVAJA SUIZA DEL REPORTERO

Herramientas de investigación  
en la era de los datos masivos

Elaborado por David Hidalgo y Fabiola Torres







OjoPúblico

# LA NAVAJA SUIZA DEL REPORTERO

Herramientas de investigación  
en la era de los datos masivos

Elaborado por David Hidalgo y Fabiola Torres



Consejo de la Prensa Peruana

## ÍNDICE

### La navaja suiza del reportero

© David Hidalgo y Fabiola Torres

© Asociación de Periodismo de Investigación OjoPúblico  
Jr. Pablo Bermúdez Nro. 150, Int. 11-A, Urb. Santa Beatriz  
(Frente Al Parque De La Reserva). Lima - Lima - Lima.

© Consejo de la Prensa Peruana  
Calle Ignacio Merino Nro. 616. Lima - Lima - Miraflores

Esta publicación ha sido posible con el apoyo de la Fundación Hivos y el Instituto para la Democracia y la Asistencia Electoral - IDEA.

Investigación: Fabiola Torres y David Hidalgo  
Asistente de investigación: Karina Valencia  
Diseño y diagramación: Kati Sanabria

1a edición, Febrero 2016  
Hecho el Depósito Legal en la Biblioteca Nacional del Perú N° 2016-01479

Tiraje: 500 ejemplares  
ISBN: 978-612-47141-0-8

Impreso en:  
HAPANA CORP S.A.C.  
Mza. R Lote. 43, Urb. Santa Elisa, Lima - Lima - Los Olivos  
Febrero, 2016.

### Presentación

5

...

### I. El nuevo alfabeto del periodista: de cómo el encuentro con un hacker aceleró la reinención del periodismo.

11

.....

Caja de seguridad: herramientas para evitar que te espíen en Internet / Guía para investigar con datos / Open Refine, un software con esteroides: cómo detectar errores entre millones de datos / Catálogo en línea: aplicaciones desarrolladas en el Perú / Recursos digitales para cazar historias: los utensilios de escritorio para el reportero de investigación.

### II. ¿Cómo rastrear delitos en una base de datos? Veinte investigaciones que cambiaron la manera de hacer periodismo.

39

.....

Investigaciones con data filtrada / Investigaciones con data pública / 6 instrumentos para contar mejores historias: cómo enriquecer un relato con gráficos, infografías y hasta frases sonoras / Investigaciones con bases construidas / El poderoso Neo4J: o cómo descubrir un fraude global con nodos y aristas / Cartografía digital: recursos para ubicar los hechos y personajes en el lugar exacto.

### III. El camino hacia una cultura de innovación: los laboratorios digitales del periodismo de investigación en el Perú.

69

.....

Dos tecnólogos apoyan al periodismo / Estudio de caso: Cuidados Intensivos. News apps o la noticia que nunca muere / Workshops para reporteros: organizaciones aliadas para combinar periodismo y tecnología / Habemus o no habemus data: Ley de transparencia vs. Ley de protección de datos personales / El dilema existencial: ¿Cuándo un dato privado es de interés público? / La letra pequeña: cómo hacer un pedido de información eficaz.



## PRESENTACIÓN

**C**uando OjoPúblico salió en línea en el 2014 nos propusimos innovar en los diferentes ámbitos del periodismo en Perú y Latinoamérica. Desde mejorar en la publicación de contenido de investigación e implementar nuevas narrativas digitales para quebrar los tradicionales formatos, hasta explorar el modelo de negocio ideal mientras intentamos construir una moderna redacción integrada por periodistas y tecnólogos. Todos agrupados bajo la siguiente prédica: contar historias de calidad que otros medios prefieren obviar, reportajes incómodos para el poder, pero relevantes para la opinión pública y necesarios para el ciudadano.

“La navaja suiza del reportero: herramientas de investigación en la era de los datos masivos” es la primera prueba en formato editorial de este precepto periodístico. Patrocinado por Hivos de Holanda en alianza con el Consejo de la Prensa Peruana (CPP), y escrito por David Hidalgo y Fabiola Torres, es una poderosa radiografía del trabajo del reportero en esta época de información calculada en terabytes. Didáctico, riguroso y ágil, este manual profundiza en los mejores trabajos realizados por el periodismo de datos contemporáneo.

En la guía que usted tiene en sus manos podrá identificar a los equipos de élite de la prensa mundial mientras marcan tendencia a nivel planetario con investigaciones basadas en el análisis y la visualización de millones de documentos. Desde la filtración del Caso Wikileaks en el año 2010 -que originó la primera alianza entre dos mundos para contar una historia sin precedentes: la organización del hacker y programador Julian Assange y la redacción de The Guardian, uno de los diarios más innovadores del mundo-, hasta los desfiles mundiales del Consorcio Internacional de Periodistas de Investigación (ICIJ) sobre los paraísos fiscales, en los años siguientes.

El lector de “La navaja suiza del reportero” recorrerá una selección de los principales casos de periodismo de datos a nivel inter-

nacional y conocerá las herramientas digitales más usadas por las mejores redacciones del mundo para limpiar, analizar y visualizar grandes cantidades de información. Desde las más conocidas como Open Refine, creada por Google hace más de un lustro y una de mis favoritas, hasta otras más complejas como el Neo4J, utilizada por el equipo de OjoPúblico -este año, en plena campaña electoral- para crear un especial inédito sobre el financiamiento de los partidos políticos y las campañas presidenciales del Perú en la última década.

En el manual de casi 100 páginas -una suerte de crónica sobre los reporteros en tránsito al uso de la tecnología- el lector descubrirá que los periodistas de investigación ya no pueden trabajar aislados como lobos solitarios. La simbiosis creada con los escritores de código ha llevado esta profesión a horizontes desconocidos en la exploración de documentos que antes hubieran sido imposibles de analizar por cualquier medio de comunicación en el mundo. También recordamos la esforzada labor de los pocos periodistas nacionales que intentaron trabajar con datos en la década secretista de los '90, en pleno gobierno autoritario de Alberto Fujimori, hasta los inicios del nuevo siglo.

En tiempos en los que los periodistas escriben bajo la sombra de una industria de medios declarada en vías de extinción, presionados por la generación inmediata de contenido viral y sometidos a la dictadura del tráfico fácil en internet, “La navaja suiza del reportero” aporta ideas y soluciones para descubrir las mejores prácticas del periodismo en la era digital. Si en aquella breve imagen de la película *Spotlight*, en la que aparece un periodista del Boston Globe frente a una hoja de cálculo, podemos avistar los orígenes del periodismo de investigación en la era de los datos masivos, en “La navaja suiza del reportero” podremos vislumbrar nuestro presente y, quizá, nuestro futuro.

*Óscar Castilla C.*  
Director Ejecutivo  
OjoPúblico

## PRÓLOGO

“ La navaja suiza del reportero: herramientas de investigación en la era de los datos masivos” fue un esfuerzo inicialmente concebido como una guía práctica para difundir y promover el uso de data por parte de periodistas y activistas interesados en explorar, interpretar y contrastar la información disponible en el infinito universo digital. Sin embargo, la publicación que usted tiene en las manos, elaborada por David Hidalgo y Fabiola Torres –Director Periodístico y Editora de Análisis de Datos de la Asociación de Periodismo de Investigación OjoPúblico–, superó nuestras expectativas. Para el Consejo de la Prensa Peruana representa, en principio, una antología sobre el desarrollo y potencia de un género de periodismo insuficientemente desarrollado en nuestro medio. Pero sobre todo es un manual que revela cómo esta profesión, al incorporar el uso de la tecnología, ha dado como resultado algunas de las mejores investigaciones de los últimos tiempos y tiene un amplio margen para elevar sus estándares en beneficio de los ciudadanos y de la cultura democrática.

“La navaja suiza del reportero” entra en circulación en un momento oportuno: el año 2016 marca un punto de quiebre en materia del derecho ciudadano de acceso a la información pública. En la esfera oficial observamos la creciente erosión de los mecanismos anticorrupción establecidos para impulsar temas fundamentales como la lucha contra la cultura del secreto, intensamente arraigada en el Perú. Del mismo modo, se hace evidente el incumplimiento de los compromisos de transparencia, participación ciudadana y rendición de cuentas asumidos por el Perú en el marco de la Alianza por un Gobierno Abierto.

Lo paradójico es que este retroceso coincidió con un giro en las prioridades establecidas por el ente rector en esta materia, la Secretaría de Gestión Pública de la Presidencia del Consejo de Ministros.

Desde mediados del 2015, esta entidad priorizó la implementación de la Estrategia de Apertura y Reutilización de Datos Abiertos Gubernamentales en el Perú, en el contexto de las iniciativas desplegadas por el Gobierno peruano con el fin de alcanzar los estándares establecidos para integrar la Organización para la Cooperación y el Desarrollo Económico (OCDE).

Pese al tiempo transcurrido, las iniciativas planteadas desde el Estado son insuficientes e inconsistentes.

Por suerte no sucede lo mismo en otros escenarios del quehacer nacional. El país hoy cuenta con un joven, pero dinámico y creciente grupo de activistas que utilizan la data disponible para desarrollar herramientas y producir información y liberarla con el fin de mejorar la calidad de vida de los ciudadanos. Ejemplo de ello son los cada vez más frecuentes hackatones, concursos y retos para premiar a quienes desarrollen aplicaciones en todos los ámbitos. En nuestro país, buena parte de ese esfuerzo corre a cargo de Open Data Perú, una comunidad que ya es parte del movimiento internacional de datos abiertos y celebra cada marzo el Open Data Day.

Para el Consejo de la Prensa Peruana ha sido una experiencia gratificante unir esfuerzos con OjoPúblico en el Proyecto para la Promoción de Datos Abiertos en el Perú, a través de actividades que fueron auspiciadas por la Fundación HIVOS y que incluyeron, por ejemplo, la organización del seminario “Periodismo de datos abiertos: la agenda pendiente de la prensa peruana”, en agosto del 2015.

No dudamos de que “La navaja suiza del reportero” se convertirá en una herramienta de consulta para los periodistas que procuren encontrar antecedentes exitosos, ideas sugerentes y métodos eficaces para investigar historias relevantes e insospechadas a partir de las bases de datos.

Por ello extendemos nuestro sincero agradecimiento a la Fundación HIVOS y a IDEA Internacional, que se sumó a la iniciativa e hizo posible la estuenda edición que hoy presentamos.

Para nuestra institución, este esfuerzo coincide con un nuevo ámbito de actividades que iniciamos en el 2015 y seguiremos desarrollando en los próximos años: se trata del proyecto Escuela de Periodismo de Datos, cuyo objetivo es capacitar a periodistas peruanos en el uso de nuevas tecnologías y fuentes de información para enriquecer su trabajo y el impacto del mismo en la sociedad. Es una iniciativa que cuenta con el impulso de IBM, OjoPúblico, Open Data Perú y el Consejo de la Prensa Peruana. Estamos seguros de que este esfuerzo tendrá grandes y edificantes frutos.

*Kela León Amézaga*

Directora Ejecutiva

Consejo de la Prensa Peruana




**Junta Directiva 2016-2018**

Presidente: Bernardo Roca-Rey Miró Quesada.

Vicepresidenta: María Eugenia Mohme Seminario.





# EL NUEVO ALFABETO DEL PERIODISTA

De cómo el encuentro con un hacker aceleró la reinención del periodismo

1

## El ecosistema del periodismo digital

---

David Leigh  
Paul Radu NSA Investigación  
Programación Simon Rogers Hack Hackers  
The Guardian Snowden Pro Publica  
Scott Klein Excel Dan O’Huiginn  
Encriptado Wikileaks Greenwald  
Assange Base de datos Big Data  
Paul Bradshaw Informática

Cuando ya se había convertido en el hombre más buscado del planeta, el analista estadounidense Edward Snowden se sentó junto a un par de periodistas y grabó con voz tranquila una frase que podría ser un salmo del futuro. “La tecnología es el más grande ecualizador de la historia humana”.<sup>1</sup> Al momento en que lo dijo, estaba oculto en la habitación de un hotel de Moscú, a riesgo de ser capturado por haber revelado la mayor maquinaria de vigilancia clandestina que haya existido jamás. En lugar de un pedido de ayuda, era un mensaje sobre el verdadero sentido de la revolución digital. “Nos ayuda a adoptar nuevos rostros, entrar en nuevas comunidades, participar en nuevas conversaciones y descubrir quiénes somos y en qué nos queremos convertir”. Snowden proclamaba una lucha contra los poderes que pretenden usar la tecnología para decir si somos buenas o malas personas. “No son los gobiernos quienes deben decidirlo. Somos nosotros”. Su modo de contribuir a esa lucha fue convertirse en la mayor fuente periodística de todos los tiempos.

Snowden es el arquetipo de esta era de informantes globales. Cuando salió a la luz, el soldado Bradley Manning ya estaba preso por haber filtrado una gran cantidad de secretos y Julian Assange estaba asilado en la embajada de Ecuador en Londres. A diferencia de Manning, que dejó un reguero de pistas personales antes de la filtración, y de Assange, que hizo del activismo un culto a la personalidad, Snowden solo hizo su primer movimiento después de un cálculo al milímetro que le permitió aparecer de la nada y hacer planteamientos con cierto aire visionario, muy bien estructurado, sobre las repercusiones de su filtración ya no para la sociedad estadounidense, sino para cada individuo en el planeta. El eje de esa estrategia era

---

<sup>1</sup> PLATON. “The Most Wanted Man in the World: Edward Snowden in His Own Words – WIRED”. Video. En: <https://www.youtube.com/watch?v=LgA7DoptXg0> [Visualizado: 21 de noviembre de 2015]



asegurarse la intervención de Glenn Greenwald, un periodista conocido por sus antecedentes como abogado de derechos humanos y por su cobertura del espionaje masivo cometido por la NSA, la agencia de inteligencia especializada en el recojo de información y análisis de datos. El contacto entre ambos es una prueba a escala de los desafíos del periodismo para entender y registrar hechos de interés público en la sociedad digital de este tiempo. El informante más famoso desde Garganta Profunda no solo filtró secretos al periodista. Tuvo que alfabetizarlo en recursos técnicos antes de ponerse a trabajar.

El primer contacto se produjo en diciembre del 2012. Snowden envió a Greenwald un e-mail con seudónimo que empezaba con una defensa de la seguridad de las comunicaciones entre las personas. Siempre desde el anonimato, el autor del mensaje explicaba al investigador que utilizar una cuenta normal de correo ponía en peligro a la gente que quisiera transmitirle información sensible. No era una sorpresa, las historias sobre el espionaje a través de Internet habían generado titulares durante casi toda la primera década del siglo XXI y en los últimos años se habían concentrado en el papel de los gigantes de la comunicación en línea para facilitar la vigilancia gubernamental en países de poca tradición democrática como China o Siria. La fuente misteriosa sugería que el periodista se instalara un programa de encriptación, un software que permite codificar desde las contraseñas hasta los mensajes para hasta hacer imposible que una tercera persona, agencia o gobierno, interceptara sus comunicaciones. Incluso se ofrecía a darle una mano si le resultaba difícil. “Hacía tiempo que yo quería usar software de encriptación”, escribe Greenwald en el libro *Sin lugar donde esconderse*, que cuenta detalles de la investigación. “Sin embargo, el programa es complicado, sobre todo para alguien como yo, poco ducho en programación y ordenadores. Era una de estas cosas para las que nunca encuentras el momento”.<sup>2</sup>

<sup>2</sup> GREENWALD, Glenn. “Sin lugar donde esconderse. Edward Snowden, la NSA y el estado de vigilancia de EEUU.”. Barcelona: Ediciones B, 2014.

**“Las computadoras no hacen bueno a un mal reportero. Lo que hacen es convertir a un buen reportero en uno mejor”.**

.....  
**Elliot Jaspin**, Premio Pulitzer 1979  
de Periodismo de Investigación  
.....

## CAJA DE SEGURIDAD

[Herramientas para evitar que te espíen en Internet]



### Tor Project

<https://www.torproject.org/>

Software gratuito para comunicaciones seguras. Oculta la dirección de IP de los dispositivos utilizados. Permite navegar sin ser detectado ni dejar rastro de sitios visitados o la ubicación geográfica del usuario.

### Cryptocat

<https://crypto.cat/>

Servicio de mensajería instantánea privada en la web que puede ser usado desde cualquier navegador y en celulares.

### Mozilla Thunderbird

<https://www.mozilla.org/en-US/thunderbird/>

Programa de correo libre y de código seguro para recibir, enviar y almacenar mensajes electrónicos. Se puede gestionar varias cuentas de correo con un solo programa.

### Enigmail

[www.enigmail.net](http://www.enigmail.net)

Complemento del Thunderbird que permite enviar correos protegidos con claves cifradas. El usuario mantiene su clave. Para usar Enigmail, se debe instalar también GNU Privacy Guard (GnuPG).

El programa al que se refería Greenwald se llama PGP, siglas en inglés de *Pretty Good Privacy* (Seguridad Muy Buena). Es una herramienta popular entre hackers y toda clase de gente que vive en riesgo de ser espiado. Funciona con una clave especial que uno debe intercambiar con el remitente para establecer un contacto seguro. “En esencia, el programa envuelve los e-mails con un escudo protector que es una contraseña compuesta por centenares, incluso miles de números aleatorios y letras sensibles a las mayúsculas”, cuenta Greenwald en su libro. Incluso los programas de desciframiento más avanzados de las agencias de inteligencia más poderosas tardarían años en vulnerar esa protección. Aunque Greenwald conocía sus beneficios al haber escrito sobre casos como el de Wikileaks o Anonymous, no lo había incorporado a sus herramientas ni estaba dispuesto a dedicarle tiempo.

Días después, el remitente anónimo volvió a escribirle con una serie de instrucciones para instalar el programa. Incluso le ofreció contactarlo con algún experto para que le ayudara a empezar. Greenwald ofreció hacerlo, pero tampoco movió un dedo. Tenía mucha carga de trabajo y nada le garantizaba que hacer el esfuerzo trajera a cambio una gran historia. Semanas después, el personaje desconocido insistió en facilitarte las cosas con un video tutorial que llevaba por título: “PGP para periodistas”. Ni siquiera por eso, Greenwald se puso en acción. Tampoco lo haría en los dos meses siguientes. Para entonces, el informante había buscado otra vía para seguir con sus planes: buscó a la documentalista Laura Poitras por la sencilla razón de que ella sí usaba programas de encriptación. Fue Poitras quien supo por primera vez las dimensiones de la fuente y sus secretos. “Así de cerca estuve de perder las filtraciones más importantes y trascendentales para la seguridad nacional en la historia de Estados Unidos”, reconocería Greenwald. Su suerte estuvo en que Snowden insistió en trabajar con él.

Tiempo después, tras una serie de medidas de seguridad que incluyeron crear nuevos e-mails encriptados, más claves y la ayuda



de un experto en seguridad informática como intermediario, la propia Laura Poitras encargó a este tecnólogo aliado que le enseñara a Greenwald un sistema todavía menos conocido llamado *Tails* (siglas de *The Amnesic Incognito Live System*), que solo se usa desde un dispositivo portátil.<sup>3</sup> El experto preparó una versión especial para el reportero en un USB azul y se lo envió por correo hasta Brasil. Es el accesorio que aparece conectado a la computadora de Greenwald mientras entrevista a Snowden en una habitación de Hong Kong para el documental *Citizenfour*. Así comenzó la historia. En su actual perfil de Twitter, Glenn Greenwald señala que tiene disponible su llave pública de PGP y su respectiva huella digital, un número más corto, de apenas 40 dígitos, que facilita la confirmación de la llave. También es una huella del aprendizaje que representó su encuentro con Snowden en un fragmento de la historia en que la tecnología ya no es un accesorio, sino parte del hábitat en que transcurre la experiencia humana.

Desde la filtración más impactante del siglo, existe la idea de que un día los periodistas tendremos que imitar a los astrónomos para capturar certezas en el expansivo universo digital. No todos vamos a escribir sobre el espionaje de los servicios secretos occidentales ni sobre los planes de ciertos gobiernos para capturar Internet –aunque deberíamos tenerlo en agenda–, pero tan valioso como eso es tener los instrumentos para entender esta era de los datos. Incluso antes de los atentados del 11 de septiembre, ya se sabía que la mayor agencia de inteligencia estadounidense estaba interceptando 1.700 millones de comunicaciones en un día. Con las revelaciones de Snowden se supo que en el 2013 la misma agencia alcanzó a capturar un billón de metadatos que permiten desde saber qué busca la gente en Internet, cuáles son sus aficiones e incluso cuál será su comportamiento en el futuro inmediato. Ese mismo año, un catedrático estadounidense

<sup>3</sup> LEE, Micah. “Ed Snowden taught me to smuggle secrets past incredible danger. Now I teach you”. The Intercept. En: <https://theintercept.com/2014/10/28/smuggling-snowden-secrets/> [Visualizado: 22 de noviembre de 2015]

calculó que el volumen de información almacenado en el mundo era de 1.200 exabytes, equivalente a cubrir de libros toda la superficie de Estados Unidos unas cincuenta y dos veces.<sup>4</sup> Casi no hay proceso que no pueda ser cuantificado. ¿Cómo entender una cantidad de datos que, puestos en CD, daría para formar una torre hasta la Luna?

En esencia, se requiere un cambio en las capacidades técnicas y en el pensamiento operacional del periodista de investigación. La clásica metáfora del oficio tendría que cambiar de tener una caja de herramientas a manejar un equipo de laboratorio. “Los periodistas no necesitan aprender a programar, pero sí necesitan desarrollar mentalidad de datos masivos, de manera en que entiendan que los datos contienen historias que pueden quedar sin ser contadas”, dice el experto en Internet y catedrático de Oxford Víktor Schönberger.<sup>5</sup> En la anterior gran revolución del periodismo -el llamado Nuevo Periodismo- se hablaba de las escenas, el diálogo y la perspectiva como las técnicas experimentales que habían cambiado la forma de contar buenas historias. La nueva artesanía incluye dominar suficientes conceptos para pensar en una redacción periodística como un taller de robótica: encriptar, scrapear, refinar, visualizar. Un metalenguaje para definir cómo manejamos los datos.

No son las aplicaciones ni el software los que definen el nuevo momento del periodismo, sino la posibilidad que nos ofrecen de encontrar respuestas a preguntas que antes parecían de ciencia ficción. “Los datos pueden revelar secretos a quienes tengan la humildad, el deseo y las herramientas para escuchar”, escribió Víktor Schönberger junto a Kenneth Cukier en el libro *Big data, la revolución de los datos masivos*. También en eso nos parecemos a los astrónomos: nos enfrentamos a un universo tan grande que exige mejorar nuestros instrumentos de un día para el otro.

4 SCHÖNBERGER, Viktor y Kenneth CUKIER. “Big data, la revolución de los datos masivos”. Madrid: Turner, 2013.  
5 GONZALO, Marilín. “Los datos masivos (o big data) son el nuevo oro”. En eldiario.es (España), 5 de agosto del 2013. En: [http://www.eldiario.es/turing/Big-data\\_0\\_161334397.html](http://www.eldiario.es/turing/Big-data_0_161334397.html) [Visualizado: 22 de noviembre de 2015]



**“Necesitamos humanizar y personalizar los grandes datasets en una forma que no desvirtúe la complejidad o la escala de los temas que se trata”.**

.....  
**Paul Bradshaw**, autor del *Online Journalism Blog*.  
.....

## BIG (REALLY BIG) DATA

A inicios del siglo XXI, el economista estadounidense Steven Levitt imaginó una economía de lo raro. Se trataba de encontrar verdades sorprendentes a partir de la forma en que se analizaban los datos. Levitt lo hizo llevando al extremo la lógica de hacer preguntas: ¿Por qué continúan viviendo los traficantes de drogas con sus madres? o ¿qué resulta más peligroso, una pistola o una piscina? En uno de sus ensayos planteó a sus lectores el caso de una pareja que evita enviar a su hija a casa de una vecina cuyo padre guarda una pistola en casa, pero la deja ir donde otra que vive en una casa con piscina en el patio. La pregunta fue cuál decisión era la correcta para la seguridad de la niña. Levitt encontró que de acuerdo a las estadísticas, hay una muerte infantil por cada 11 mil piscinas, lo que en un país con 6 millones de piscinas supone una media de 550 niños ahogados al año. En cambio, un niño muere de un disparo por cada millón y medio de armas. En un país con 200 millones de armas, la proporción de 175 niños muertos de un disparo al año. Traducción: una niña tiene estadísticamente más probabilidades de morir en una piscina que jugando con el arma del papá de su vecina. Si el ser humano suele modificar sus comportamientos con muestras a escala, ¿qué pasa cuando la información desborda nuestra capacidad de almacenarla? “La era de los datos masivos pone en cuestión la forma en que vivimos e interactuamos con el mundo”, explican Schönberger y Cukier.

Su caso predilecto fue cuando el buscador más famoso del mundo salvó a Estados Unidos de una epidemia global. Cuando la gripe aviar llegó a ese país, los sistemas de salud colapsaron ante la falta de información oportuna para plantear estrategias. El sistema de alertas era demasiado lento para entender la propagación de la enfermedad. Por esos días, la revista científica *Nature* publicó un artículo en que ingenieros de Google demostraban haber encontrado la manera de predecir la propagación de la gripe común. El método consistía en combinar las tendencias de búsqueda acerca de los síntomas con in-



# GUÍA PARA INVESTIGAR CON DATOS

El periodista Paul Bradshaw, autor del Online Journalism Blog, plantea que el trabajo periodístico con bases de datos comprende cinco etapas: recojo, depuración, análisis, verificación y presentación de hallazgos. Podemos tomar esa secuencia para plantear el siguiente ejercicio a la hora de iniciar un proyecto.

## El recojo de los datos



Debes conocer los formatos de los archivos que contienen la información y las herramientas para recopilarlos. Se puede obtener datos masivos mediante un script, un programa simple que permite bajar información de manera automatizada. A este proceso se le conoce como scraping.

¿Hay bases de datos sobre el tema? ¿Cómo y para qué fueron hechas? ¿Están en una web oficial o debo hacer un pedido de acceso a la información?

Si la base está en una web, ¿es descargable o debo hacer scraping?

¿Cuál es el mejor formato (Excel, CSV, Json) para solicitar una copia de esas bases de datos? Si la información está en PDF o JPG, ¿cómo la paso a un Excel?

Si debo construir una nueva base, ¿qué variables debo incluir y qué podré demostrar?

## Depuración y contexto



Puede haber errores de varios tipos: registros duplicados, casillas incompletas, palabras mal escritas, etc. Necesitas herramientas para identificar y resolver esos problemas. A esto se le llama limpiar la data.

¿La base de datos está completa? ¿Cuántas líneas de información tiene? ¿Puedo limpiarla con Excel u Open Refine? ¿En qué casos debo hacerlo a mano? ¿En qué caso debo usar gestores con más capacidad, como MongoDB?

¿Conozco y entiendo todos los términos, variables y siglas que aparecen en las bases? ¿Son los mismos que los usados en bases similares? ¿Los criterios apuntan al sentido de la pregunta que quiero responder o necesito ver esos mismos datos en sentido inverso?

## Cruce y análisis



En esta fase, el valor de sus hallazgos depende de la calidad de las preguntas y de la combinación de dos o más registros para encontrar coincidencias reveladoras.

¿Tienen mis bases de datos un concepto o código común que me permita cruzarlas: DNI, RUC, nombres completos?

¿El cruce de las bases de datos muestra tendencias, patrones, procesos evolutivos en un periodo determinado? ¿En qué contexto?

O, por el contrario, ¿revela comportamientos atípicos? ¿En qué contexto?

## Verificación



El periodista de investigación debe aplicar la metodología tradicional: ir a los lugares necesarios, entrevistar a las personas involucradas, revisar nuevos documentos para detectar las debilidades y fortalezas de la base de datos.

¿Reflejan los datos la condición real de las personas? ¿Ha variado algo en la vida del aludido, en su salud, estabilidad económica, situación legal o sus vínculos?

¿Influye eso en el sentido del hallazgo? ¿Confirma su relevancia, la acentúa o la relativiza?

¿Con qué experto puedo validar la metodología del cruce? ¿Es posible que el hallazgo sea correcto, pero admita más de una interpretación?

## Presentación



Hay que pensar desde el inicio en la forma más eficiente de presentar los hallazgos. Existen librerías como d3.js.org y repositorios de software como Github, con ejemplos que uno puede adaptar a lo que necesita.

¿Conviene más una visualización o una aplicación? ¿Cuál aporta al sentido de la historia?

¿Cómo debe ser la experiencia del usuario? ¿Qué es lo que el gráfico o la herramienta debe generar en el lector? ¿Qué elementos de mi aplicación o visualización la hacen necesaria para el usuario?

¿Es responsiva? ¿Se verá bien en teléfonos móviles y tablets? ¿Se puede compartir? ¿Se puede embeber?

## OPEN REFINE, UN SOFTWARE CON ESTEROIDES

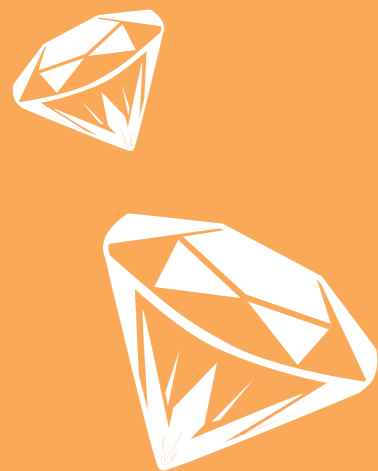
[Cómo detectar  
errores entre  
millones de datos]

Todo el que trabaje con hojas de cálculo sabe que existen **cuatro problemas comunes**: faltas ortográficas, nombres o palabras escritas de varias formas, caracteres invisibles o espacios en blanco. No parecen complicados para quien use un archivo de Excel personal, pero son una pesadilla a la hora manejar bases de datos con millones de líneas. En estos casos, **lo más útil es trabajar con Open Refine**, una herramienta de código abierto que **permite depurar y organizar datos en pocos pasos**. Tomemos como muestra un registro de empresas exportadoras de oro. Es posible que el nombre de una compañía haya sido llenado de varias formas (OroGoldSA, OroGoldS.A. y OroGold). Un análisis inicial las contabilizará como empresas diferentes. **Open Refine encuentra las coincidencias y permite editarlas con una sola acción** para darles uniformidad.

Además, **si se comete algún error al editar la base de datos, es posible volver al estado anterior** con la opción que muestra el historial de cambios realizados.

Este programa se puede descargar de <http://openrefine.org/download.html>, es compatible con cualquier navegador y **está disponible para Windows, Mac y Linux**. Además permite transformar archivos de diferentes extensiones como XLS, CSV, JSON, XML, TSV y las spreadsheets de Google.

**Es un recurso muy necesario en la caja de herramientas del reportero de investigación**. Algunos periodistas suelen definirlo como “la versión de Excel con esteroides”.



**300**

mil registros de  
compañías inscritas en  
Panamá fueron descargados  
por el programador Dan  
O’Huiginn para crear un  
sitio que facilita las  
búsquedas.

formación histórica sobre la evolución de la enfermedad. “Otros ya habían intentado hacer esto con los términos de búsqueda de Internet, pero nadie disponía de tantos datos, capacidad de procesarlos y *know-how* estadístico como Google”, escriben los expertos. Tan solo validar las palabras o frases clave supuso recrear 450 millones de modelos matemáticos diferentes. El resultado fue un grupo de 45 términos que mostraban la relación entre las búsquedas de los posibles pacientes y la evolución de la gripe. El detalle es que, a diferencia del modo tradicional, que podía tomar semanas para recuperar la información, Google elaboró un software que lograba esa precisión en tiempo real.

¿Qué podemos aprender los periodistas de este tiempo en que un algoritmo puede predecir el momento en que millones de personas se limpiarán la nariz? “Los *big data* -los datos masivos- se refieren a cosas que se pueden hacer a gran escala [...] para extraer nuevas percepciones o crear nuevas formas de valor, de tal manera que transforman los mercados, las organizaciones, las relaciones entre los ciudadanos y los gobiernos”, señalan Schönberger y Cukier. Desde que el psiquiatra Carl Jung descifró cómo sueñan los seres humanos ninguna operación para encontrar historias en el mundo abstracto fue tan poderosa.

“Los datos son sagrados”, dice un viejo lema del diario inglés *The Guardian*. Ahora es el principio que guía el trabajo de su antiguo editor digital Simon Rogers el periodista que convirtió la estadística en un cuadro de Mondrián. “La mayor parte del tiempo actuamos como el puente entre los datos (y aquellos que están desesperados en explicarlos) y la gente del mundo real que trata de entender de qué va realmente una historia”, dice Rogers en *Facts are sacred*, una guía de cómo convertir datos en conceptos visuales. En el 2011, el equipo liderado por Rogers –quien ahora dirige el laboratorio digital de Google– explicó el aparato administrativo civil de Inglaterra como si fuera un racimo de globos de colores. De un vistazo, los lectores pudieron entender una estructura burocrática que original-

# DIEZ BASES DE DATOS PARA RASTREAR EL DINERO Y OTRAS PISTAS

Una sugerente selección recogida por **OjoPúblico** entre los asistentes a la Conferencia Global de Periodismo de Investigación 2015.

## ICIJ Offshore Leaks Database

<http://offshoreleaks.icij.org/search>

Tiene datos de más de 100 mil compañías y fondos guardados en paraísos fiscales. Es parte de un archivo de 2,5 millones de documentos filtrados al Consorcio Internacional de Periodistas de Investigación (ICIJ).

## OpenCorporates

<https://opencorporates.com/>

Contiene información sobre 80 millones de compañías y 90 millones de directores en más de 100 países. Se puede buscar por nombre de empresa, dirección y nombres de los directores.

## Personas de interés

<https://www.personadeinteres.org>

Da acceso a archivos judiciales, registros de propiedades e informes de inteligencia sobre gente vinculada al crimen organizado, narcotráfico, corrupción, etc. Tiene información sobre procesados por narcotráfico en Perú.

## Open Spending

<https://openspending.org/>

Sitio web que sigue el gasto público de los gobiernos en todo el mundo y lo presenta en distintas formas de visualización.

## Registro de propiedades de Miami

<http://www.miamidade.gov>

Base de datos libre del Condado de Miami Dade, que permite localizar propiedades en Miami, por nombre del dueño.

## Investigative Dashboard

<https://www.investigativedashboard.org/>

Permite buscar información de los accionistas, directores y los reportes financieros de compañías alrededor del mundo. Hay enlaces a más de 450 bases de datos en línea de 120 países. Plataforma construida por el Organized Crime and Corruption Reporting Project (OCCRP).

## Search Systems

<http://publicrecords.searchsystems.net>

Portal especializado en la búsqueda de personas. Contiene más de 55 mil bases de datos organizadas por fecha de nacimiento, fecha de defunción, matrimonio, licencias, acciones, hipotecas, entre otras subdivisiones.

## Registro de compañías de Panamá

<http://ohuiginn.net/panama/>

Sitio independiente que reordenó la data del registro público de Panamá para facilitar la investigación sobre más de 300 mil empresas inscritas en este país. Permite hacer búsquedas por nombre de personas.

## FlightAware

<https://es.flightaware.com/>

Es una plataforma para rastrear los vuelos y su estado en todo el mundo. Ofrece una aplicación descargable en el celular. Permite buscar por el nombre de la línea aérea, el número de vuelo, la ruta y la matrícula de la aeronave.

## Marine traffic

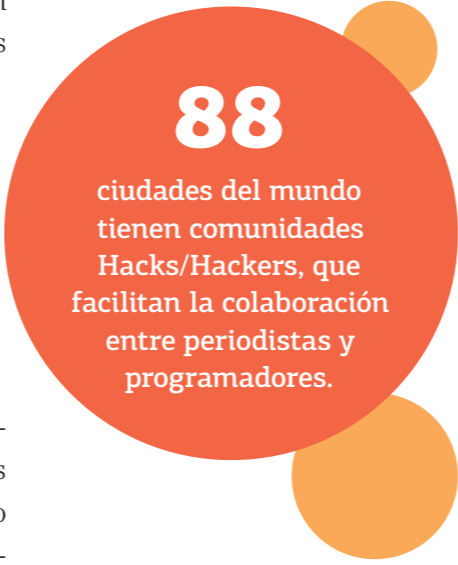
<http://www.marinetraffic.com/>

Es una base de datos actualizada en tiempo real que permite rastrear la localización de cualquier tipo de barco o navío, además de las salidas, llegadas y recorridos.

mente era una lista con más de 200 mil nombres, sus posiciones y su salario. El análisis permitió saber que al menos 90 burócratas ganaban más que el Primer Ministro inglés.

En otra ocasión, los periodistas David Leigh y Nick Davies, del equipo de investigación de *The Guardian*, obtuvieron de Wikileaks un archivo con información detallada acerca de todos los incidentes militares registrados por el ejército estadounidense durante la guerra en Afganistán. La habían creado soldados encargados de monitorear las acciones. El primer paso fue recibir la información encriptada por e-mail. El problema fue revisar los datos y encontrar la historia: el archivo contenía una hoja de cálculo con 91 mil 201 filas. Era un volumen demasiado abrumador para ser analizado por los periodistas y hasta los expertos en sistemas del periódico encontraban dificultades para manejarlo. “Es como encontrar pequeñas pepitas de oro en medio de una montaña de datos”, llegó a decir Leigh.<sup>6</sup>

Para ese momento, el diario tenía experiencia con grandes bases de datos liberadas por el gobierno e incluso había creado exploradores internos que permitían a los periodistas hacer búsquedas. Esta vez volvieron a hacerlo. Los datos fueron filtrados de acuerdo a un orden que incluía fecha, hora, descripción de los ataques, número de víctimas y las coordenadas donde habían ocurrido. El análisis permitió establecer que el número de ataques con artefactos caseros –los más impredecibles y letales– habían aumentado, y que ese aumento se había producido justamente en las zonas controladas por los ejércitos de Reino Unido y Canadá. Ahora el reto era encontrar la mejor manera de contar esa historia. Fue entonces que un equipo de visualizadores, encabezado por Simon Rogers, se incorporó al trabajo. “El proyecto Wikileaks estaba produciendo nuevos tipos de datos, así que necesitaban ser extraídos con nuevos tipos de periodismo”, escribió Leigh junto a Luke Harding, otro de los reporteros que participó en la investigación.<sup>7</sup> El resultado fue



**88**  
ciudades del mundo  
tienen comunidades  
Hacks/Hackers, que  
facilitan la colaboración  
entre periodistas y  
programadores.

**“Nunca comienzo un proceso si no desarrollo un plan. Utilizo una matriz propia, que nos ayuda a apuntar a preguntas concretas”.**

.....  
**Ginna Morelo**, Editora de Datos de *El Tiempo* (Colombia)  
.....

un mapa que mostraba por primera vez la evolución de seis años de atentados en ese país: entonces se confirmó que la racha sangrienta había dejado más muertos civiles que militares.<sup>8</sup>

“La historia de Wikileaks es una combinación de las dos cosas: los conocimientos del periodismo tradicional y el poder de la tecnología, unidos para contar una historia alucinante”, escribió la dupla que reconstruyó el caso.

Aún con estas muestras del potencial de los datos, hasta hace muy poco muchos periodistas, incluso los de investigación, tendían a pensar que la tecnología era un idioma extraterrestre. La sola idea de indagar en un archivo de Excel con más de mil filas desanimaba al tradicional cazador de documentos confidenciales. “No tienes que ser un programador”, precisa Rogers en su libro. “Puedes convertirte en codificador de lujo si así lo quieres, pero la tarea principal es pensar en los datos como un periodista en vez de como un analista”. Ninguna herramienta digital va a reemplazar el ejercicio de preguntar qué información relevante pueden brindar los datos o qué pasaría si uno mezcla una base con otra, a la manera del original economista de lo raro. Lo que uno no pueda hacer por su cuenta, puede hacerlo con aliados del universo paralelo.

## EL CÓDIGO DE LA FUERZA

Era cuestión de tiempo que alguien inventara un espacio para integrar dos maneras de ver la información. La idea partió de una coincidencia entre un joven corresponsal y dos periodistas veteranos. Burt Herman era un reportero de la agencia Associated Press que llevaba doce años viajando por zonas sensibles del mundo, desde Corea y algunos países de la ex Unión Soviética hasta los convulsionados Irak y Afganistán. Entre el 2008 y el 2009, Herman dejó la agencia y optó por una beca para explorar innovaciones periodísticas en la Universidad de Stanford. Desde allí, en la vorágine digital de Silicon Valley, empezó a organizar reuniones de gente interesada

6 LEIGH, David y Luke HARDING. “Wikileaks y Assange”. Barcelona: Ediciones Deusto, 2011.

7 Op. cit.

8 ROGERS, Simon. “Facts are sacred”. Londres: Faber and Faber Limited, 2013.



en el periodismo y la tecnología. Por esa misma época, Aron Pilhofer, editor del *The New York Times*, y Rich Gordon, profesor de la Northwestern University, lanzaron desde Massachusetts una convocatoria para formar una red que desarrollara aplicaciones y herramientas digitales para procesar información. Ambas iniciativas coincidían en un concepto: unir a los hacks, un término que alude a la capacidad de los periodistas para producir textos en serie, con los hackers, que son prolíficos escritores de código fuente, el conjunto de instrucciones que hace funcionar las máquinas.<sup>9</sup>

Semejante cruce de lenguajes daría para un episodio de *Star Wars*: es como si dos razas alienígenas –una respecto de la otra, al menos– hubieran llegado a un acuerdo para cumplir una misión. El único modo posible es intercambiar conocimientos: los periodistas aprenden de los hackers la jerga y principios que rigen el ciberespacio y a cambio los entrenan para usar sus habilidades con el fin de dar sentido a la información. La prueba está en la experiencia del propio Burt Herman. Mientras realizaba su beca, entró en contacto con el programador belga Xavier Damman y juntos se propusieron crear una herramienta que aprovechara el potencial informativo de las redes sociales. El resultado fue Storify, una aplicación que permite reunir fotos, videos, tuits y links para contar una historia que puede ser insertada en cualquier sitio web. “La manera de dar sentido a los medios de comunicación social es a través de la curaduría humana con la ayuda de la tecnología”, ha dicho Herman.

Esta alianza ya está generando cambios en el periodismo global: la comunidad HacksHackers tiene capítulos en ciudades de todos los continentes. En cada lugar ha facilitado la creación de herramientas que permiten procesar grandes cantidades de información. A inicios del 2014, por ejemplo, el capítulo de Rosario, en Argentina, reunió datos del ministerio de Justicia, reportes policiales y artículos de prensa y construyó un mapa interactivo en que uno puede ver el

**“Cuando diseño una hoja de cálculo para una nota pienso en las metas: qué quiero saber y cuáles son los posibles patrones en los datos”.**

.....  
**Lise Olsen**, reportera de investigación del *Houston Chronicle*.  
.....

## CATÁLOGO EN LÍNEA

[Aplicaciones periodísticas desarrolladas por OjoPúblico]



### Fondos de Papel

<http://fondosdepapel.ojo-publico.com/>

Reportaje interactivo que ofrece una radiografía inédita sobre el financiamiento privado de los partidos políticos del Perú y sus candidatos presidenciales en las campañas del 2006, 2011 y 2016.

### Billeteón electoral

<http://ojo-publico.com/sites/apps/billeteon-electoral/>

Aplicación que permite evidenciar la desigualdad económica al comparar los ingresos de los usuarios con los de los principales candidatos presidenciales peruanos en las elecciones del 2016.

### Congreso Airlines

<http://ojo-publico.com/sites/apps/congreso-airlines/>

Aplicación que muestra los viajes internacionales de 113 congresistas del Perú y los gastos que realizó el Poder Legislativo para cubrirlos entre el 2011 y 2015.

### Cuentas Juradas

<http://cuentasjuradas.ojo-publico.com/>

Plataforma que muestra la evolución del patrimonio declarado entre el 2003 y 2014 por 38 alcaldes de Lima que buscaban la reelección y de 23 exalcaldes que querían volver al poder en las elecciones municipales del 2014.

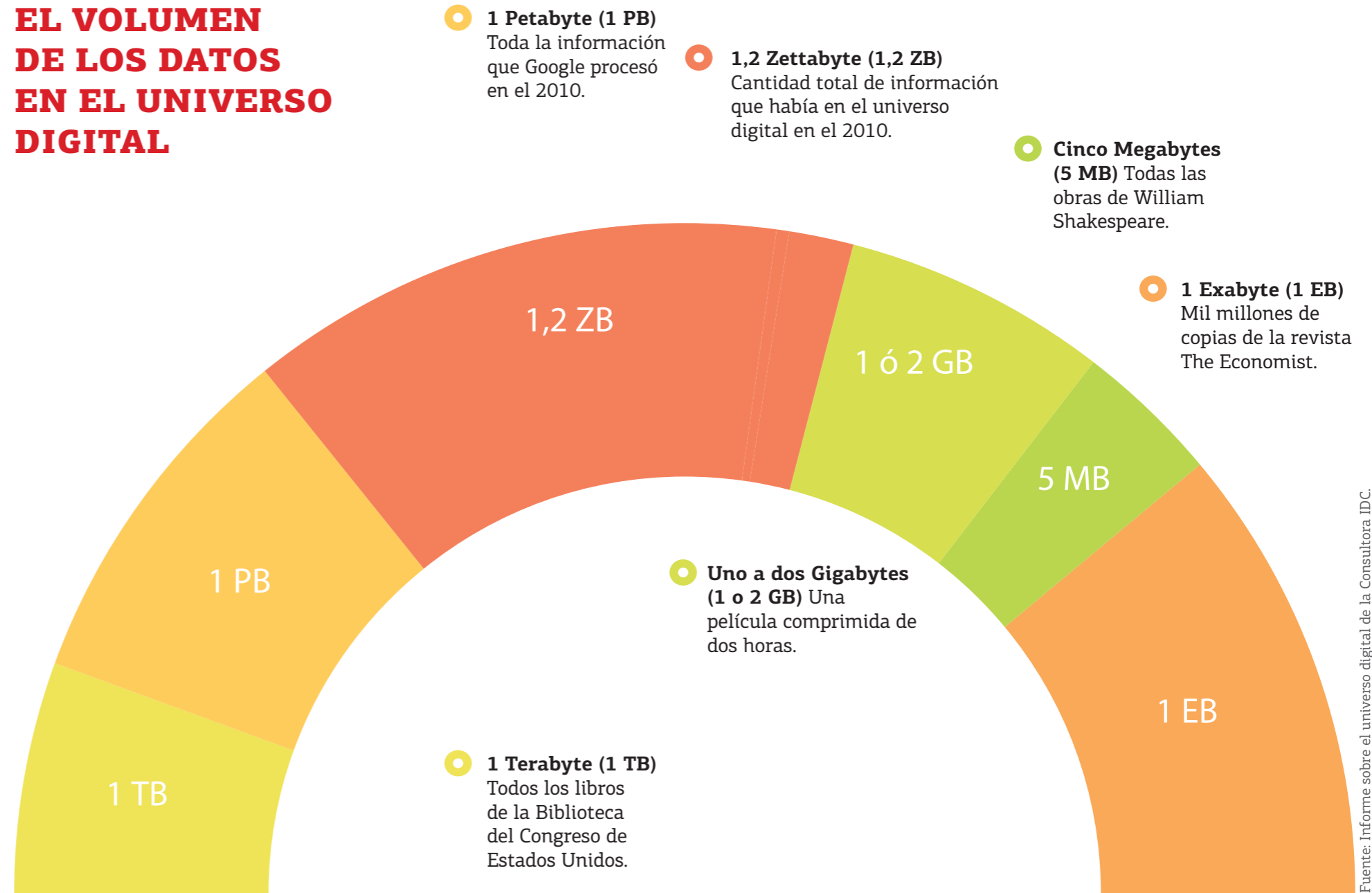
<sup>9</sup> HIDALGO, David. “Periodistas buscan hackers (de los buenos)”. En: <http://hhlima.info/node/8> [Visualizado: 25 de noviembre de 2015]

punto exacto en que se produjo cada homicidio ocurrido el año anterior. “El propósito del proyecto era crear una plataforma que permitiera demostrar, a través de la visualización de datos, el incremento de la violencia social en la ciudad”, ha escrito Ezequiel Clerici, uno de los organizadores de esa comunidad. Tiempo atrás, en el 2011, el capítulo de Buenos Aires creó una aplicación que permitía seguir en tiempo real los resultados de las elecciones presidenciales: basta marcar en un mapa el lugar que a uno le interesa y obtiene el dato correspondiente. En cada caso se cumple el último axioma de la era de la información: el problema no son los cambios en los métodos del periodista, sino qué entendemos por hacer periodismo.

Como es de esperarse en la época de los datos masivos, el potencial es enorme. En junio del 2014, varios capítulos de América Latina se unieron en una hackatón regional para generar herramientas que permitieran a los periodistas vigilar el uso de los fondos públicos. La actividad fue bautizada con la expresión que ha guiado siempre al mejor periodismo: “La ruta del dinero”. El capítulo de Lima reunió a más de cincuenta miembros que se encerraron durante doce horas en el auditorio de un instituto dedicado a la enseñanza en tecnología. La jornada consistió en formar equipos mixtos, de periodistas y desarrolladores, para analizar información dura de distintas bases de datos y convertirla en noticia. En esencia, se trataba de buscar el lado sexy de una tabla de Excel.

Al final de la tarde, la comunidad presentó siete proyectos, que iban desde el análisis de cómo se invierten los fondos del Ministerio del Ambiente hasta un cálculo del dinero destinado por el Estado a la Iglesia Católica. Una de las herramientas más interesantes era una aplicación que permitía realizar búsquedas en bases de datos y páginas especializadas para identificar vínculos de funcionarios con el crimen organizado y su posible relación con fondos públicos. Otra era una aplicación que permitía sistematizar los datos sobre los principales proveedores del Estado y cuánto dinero han facturado al país. De no haber sido por ese encuentro, que funcionó como un

## EL VOLUMEN DE LOS DATOS EN EL UNIVERSO DIGITAL



Fuente: Informe sobre el universo digital de la Consultora IDC.

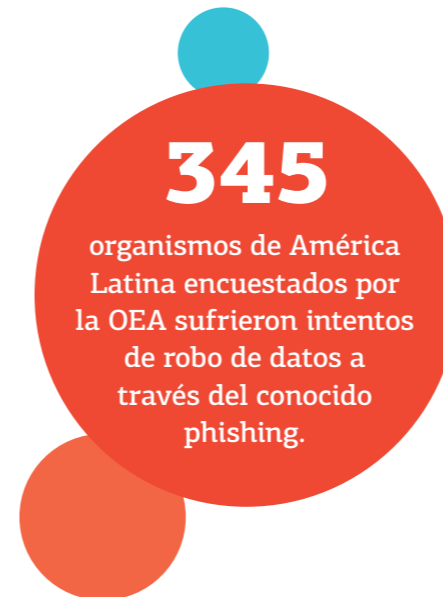
taller para curiosos, la información seguiría sepultada en archivos inconexos. “Este grupo reúne a todas estas personas: los que están trabajando para ayudar a la gente a darle un sentido a su mundo”, dice una declaración en la página original de HacksHackers.

## LOS ALIADOS

Una tarde de octubre de 2010, el ingeniero costarricense Rigoberto Carvajal decidió renunciar a su trabajo en una corporación de desarrollo de software para enrolarse en el periodismo. Por medio de un amigo se había enterado de que el diario La Nación buscaba a un programador para su Unidad de Investigación. Carvajal se presentó por la curiosidad de saber cómo un tipo como él podía ser útil para destapar secretos. Tras una primera entrevista, quedó convencido de que era el lugar correcto. “Estudié programación porque me gusta mucho resolver problemas y desde el periodismo uno puede hacerlo con fines más nobles que incrementar las ganancias de una empresa”, dice Carvajal, quien ahora es uno de los expertos en bases de datos del Consorcio Internacional de Periodistas de Investigación (ICIJ).<sup>10</sup>

El proyecto para el que estaba convocado era un desafío para un periódico latinoamericano en esa época: reunir todas las bases de datos públicas de su país para analizarlas, obtener conocimiento nuevo e historias relevantes por investigar. Los informáticos del equipo debían tener la misma audacia y pasión por la verdad que los periodistas. En la entrevista de trabajo, Carvajal comentó que en cierta ocasión logró ubicar el paradero de una persona de la que solo tenía el nombre, gracias a que se las ingenió para contactar personas que le dieron la información adecuada. Su primera prueba demostró que mantenía la habilidad. Cuando le pidieron detectar las propiedades de Shakira en EE.UU., el programador buscó el verdadero nombre de la cantante y en seguida lo rastreó en registros de propiedades hasta lograr su objetivo. Con el tiempo, Rigoberto Carvajal –un fanático de la ciencia ficción– se transformó en una suerte de *Spock*,

<sup>10</sup> Entrevista a Rigoberto Carvajal.



el personaje entre humano y extraterrestre de *Star Trek*: un híbrido entre la computación y el periodismo.

“Volver a lo que hacía antes en la programación, para la parte comercial, sería como volver al lado oscuro de la fuerza”, dice en alusión a otra de sus películas favoritas.

El mismo espíritu animó al británico Dan O’Huiginn cuando se planteó descargar todo el registro público de Panamá en el 2008 para una investigación sobre traficantes de armas. Hasta ese momento, la web oficial solo permitía búsquedas en base de los nombres de las empresas. Eso representaba una limitación para el trabajo de los reporteros que seguían la pista de personajes sospechosos. O’Huiginn extrajo los datos de más de 300 mil compañías, ordenó la información, y la utilizó para crear un sitio que admitía las búsquedas basadas en nombres de individuos.<sup>11</sup> El programador suele explicar que ese trabajo no tuvo nada que ver con la piratería ilegal. Solo usó sus habilidades técnicas para automatizar la recopilación de datos. “No me importa que me llamen hacker en el sentido literal”, dijo cuando los medios panameños se sorprendieron de que un sujeto desconocido, desde un terminal en otro continente, lograra ese acceso. Su esencia, explicó, es la de “una persona que disfruta explorando los detalles de los sistemas programables y cómo extender sus capacidades, a diferencia de la mayoría de los usuarios, que prefieren aprender solo el mínimo necesario”.<sup>12</sup>

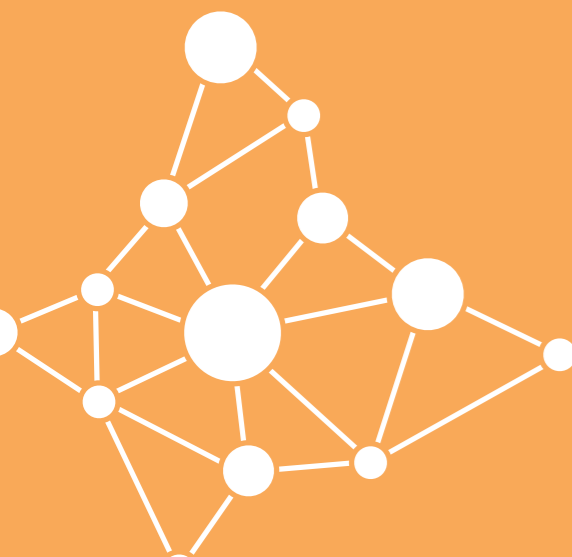
El sitio creado por Dan O’Huiginn, que recibe 2 mil visitas diarias, permitió a periodistas de investigación de muchos países verificar si funcionarios bajo sospecha de enriquecimiento ilícito y corrupción tenían propiedades registradas en secreto en Panamá. Con esta herramienta, en el 2011 la reportera Khadija Ismayilova demostró que las hijas del presidente de Azerbaiyán, Ilham Aliyev, dirigían una compañía de telecomunicaciones a través de empresas *off-shore*.

<sup>11</sup> <http://ohuiginn.net/panama/> [Visualizado: 22 de noviembre de 2015]

<sup>12</sup> “El Registro Público solicitó que eliminara mi web”. En Diario La Estrella (Panamá), 6 de octubre de 2013. En: <http://laestrella.com.pa/panama/nacional/registro-publico-solicito-eliminar/23502395>. [Visualizado: 22 de noviembre de 2015]

## RECURSOS DIGITALES PARA CAZAR HISTORIAS

[Los utensilios de escritorio para el reportero de investigación]



### ScraperWiki

<https://scraperwiki.com/>

Plataforma online que permite descargar información de la web y agruparla de forma ordenada en una base de datos (Excel, CSV, etc.). Ofrece la posibilidad de que cualquier persona cree su propio script de acuerdo a sus intereses.

### DocumentCloud

<https://www.documentcloud.org/>

Plataforma para gestionar documentos. Extrae el texto de una imagen mediante software OCR. Permite resaltar datos, hacer anotaciones y organizarlas en links de fácil acceso. Ayuda a buscar por temas, embeber documentos, y colocarlos en un catálogo público. Se accede previa postulación.

### Visual Investigative Scenarios (VIS)

<https://vis.occrp.org/>

Una herramienta de especial interés para los periodistas de investigación. Permite establecer relaciones entre personas y organizaciones y adjuntar documentos que prueben esa relación.

Dicha corporación tenía más de un millón y medio de suscriptores, cubría el 80 por ciento del territorio de ese país y en aquel tiempo era la única proveedora de servicios 3G. El sitio también permitió identificar a las empresas *off-shore* del expresidente de Egipto, Hosni Mubarak, y además proporcionó evidencia que permitió conectar a cinco personas con el asesinato del exgobernador de la provincia de Panamá, Darío Fernández. Todas fueron condenadas.

Desde el 2010, el programador O'Huiginn decidió entregarse de lleno al trabajo con periodistas de investigación. Ha sido becario del *African Network of Centers for Investigative Reporting* y colaborado con proyectos del *Organized Crime and Corruption Reporting Project (OCCRP)*. Hoy vive en Alemania y trabaja en el proyecto Openoil, el primer mapa de datos abiertos sobre concesiones petroleras en 18 países de Medio Oriente.

“Si se hace bien, la gente tiene realmente mucho apetito por ver los datos”, dice Scott Klein, un referente del periodismo de investigación con datos masivos.<sup>13</sup> “Es suficiente ver cuánta gente entiende –y le encantan– estadísticas deportivas increíblemente sofisticadas e impenetrables”. Si un lector promedio está dispuesto a leer páginas deportivas que parecen reportes de la bolsa de valores, ¿por qué no encontraría la utilidad de una herramienta que examina el sistema de salud o la calidad del agua que se toma en su localidad?

Klein es editor de *Propublica*, uno de los medios más innovadores de Estados Unidos. En el 2010 recibió la misión de montar un proyecto que parecía sacado de un laboratorio de Silicon Valley: el departamento de aplicaciones de noticias. Se trataba de un equipo formado por reporteros y tecnólogos que trabajaría de manera conjunta para hacer periodismo usando software. Uno de sus proyectos más impactantes fue *Dollars for Docs* (Dólares para los doctores), que reveló pagos por 258 millones de dólares a médicos que promovían entre sus pacientes los productos de siete empresas farmacéu-

<sup>13</sup> Citado en HOWARD, Alexander B. “El arte y la ciencia del periodismo de datos”. Tow Center de Periodismo Digital (Traducción de La Nación DATA).



ticas. El equipo aprovechó una ley que obligaba a los laboratorios a difundir la información sobre el dinero que dan a los médicos por comisiones, almuerzos y subvención para convenciones, entre otros conceptos. Todos los documentos habían sido publicados en los sitios web de las farmacéuticas, pero en formatos complicados de manejar. Los programadores de *Propublica* elaboraron un script –un pequeño *software* que automatiza procesos– para recoger toda esa información. Con ese material elaboraron una aplicación que permite a cualquier persona en Estados Unidos buscar a su doctor y saber si ha recibido dinero de los laboratorios, cuánto y por qué.

Klein suele referirse a su equipo de cinco personas como “programadores periodistas que piensan como reporteros”. Quiere decir que tienen habilidades para manejar herramientas digitales, pero a la vez tienen el instinto para detectar una buena historia en una montaña de datos. “Algunos han dicho que los periodistas con habilidades para desarrollar *software*, o viceversa, son unicornios, raros de encontrar. Eso no es verdad, usted puede desarrollar en ellos las habilidades de periodismo e ingeniería que se requieren hoy en día”, precisa.<sup>14</sup> El potencial es mayor al de solo hacer una visualización, que permite contar una historia en un gráfico, e incluso que una aplicación, que permite contar varias historias y que el lector haga sus propios hallazgos. Es la posibilidad de unir dos metodologías rigurosas y afines que expandan las fronteras del conocimiento. “El periodismo de investigación es el departamento de estudio y desarrollo de la profesión”, escribió Brant Houston, cofundador de la *Global Investigative Journalism Network*.<sup>15</sup>

Quizá la imagen que mejor refleja este momento ha sido la planteada por Evan Smith, cofundador del Texas Tribune. Smith dice que el periodista de hoy debe ser como una navaja suiza.<sup>16</sup> La idea



de mantener competencias separadas es un anacronismo comparable al de quienes en su momento se resistieron a dejar la máquina de escribir por la computadora. El editor lo explica con una escena que todavía debe enfrentar a pesar de que su medio trata de incidir en proyectos multimedia. “La gente viene y me dice: ‘Quiero ser un redactor’. ¿De qué tipo? ‘Solo quiero escribir, nada más que escribir’. ¿No quieres grabar un video con tu teléfono, editarlo y postearlo? ‘No’. ¿No quieres grabar audio en tu teléfono, editarlo y postearlo? ‘No’. ¿No quieres hacer nada en HTML básico? ‘No’. ¿No quieres encargarte de las redes? ‘No’. Entonces esto es lo que vas a hacer: ve a Home Depot (una tienda de materiales), compra mucha madera, construye una máquina del tiempo y regrésate al *Esquire* de 1964, porque esa fue la última vez que alguien tuvo ese trabajo”. La figura de la navaja no es retórica: puede que el reportero no use esas herramientas todos los días, pero la idea es que estén ahí cuando se necesiten.

14 FALLAS, Hassell. “Simplificar es clave para crear aplicaciones de noticias”. Blog La Data Cuenta. En: <http://hasselfallas.com/2014/09/11/simplificar-es-clave-para-crear-aplicaciones-de-noticias/> [Visualizado: 25 de noviembre de 2015].

15 Citado en: KAPLAN, David. “Periodismo de Investigación Global: Estrategias para su Financiamiento”. Center for International Media Assistance, 2013.

16 “Journalists today have to be swiss army knives”. Entrevista a Evan Smith en The future of news. En:

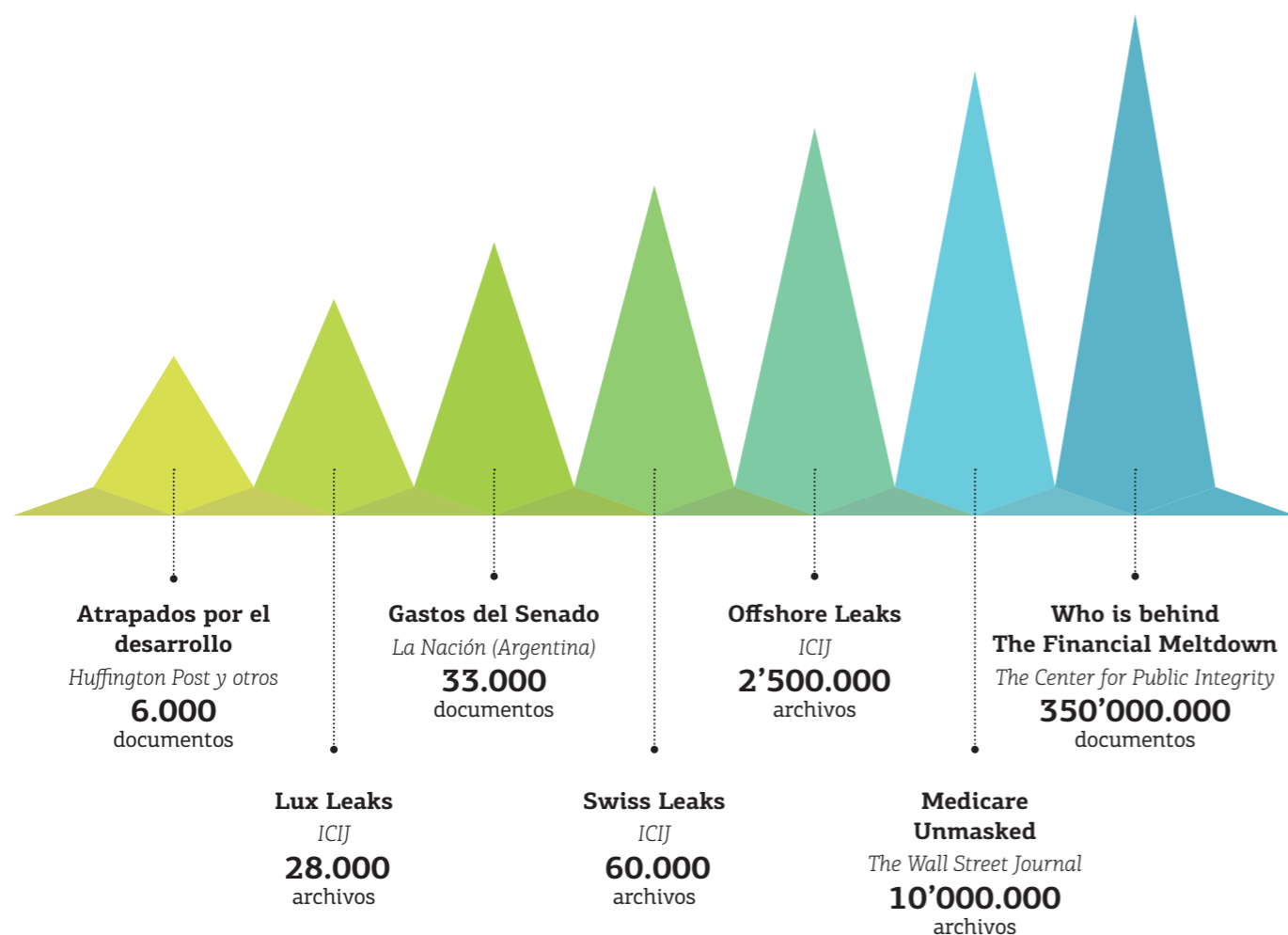
[http://futureof.news/episodes/evan-smith-2/?utm\\_campaign=refdotfonesmithvideo2&utm\\_source=Twitter&utm\\_medium=esmith2\\_jrsch\\_hd\\_flw](http://futureof.news/episodes/evan-smith-2/?utm_campaign=refdotfonesmithvideo2&utm_source=Twitter&utm_medium=esmith2_jrsch_hd_flw) [Visualizado: 25 de noviembre de 2015].



# ¿CÓMO RASTREAR DELITOS EN UNA BASE DE DATOS?

Veinte investigaciones que cambiaron  
la manera de hacer periodismo

## La dimensión de lo investigado



**D**esde que Wikileaks dinamitó la industria global de los secretos, el periodismo de investigación se ha visto envuelto en una fiebre por los datos. Ahora es posible rastrear la corrupción en varios continentes, detectar empresas y personajes que buscan evadir impuestos en todo el mundo, o entender los movimientos internacionales del crimen organizado. En setiembre de 2011, el periodista australiano Gerard Ryle, del Consorcio Internacional de Periodistas de Investigación (ICIJ), recibió un disco duro con 2.5 millones de archivos. Dos ingenieros en computación convirtieron esa marejada en una base de datos confiable. A partir de allí se realizó un intenso reporte que reveló las operaciones de más de 122 mil empresas y 130 mil personas en las sombras del sistema financiero mundial.

El hallazgo despertó elogios, pero también alertas. “El periodismo de investigación no debe confundirse con lo que ha sido etiquetado como ‘periodismo de filtraciones’”, dice David Kaplan. Esta observación es un misil al centro del debate sobre el encuentro entre periodismo y tecnología. ¿Son válidos los documentos obtenidos por hackeo? ¿Cómo hacer que nos digan lo que en verdad necesitamos saber? Hay que ver las hojas de cálculo como fichas forenses de la realidad: ofrecen detalles, pero la verdad requiere trabajo. “Las habilidades básicas de los periodistas de investigación –señala Kaplan– son similares a las de los más calificados fiscales y policías, antropólogos de campo e investigadores privados: el uso de fuentes primarias, verificación de la evidencia, entrevistas a testigos de primera mano, y el seguimiento a los rastros de personas, documentos, y el dinero”.

Las mejores muestras del periodismo de investigación reciente se han generado también por el acceso a información pública o la construcción de bases de datos nuevas, con información recopilada de diferentes fuentes, para responder a una pregunta que nadie se había hecho antes. A continuación algunos ejemplos notables.

# INVESTIGACIONES CON DATA FILTRADA

Pocos procesos muestran las peculiaridades de la época como la filtración de datos. A las nuevas formas de interconexión con las fuentes, mediante programas informáticos especializados, se suma el desafío de entender cantidades de información que podrían sepultar cualquier entusiasmo si no contáramos con la ayuda de los programadores. Estos casos lo demuestran.

## Offshore Leaks

### MEDIO

#### Consorcio Internacional de Periodistas de Investigación (ICIJ)

Fue un esfuerzo global liderado por el Consorcio Internacional de Periodistas de Investigación (ICIJ) con la colaboración de The Guardian, BBC, Le Monde, The Washington Post y una treintena de medios. Contó con el trabajo de 112 reporteros de 56 países.

Fecha: Abril 2013

### REVELACIÓN

Políticos, aristócratas, banqueros y criminales de diversos países usaron paraísos fiscales para crear compañías o fideicomisos con el fin de ocultar sus bienes o capitales, y en muchos casos evitar el pago de impuestos. En la lista aparecen el presidente de Azerbaiyán, Ilham Aliyev y su familia; Jean-Jacques Augier, tesorero de la campaña electoral del presidente francés Francois Hollande; la baronesa española Carmen Thyssen-Bornemisza, quien utilizó canales *offshore* para comprar obras de arte; y María Imelda Marcos, hija del exdictador filipino Ferdinand Marcos.

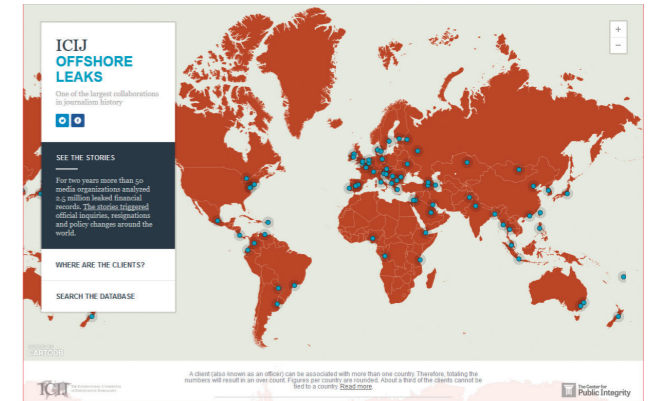
### ANÁLISIS DE DATOS

La filtración era 160 veces más grande que los documentos diplomáticos liberados hasta entonces por Wikileaks. El ICIJ contó con la ayuda de expertos en informática como Sebastian Mondial, de Alemania; Duncan Campbell y Matthew Flower, de Inglaterra; el costarricense Rigoberto Carvajal y el maltés Matthew Caruana. Ellos hicieron la limpieza y organización de los datos mediante la herramienta OpenRefine. El programa dtSearch los ayudó a rastrear nombres entre 260 gigabytes de datos. Y con el software Nuix obtuvieron conexiones de palabras clave incluidas en los archivos adjuntos de los correos electrónicos de varias personas, sin necesidad de abrir los documentos. También usaron el *software* libre Talend Open Studio para integrar y organizar los datos en gráficos de relaciones.

Los programadores consiguieron reconstruir el sistema de *software* de las empresas que prestaban sus servicios de creación de *offshore*. Esta crucial tarea allanó el terreno para que los periodistas iniciaran sus pesquisas, ya que pudieron navegar por archivos completamente estructurados y conocer quiénes estaban detrás de las empresas creadas, quiénes eran sus socios, intermediarios y beneficiarios.

El análisis del disco duro detectó más de 100 mil empresas extranjeras o fideicomisos localizados en lugares como las Islas Vírgenes, Hong Kong, Islas Caimán, entre otros. Los documentos revelaron la participación de 12.000 agentes intermediarios y 130.000 personas de 170 países.

En junio de 2013, el ICIJ y la Unidad de Investigación del diario La Nación de Costa Rica, dirigida por Giannina Segnini, lanzaron la aplicación Offshore Leaks Database, que permite buscar por nombres o por países.



### IMPACTO

El caso sacudió Europa y provocó dimisiones de alto nivel, como las del ministro de Economía de Francia, Jérôme Cahuzac, y el portavoz adjunto del Parlamento de Mongolia, Bayartsogt Sangajav, por ocultar cuentas bancarias en Suiza. Se abrieron investigaciones judiciales contra funcionarios y empresarios en Filipinas, India, Grecia y Corea del Sur. Distintos colectivos sociales promovieron campañas contra los paraísos fiscales. En febrero del 2015, el ICIJ fue reconocido con el premio George Polk Award, uno de los principales de EE.UU. en la categoría Business Reporting.

<http://www.icij.org/offshore>  
<http://offshoreleaks.icij.org/>

† La Comisión de la Bolsa de Valores de EE.UU. usa el programa Nuix para fiscalizar los correos electrónicos que confisca a sociedades anónimas cuando sospecha de un presunto comportamiento ilícito.



## Lux Leaks

### MEDIO

#### Consortio Internacional de Periodistas de Investigación (ICIJ)

Otro proyecto colaborativo liderado por el ICIJ. Contó con 80 reporteros de 26 países.

Fecha: Noviembre 2014

### REVELACIÓN

Más de 340 compañías, entre las que figuran Apple, JP Morgan, FedEx, Amazon y Pepsi, firmaron acuerdos fiscales secretos con Luxemburgo para evadir impuestos. Esos tratos, aprobados entre 2002 y 2010, representaron miles de millones de dólares de ingresos fiscales perdidos para los Estados donde estas empresas obtuvieron beneficios. Los acuerdos fueron firmados cuando el actual primer ministro de Luxemburgo, Jean-Claude Juncker, era ministro de Finanzas del ducado.



### ANÁLISIS DE DATOS

El ICIJ accedió a un archivo de 4.4 gigabytes que contenía 28.000 páginas de documentos. Durante seis meses, 80 periodistas se unieron en una plataforma de comunicación segura llamada Enterprise, implementada por el ICIJ para organizar el análisis del contenido. Esta herramienta permitía compartir desde transcripciones de entrevistas y fotos hasta material confidencial. “Era lo más parecido a una redacción global”, ha dicho Marina Walker, subdirectora del consorcio. Cada reportero revisó documentos financieros muy complejos, relacionados a empresas de su país. El ICIJ recibió asesoría de especialistas en finanzas y tributación. En paralelo a la investigación de los periodistas, el ICIJ formó un equipo de reporteros e ingenieros en computación que elaboró una base de datos de exploración pública.

### IMPACTO

La Comisión Europea investigó si las prácticas de Luxemburgo constituyeron un sistema fiscal adaptado a la medida de las grandes corporaciones en detrimento del derecho comunitario. El ICIJ recibió el George Polk Award en la categoría Business Reporting, y el Data Journalism Award a Mejor Investigación del 2015, junto a los Swiss Leaks.

<http://www.icij.org/project/luxembourg-leaks>

## Swiss Leaks

### MEDIO

#### ICIJ, The Guardian, CBS, Le Monde, Süddeutsche Zeitung y The Washington Post

Tercer proyecto colaborativo internacional del ICIJ, en colaboración con The Guardian, CBS, Le Monde, Süddeutsche Zeitung y The Washington Post. Incluyó a 154 periodistas de 47 países.

Fecha: Febrero 2015

### REVELACIÓN

La filial suiza del Banco HSBC estableció un sistema de evasión fiscal que ayudó a más de 100 mil clientes acaudalados de 203 países del mundo a ocultar su dinero para evitar el pago de impuestos entre los años 2005 y 2007. En la lista de beneficiarios figuraban miembros de la realeza, políticos, celebridades, traficantes de drogas y empresarios con cuentas que, en conjunto, superaron los 100 mil millones de dólares.

### ANÁLISIS DE DATOS

Se basa en registros de cuentas bancarias secretas de clientes del banco HSBC sustraídos por el informático y ex empleado de su filial suiza, Hervé Falciani. El disco duro contenía 2.5 millones de archivos. A principios de 2014, el diario francés Le Monde tuvo acceso a los datos y los pasó al ICIJ para diseñar una forma de investigarlos. El primer paso fue recrear la base de datos de clientes del HSBC a partir de los archivos planos de Excel disponibles. En seguida se hicieron relaciones entre los nombres y los países. Luego se usó el software Talend para transferir la base de datos original a la base de datos gráfica Neo4j, que permite organizar relaciones. La herramienta Linkurious facilitó la visualización y el proceso de análisis. Los reporteros participantes se comunicaban a través de la plataforma Voyager –creada con un software de código abierto llamado Oxwall–, que permite desde armar foros temáticos hasta compartir archivos. La base de datos dio pie a un diagrama de 275 mil nodos con 400 mil relaciones entre ellos.

<http://www.icij.org/project/swiss-leaks>

### IMPACTO

En febrero de 2015, la Fiscalía de Suiza abrió una investigación penal contra el banco HSBC por lavado de dinero agravado. En Reino Unido, la agencia recaudadora de impuestos recuperó 236 millones de dólares de algunos de los 3.600 británicos identificados como usuarios de la sucursal del HSBC en Ginebra, pero solo abrió proceso a una persona. Francia inició 103 acciones judiciales contra el mismo número de personas. En junio de 2015, el ICIJ recibió el Data Journalism Award en la categoría Mejor investigación del año.



## INVESTIGACIONES CON DATA PÚBLICA

Noventa países del mundo -catorce en América Latina y El Caribe- tienen leyes de acceso a la información pública, según la Alianza Regional para la Libre Expresión e Información. Este escenario representa una ventaja para el periodismo de investigación: es posible rastrear la gestión del Estado o cómo influye el poder corporativo en sus decisiones y cómo esto impacta en la vida de los ciudadanos.

### Atrapados por el Desarrollo

*Evicted and Abandoned:  
The World Bank's Broken Promise to the Poor*

#### MEDIO

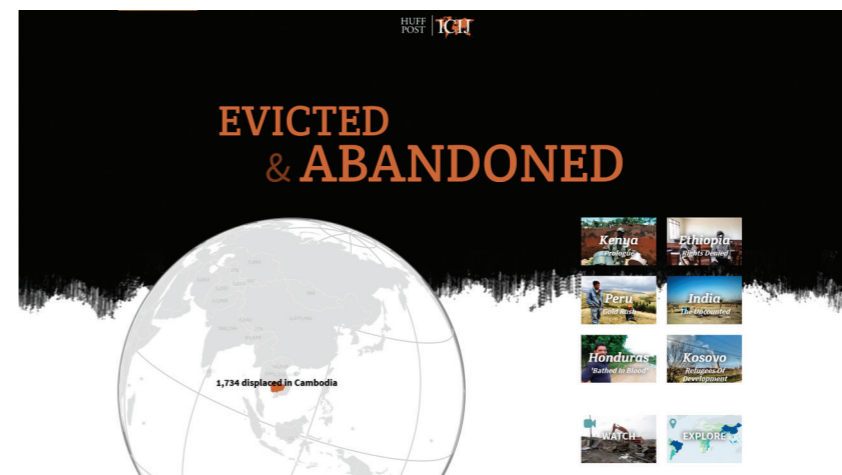
### The Huffington Post, The Guardian, The Ground Truth Project, The Investigative Fund

Trabajo en conjunto con medios y organizaciones como el Huffington Post, The Guardian, el Ground Truth Project y el Investigative Fund. Involucró a 50 periodistas de 21 países.

Fecha: Abril 2015

#### REVELACIÓN:

Más de tres millones de personas de condición vulnerable fueron desplazadas de las zonas donde vivían por cerca de mil proyectos financiados por el Banco Mundial en 124 países entre los años 2004 y 2013.



#### ANÁLISIS DE DATOS:

A comienzos de 2014, el periodista estadounidense Sasha Chavkin observó que los reportes del Defensor del Pueblo del Banco Mundial, que supervisa las actividades del organismo, contenían decenas de reclamaciones de comunidades desalojadas por proyectos financiados por el banco en varios países en desarrollo. Chavkin descargó más de 6600 documentos del BM para construir un registro de los proyectos, los beneficiarios de los préstamos y las denuncias. La información de los casos entre 2004 y 2013 estaba incompleta. Para confrontarla, se buscó fuentes dentro de la organización, exfuncionarios y expertos. Los datos permitieron identificar un patrón: el BM y la Corporación Financiera Internacional –su brazo inversor en el sector privado– no respetaron sus políticas para proteger a quienes pueden verse perjudicados por los proyectos que financia. Incluso dio préstamos a gobiernos y compañías acusados de violar derechos humanos. Los reporteros, que viajaron a países como Sudán del Sur, Etiopía, Guatemala o Perú, se comunicaban mediante la plataforma Odyssey.

<http://www.icij.org/project/world-bank>

#### IMPACTO:

El Banco Mundial anunció un plan para mejorar la supervisión de proyectos de desarrollo y evitar malas prácticas que causen desplazados. La Online News Association reconoció a esta investigación con el premio Online Journalism Award en la categoría Periodismo de investigación innovador.



# Congresistas respaldan legislación que podría beneficiarlos a sí mismos y a sus familiares

*Congress members back legislation that could benefit themselves, relatives*

## MEDIO

### The Washington Post (EE.UU.)

Fecha: Octubre 2012

## REVELACIÓN

Setenta y tres legisladores estadounidenses aprobaron normas que afectaban sus inversiones o beneficiaban a sus familiares gracias a que no estaban obligados a declarar sus potenciales conflictos de interés.

## ANÁLISIS DE DATOS

El equipo armó una base de datos en Excel con los formularios de información financiera y los registros públicos de los 535 miembros del Senado de EE.UU. Luego, contrastó las inversiones personales de los legisladores con reportes de sus actividades monitoreadas por LegiStorm, grupo de vigilancia sin fines de lucro. La información también fue cruzada con informes de la Oficina de Administración y Presupuesto de la Casa Blanca. Se detectó casos como el de un legislador que facilitó la aprobación de exoneraciones tributarias para propietarios de caballos y luego compró siete ejemplares de raza. Otro patrocinó un proyecto de ley que benefició a la compañía de gas natural en la que su esposa era accionista.



## IMPACTO

El Congreso puso a debate su reglamento de conducta ética y abrió investigaciones a los senadores con claros conflictos de intereses.

<https://goo.gl/9WbZbT>

# Señales ignoradas, consecuencias fatales

*Missed signs, fatal consequences*

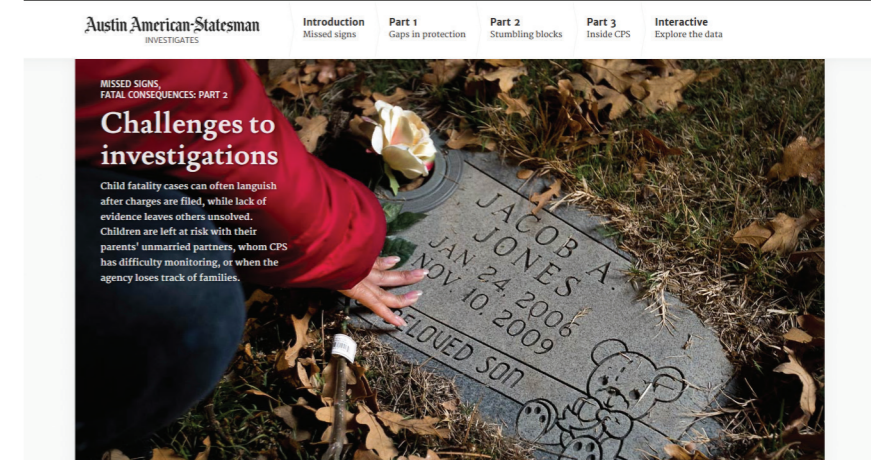
## MEDIO

### Austin American-Statesman (EE.UU.)

Fecha: Enero 2015

## REVELACIÓN

Entre 2010 y 2014, el Departamento de Familia y Servicios de Protección Infantil de Texas tuvo un sistema fallido de supervisión que permitió la muerte de 655 niños a manos de sus familiares o cuidadores. Los funcionarios no adoptaron las acciones necesarias para protegerlos.



## ANÁLISIS DE DATOS

El proyecto se inició con la obtención de 779 informes de muertes de niños por casos de violencia en sus hogares. Los documentos, solicitados por transparencia, solo estaban disponibles en PDF. Los periodistas tuvieron que transferir su contenido a un formato analizable en el que abrieron varios campos para llenar y ordenar la información. Usaron Caspio, un servicio en línea muy amigable para gestionar bases de datos, que se basa en el sistema Microsoft SQL Server. Con esta herramienta se estableció que los empleados del servicio de protección infantil habían visitado varias veces a las víctimas, pero que no tomaron en cuenta las señales de riesgo. El trabajo de seis meses incluyó la reconstrucción de los hechos con entrevistas, cotejo de documentos y visitas a los hogares afectados.

## IMPACTO

El Departamento de Familia y Servicios de Protección Infantil de Texas reformuló el sistema de investigación de abusos de menores y abrió una investigación a 50 empleados por diversos delitos.

El sitio Austin American-Statesman ganó el premio Online Journalism Awards 2015 otorgado por la Universidad de Florida en la categoría Investigative Data Journalism.

<http://projects.statesman.com/news/cps-missed-signs/>

## En casa de herrero, cuchillo de palo

### MEDIO

## La Nación (Costa Rica)

Fecha: Abril 2012

### REVELACIÓN

La mitad de ministros de la presidenta de Costa Rica, Laura Chinchilla, subvaluó sus propiedades para pagar menos impuestos entre el 2009 y 2010.



### ANÁLISIS DE DATOS

El equipo, conformado por dos reporteros y dos ingenieros en computación, organizó una base de datos en Excel con la información de las propiedades de los ministros que figuraba en los registros públicos, los valores que habían declarado en los formularios de impuestos entregados a los municipios donde se ubican sus inmuebles y los valores homogéneos de los mismos conforme a los cálculos establecidos por el Ministerio de Hacienda. Las dos primeras bases fueron descargadas por uno de los ingenieros. La tercera fue construida conforme a las operaciones aritméticas necesarias para calcular el valor de las propiedades. Al cruzarlas, detectaron a los funcionarios infractores. La investigación alcanzó a los familiares de los ministros, porque en varios casos las propiedades estaban a nombre de las cónyuges. La serie se publicó en la edición impresa de La Nación con una infografía de las revelaciones.

### IMPACTO

Se actualizaron los valores de las propiedades de todos los funcionarios públicos implicados. En el 2013, el Instituto de Prensa y Sociedad (IPYS) concedió al equipo el Premio de Periodismo de Investigación Latinoamericano por su buen manejo de la informática para potenciar el periodismo en la vigilancia de la gestión pública.

<http://goo.gl/16nd18>

# 6 INSTRUMENTOS PARA CONTAR MEJORES HISTORIAS

[Cómo enriquecer un relato con gráficos, infografías y hasta frases sonoras]



## 1. Datawrapper:

<https://datawrapper.de/>

Permite seleccionar datos de una hoja de cálculo y convertirlos en gráficos y mapas explicativos con tipos de color y fuentes personalizadas. Da opción de hacer diagramas de torta, fiebres o barras.

## 2. TimelineJS:

<http://timeline.knightlab.com/>

Se usa para crear cronologías interactivas, con fotos, videos, e hipervínculos. No se necesita abrir una cuenta. Uno inserta las fechas, textos y los URL en una hoja de cálculo de Google y la herramienta los organizará para visualizarlos de una forma muy atractiva.

## 3. Infogr.am:

<https://infogr.am/>

Desarrolla infografías en línea. Ofrece plantillas que permiten mostrar datos en barras, círculos, fiebres. Uno puede insertar la información en su formato o importar archivos en Excel o Csv. Tiene una versión gratuita y una Premium de paga. Es posible compartir los gráficos a través de Facebook, Twitter y Pinterest.

## 4. Tableau Public:

<https://public.tableau.com/>

Convierte los datos de una hoja de cálculo en gráficos interactivos (mapas, tablas, barras) y crea filtros para que los usuarios hagan consultas y tengan resultados personalizados. No se necesita saber programación para usar esta herramienta.

## 5. StoryMapJS:

<http://storymap.knightlab.com/>

Crea historias basadas en lugares identificados en un mapa. Permite insertar videos, tuits, textos o imágenes para mostrarlos en forma de una galería asociada a cada sitio seleccionado. También se inserta la información en una hoja de cálculo de Google.

## 6. Soundcite:

<https://soundcite.knightlab.com/>

Coloca sonidos a una palabra o una frase de un texto. Se sube el archivo MP3 a SoundCloud, luego hay que seleccionarlo desde Soundcite y se ajusta el fragmento deseado. Se generará un código para insertarlo en la frase escogida, lo que enriquece la experiencia de leer con sonido.



## INVESTIGACIONES CON BASES CONSTRUIDAS

En los últimos años, grandes casos nacieron de proyectos de investigación con bases de datos que los propios periodistas se encargaron de construir. Aquí reunimos trece experiencias que pueden inspirar el desarrollo de nuevos temas.

### ¿Quién está detrás del derrumbe financiero?

*Who's Behind the Financial Meltdown?*

#### MEDIO U ORGANIZACIÓN

**The Center for Public Integrity (EE.UU.)**

Fecha: Mayo 2009

#### REVELACIÓN

Las corporaciones estadounidenses Lehman Brothers, Merrill Lynch, JP Morgan & Co., Citigroup, Goldman Sachs & Co. y el banco suizo Credit Suisse First Boston fueron parte del negocio que generó la denominada burbuja inmobiliaria que quebró el sistema financiero. Esas corporaciones eran propietarias de 21 de las 25 principales empresas de la industria *subprime* que otorgó créditos hipotecarios de vivienda de alto riesgo, lo que desataron la crisis económica de 2008, y luego se beneficiaron del rescate financiero.



#### ANÁLISIS DE DATOS

El editor de datos del CPI, David Donald, comenzó un análisis de 350 millones de solicitudes de hipotecas aprobadas desde 1994 hasta el 2007. La información había sido previamente recogida de los documentos públicos del Home Mortgage Disclosure Act (el sistema de registros de préstamos basado en la Ley de Divulgación de Hipotecas) y organizada en hojas de cálculo por el Instituto Nacional de Periodismo Asistido por Computadora (NICAR, por sus siglas en inglés). El análisis permitió identificar que la mayoría de préstamos de alto riesgo por más de un billón de dólares fue otorgada entre el 2005 y el 2007. Esto permitió establecer quiénes fueron los principales adjudicatarios. Un equipo de reporteros colaboró con perfiles de los prestamistas y se incluyó información sobre las contribuciones de las empresas involucradas a los miembros del Congreso de EE.UU. Para visualizar la ubicación de cada vivienda sujeta a préstamo *subprime* se empleó mapas de calor hechos con el Palantir Government Software, que ofrece una potente herramienta de análisis visual usada tanto en proyectos académicos como en gestión de desastres o programas de inteligencia.

<http://goo.gl/cVX1ee>

#### IMPACTO:

La serie periodística contribuyó a que el Congreso de EE.UU. formara una comisión investigadora, cuyas conclusiones fueron que el desastre pudo evitarse y que la crisis fue resultado de fallos en las regulaciones, mala gestión empresarial y el riesgo irresponsable de Wall Street.

## Los diarios secretos de Paraná

Diários Secretos

### MEDIO

## Gazeta do Povo en colaboración con RPCTV Paraná (Brasil)

Fecha: Marzo 2010

### REVELACIÓN

Entre el 2006 y 2009, la Asamblea Legislativa de Paraná (Brasil) ocultó un esquema de desvío de fondos públicos que incluía la contratación de empleados fantasmas, sobrecostos de servicios, casos de nepotismo y otros delitos. La malversación habría llegado a los 400 millones de dólares.



### ANÁLISIS DE DATOS

Los periodistas James Alberti, Katia Brembatti, Karlos Kohlbach y Gabriel Tatcheik recopilaron las gacetas legales de la Asamblea Legislativa de Paraná publicadas entre 1998 y 2009, varias de las cuales no estaban disponibles en los archivos de la propia asamblea. Digitalizaron a mano el contenido de 724 boletines legales para armar una base de datos en Excel que mostrara las contrataciones del personal y el manejo de los presupuestos señalados en las gacetas. Así pudieron identificar una veintena de empleados fantasmas, incluso personas muertas o niños. El equipo periodístico también detectó que se había empleado a parientes de los legisladores e hijos de magistrados. El análisis de datos se completó con testimonios y fuentes documentales. Luego compartieron en un buscador público todas las gacetas oficiales, incluidas las que faltaban en el archivo de la Asamblea.

### IMPACTO

Desde la publicación hasta marzo de 2015, catorce funcionarios y exempleados de la Asamblea Legislativa de Paraná fueron condenados a prisión por malversación de fondos y otros delitos. “Los diarios Secretos de Paraná” ganó el Global Shining Light Award y el Premio Latinoamericano de Periodismo de Investigación en el 2011.

<http://goo.gl/OqXwNF>

## Los rostros del desamparo

### MEDIO

## La Nación (Costa Rica)

Fecha: Febrero 2011

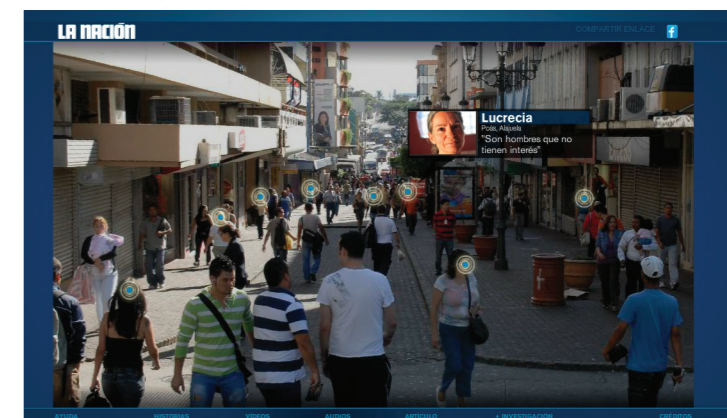
### REVELACIÓN

El programa oficial de subsidios educativos Avancemos cubre los estudios de jóvenes abandonados por padres con elevados recursos económicos.

<http://goo.gl/BfMkQC>

### ANÁLISIS DE DATOS

Esta serie investigativa se inició con la hipótesis de que podrían existir abusos en la concesión del beneficio de las becas de Avancemos, un programa de subsidios para estimular que más de 167 mil jóvenes continúen sus estudios. El equipo, liderado por Giannina Segnini, accedió a la base de datos de los beneficiarios, la completó con los nombres de sus padres, y la cruzó con los ingresos y bienes familiares. Un primer cruce reveló que 75 becarios tenían padres con salarios de entre dos mil y nueve mil dólares. Sin embargo, el hallazgo dio un giro cuando los reporteros buscaron en persona a los beneficiarios: resultó que eran hijos de personas con recursos económicos, pero que habían sido abandonados y ahora vivían en la pobreza con algún pariente. La investigación duró tres meses y derivó en nueva historia más reveladora que la hipótesis inicial: el Estado estaba subsidiando la educación de jóvenes desamparados por padres con patrimonio. En este proyecto trabajaron tres desarrolladores, tres diseñadores y 4 periodistas. Fue acompañado con un especial multimedia.



### IMPACTO

La investigación recibió una mención especial en el II Premio Regional de Periodismo, Pobreza y Derechos Humanos en Centroamérica en 2011. El Instituto Mixto de Ayuda Social (IMAS) cambió el criterio para otorgar las becas del programa Avancemos a selección por nivel académico.

# Médicos sin control: los dueños de la salud pública en Chile

MEDIO

## CIPER (Centro de Investigación Periodística)

Fecha: Septiembre 2010

### REVELACIÓN

Un seguimiento de un año a cinco hospitales de Chile demostró que no fiscalizaban la asistencia de los médicos ni el uso que hacen de la infraestructura de servicios de salud en beneficio de sus intereses personales y en perjuicio de los pacientes.



### ANÁLISIS DE DATOS

A fines de julio de 2009, CIPER solicitó los registros de asistencia de los médicos de cinco hospitales de Santiago de Chile. Para obtenerlos tuvo que superar varias trabas, porque algunos hospitales se negaron a entregar la información y otros la entregaron incompleta. El equipo logró acceder a más de 35 mil registros de asistencia. Después de pasar la información a una hoja de cálculo, la cruzó con los horarios que los profesionales deben cumplir en los policlínicos y en sus consultas particulares o en clínicas privadas. En forma paralela, visitó los hospitales seleccionados para observar en terreno el cumplimiento de los horarios. Así comprobó que varios médicos usaban la infraestructura para atender a sus pacientes privados. También se evidenció que no cumplían con las jornadas estipuladas en sus contratos.

### IMPACTO

Siete meses después de publicada esta investigación, la Contraloría General de la República de Chile publicó un informe que confirmó las infracciones de los médicos en 13 hospitales del país. La serie investigativa obtuvo una mención honrosa en el Premio Latinoamericano de Periodismo de Investigación entregado el 2011.

<http://ciperchile.cl/multimedia/medicos-son-control/>

# Dólares para los doctores

*Dollars for docs*

MEDIO

## ProPublica en asociación con The Boston Globe, Consumer Reports, NPR, Chicago Tribune y Public Broadcasting Service (PBS).

Fecha: Octubre 2010

### REVELACIÓN

Entre los años 2009 y 2010, siete compañías farmacéuticas hicieron pagos individuales de más de cien mil dólares a 17.000 médicos para que promovieran y prescribieran sus medicamentos en EE.UU.

### ANÁLISIS DE DATOS

Las siete farmacéuticas habían publicado esta información en sus sitios web, pero estaba en formatos difíciles de analizar (PDF y JPG). Dos periodistas y un desarrollador la descargaron y organizaron en un Excel que se podía desglosar en categorías como: consultorías, comidas, viajes y regalos. Por primera vez se tuvo el panorama de los pagos a médicos realizados por compañías que representaban del 36% de la industria farmacéutica en EE.UU. El equipo utilizó la herramienta Open Refine para limpiar y uniformizar los nombres de los médicos beneficiados. Luego los cruzó con las bases de datos públicas de licencias para ejercer la medicina y de expedientes disciplinarios. ProPublica presentó esta investigación con una aplicación periodística en la que cualquier persona puede buscar el nombre de su médico para conocer si recibió pagos para promover determinado producto. Tres años después, la base de datos fue actualizada hasta evidenciar pagos por cuatro billones de dólares a 681.432 médicos por parte de 1.630 farmacéuticas o fabricantes de productos médicos.

### IMPACTO

En el 2012 se promulgó una ley que convierte en obligatoria la publicación de los regalos y pagos por marketing a los médicos en EE.UU. Más de 125 organizaciones de noticias –como el Boston Globe o el Chicago Tribune– hicieron investigaciones basadas en la herramienta.

<https://projects.propublica.org/docdollars/>



# Terroristas para el FBI

Terrorists for the FBI

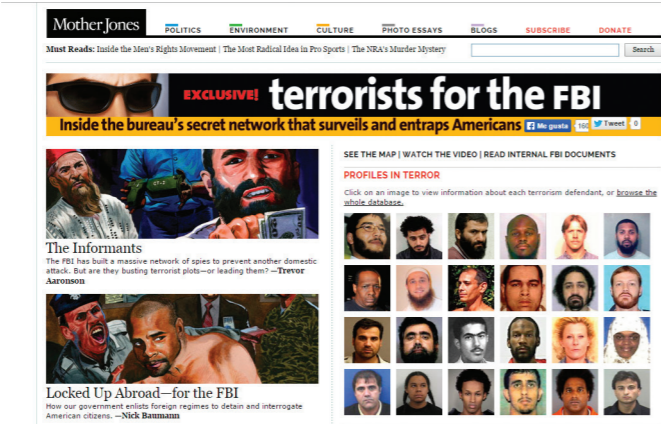
MEDIO

Mother Jones (EE.UU.)

Fecha: Agosto 2011

## REVELACIÓN

Informantes del FBI fabricaron pruebas para incriminar a sospechosos de atentar contra la seguridad de EE.UU. con el fin de reclamar recompensas de 100.000 dólares como parte de la cruzada contra el terrorismo.



## ANÁLISIS DE DATOS:

El periodista Trevor Aaronson examinó más de 500 casos de personas acusadas de terrorismo y encontró que casi la mitad había tenido como eje la participación de un informante del FBI. Aaronson combinó data extraída de archivos judiciales de distintos estados, documentos del FBI, entrevistas a agentes y los abogados de los acusados. Pasó meses trabajando con un asistente en construir una base de datos. En un primer momento utilizó Excel y el gestor MySQL, que ayuda a construir bases de datos relacionales. Luego el equipo utilizó el lenguaje Drupal para construir una herramienta de búsqueda en línea.

## IMPACTO:

FBI investigó a los agentes acusados de fabricar casos, mientras que parte de los informantes involucrados ha sido objeto de procesos judiciales. La investigación obtuvo el premio Data Journalism Award 2012 en la categoría Data-Driven Investigation.

<http://goo.gl/hKem0p>

# Metadona y la política del dolor

Methadone and the politics of pain

MEDIO

The Seattle Times (EE.UU.)

Fecha: Diciembre 2011

## REVELACIÓN:

El programa Medicaid, dirigido a personas de bajos recursos, entregó el narcótico metadona a sus pacientes para abaratar costos en sus compras de medicamentos sin tomar en cuenta los daños a la salud producidos por este analgésico. Como consecuencia, 2.173 personas fallecieron entre los años 2003 y 2011.



## ANÁLISIS DE DATOS:

Ante la alerta de varios médicos, dos periodistas buscaron todos los registros disponibles en Washington para rastrear el número y circunstancias de los fallecimientos vinculados a este narcótico. A través de solicitudes de información, obtuvieron cuatro bases de datos: los registros de los certificados de defunción, las notas forenses de cada médico, los perfiles clínicos de los pacientes y los costos de sus tratamientos con metadona en los hospitales de Washington. En paralelo, recopilaron más datos sobre el perfil socioeconómico de los fallecidos. Los archivos fueron revisados y anotados con Document-Cloud. Esta información fue pasada a Excel, luego se usó el software Access, que permite manejar mayor volumen de información, y el software ArcGIS, para hacer mapas. Para presentar los hallazgos se usó Google Fusion Tables, Tableau, Final Cut Pro y Adobe.

## IMPACTO:

Las autoridades de Washington enviaron una alerta a más de 1.000 farmacéuticos y 17.000 profesionales de la salud sobre los riesgos de la metadona. En enero de 2012, patrocinaron un programa que instruyó a médicos para limitar la metadona a casos de última alternativa terapéutica. La investigación recibió el Premio Pulitzer 2012 en la categoría Periodismo de investigación. Ese mismo año fue premiada con el Data Journalism Award.

<http://goo.gl/YoQcDs>

## Gastos del Senado

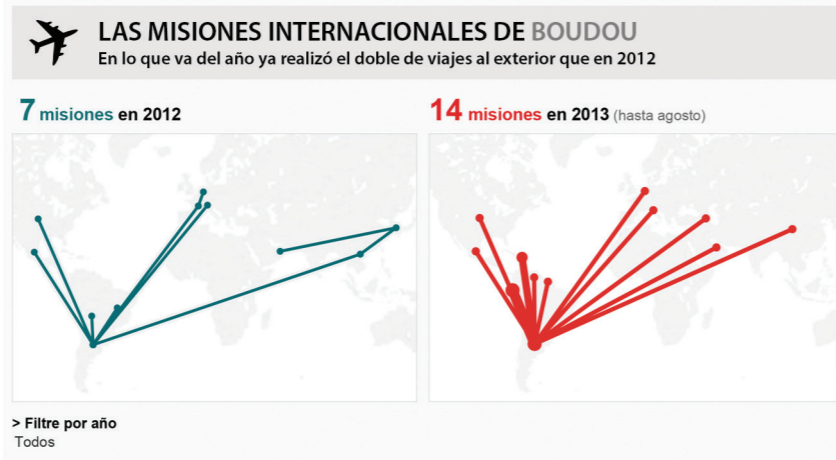
MEDIO

### La Nación (Argentina)

Fecha: Febrero 2013

#### REVELACIÓN:

El equipo descubrió contrataciones, viajes y otros gastos irregulares por más de un millón de dólares del Senado argentino entre 2010 y 2012. El vicepresidente del senado, Amado Boudou, compró muebles de lujo para su oficina e hizo rendiciones de viáticos de viajes que se superponían.



#### ANÁLISIS DE DATOS:

Un informante anónimo envió a La Nación un correo electrónico que contenía una fotografía del despacho del vicepresidente del Senado argentino, Amado Boudou, con una lujosa mesa importada desde Italia. Este hecho puso a trabajar al equipo periodístico en la descarga de 33.000 documentos de las contrataciones del senado que estaban en su web oficial. Se usó el software Omnipage 18 para convertir los PDF en archivos rastreables. La información pasó a un Excel que incluía los gastos en muebles, pagos de viajes, y gastos de personal de seguridad, entre otros. Se utilizó Tableau Public para explorar la data y hacer gráficos interactivos. Para manejar parte de los PDF se desarrolló la plataforma colaborativa VozData que permitía la participación de voluntarios para clasificar la información en un formato preestablecido.

#### IMPACTO:

La Fiscalía inició una investigación de los viajes del vicepresidente del Senado. El tema fue tratado en televisión, radios y otros periódicos. El trabajo obtuvo el Data Journalism Award 2013 en la categoría Data-Driven Investigation (Big Media), organizado por el Global Editors Network (GEN).

<http://www.lanacion.com.ar/gastos-en-el-senado-t49163>

## EL PODEROSO NEO4J

[O cómo descubrir un fraude global con círculos y líneas]

Mar Cabra, editora de la Unidad de Investigación y Datos del ICIJ, aprendió a usar desde hojas de cálculo hasta sofisticados software para seguir el rastro de la corrupción fiscal y corporativa en millones de archivos que, en apariencia, no tenían conexión directa entre sí. Para la serie Swiss Leaks usó una herramienta llamada Neo4j, que permite **identificar conexiones entre grandes cantidades de datos y mostrarlas en gráficos**. “Las conexiones eran cruciales para identificar quién hacía negocios con quién”, contó Cabra durante la Conferencia Global de Periodismo de Investigación realizada en octubre de 2015 en Lillehammer, Noruega. **En lugar de tablas, esta herramienta usa nodos y aristas, lo que hace más intuitiva la lectura de las relaciones entre datos.**

Esta función permitió al equipo de **OjoPúblico** desentrañar el rastro del financiamiento de los partidos en la campaña electoral peruana del 2016, a partir del cruce de 16 bases de datos y el análisis de 3 millones de registros. “Este sistema nos permitió comprender, analizar y cruzar simultáneamente las bases de datos”, explica Nelly Luna, la periodista a cargo de la investigación. **Lo que antes hubiera tomado varios años, con la herramienta apenas tomó seis meses.**

## Casas embargadas

Homes for the Taking. Liens, losses and profiteers

### MEDIO

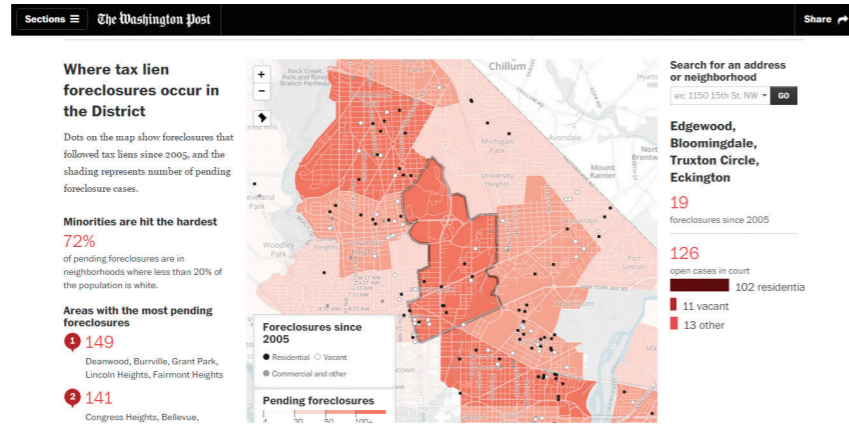
## The Washington Post (EE.UU.)

Fecha: Septiembre 2013

### REVELACIÓN

Unas 200 personas de Washington DC., en su mayoría adultos mayores, perdieron sus viviendas en ejecuciones hipotecarias irregulares por deudas tributarias de menos de mil dólares. Estos casos evidenciaron abusos en el programa de retención de impuestos del distrito que permitieron a compañías inmobiliarias hacerse de las propiedades.

<http://goo.gl/ZGtrUP>



### ANÁLISIS DE DATOS

La investigación se basó en el análisis de los documentos de ejecuciones hipotecarias entre los años 2005 y 2013 disponibles de la Oficina de Impuestos y Rentas, la Corte Superior y la Oficina del Censo de Estados Unidos. El equipo armó una base de datos de 200 propietarios adultos mayores que perdieron sus casas en subastas que hizo el programa de retención de impuestos del distrito por el atraso de pagos de menos de mil dólares. En la mayoría de los casos, las propiedades fueron vendidas a compañías inmobiliarias pese a que los dueños originales finalmente pagaron sus deudas. Cuando los reporteros buscaron a las víctimas obtuvieron más indicios de los abusos cometidos. Uno de los afectados agonizaba de cáncer y debía 1.025 dólares en impuestos. Otro tenía 95 años de edad y padecía Alzheimer, por lo que olvidó cancelar un saldo de apenas 44 dólares. Los documentos fueron examinados con DocumentCloud y los datos fueron analizados con hojas de cálculo. Los hallazgos fueron trabajados con las plataformas Mapbox y la librería de código abierto Leafletj, que permite hacer mapas amigables para móviles.

### IMPACTO

Una docena de senadores solicitó al gobierno investigar los programas de impuestos con el fin de proteger a los propietarios vulnerables de perder sus inmuebles por pequeñas deudas tributarias. La investigación ganó el premio Data Journalism Award 2014 en la categoría Data-Driven Investigation.

## Jadeantes y agobiados

Breathless and Burdened

### MEDIO

## Centro para la Integridad Pública (EE.UU.) y Unidad de Investigación de ABC News

Fecha: Octubre 2013

### REVELACIÓN

Médicos y abogados al servicio de la industria del carbón han bloqueado las ayudas a los mineros afectados por la enfermedad del pulmón negro.

### ANÁLISIS DE DATOS

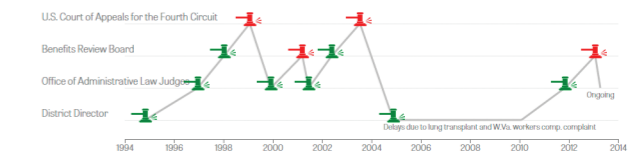
El reportero Chris Hamby se propuso construir una base de datos con información de los exámenes de rayos X practicados a más de 1.500 mineros en el Hospital Johns Hopkins desde el año 2000 hasta el 2013. Las mineras de carbón les habían negado el tratamiento de salud y pagos de beneficios sociales en base a diagnósticos firmados por un solo médico, quien descartó que tuvieran la enfermedad del pulmón negro. Los mineros judicializaron sus casos. El periodista accedió a esos expedientes para construir una segunda base de datos sobre las estrategias legales de las mineras, los reportes de otros médicos, y los veredictos de los jueces. Hamby pasó meses leyendo y procesando la información en hojas de cálculo. De esta manera pudo identificar casos y patrones que evidenciaban un sistema montado para negar beneficios a los mineros.

Breathless and Burdened

### A 19-year fight for benefits

By Chris Hamby, Chris Zubak-Skees 6:00 am, November 1, 2013 Updated: 12:19 pm, May 19, 2014

Former miner Ted Lutsek has tried for almost two decades to prove that the scarring in his lungs was caused by coal mine dust. Doctors testifying for the company have denied any link between his particular pattern of disease and his work, despite increasing recognition of this form of illness by government agencies and independent researchers. Click on a gavel to read the decision.



### IMPACTO

El Hospital Johns Hopkins suspendió y luego separó al médico responsable del programa de detección de pulmón negro. Por otro lado, senadores de EE.UU. elaboraron proyectos de ley para reformar el sistema de beneficios de los mineros de la industria del carbón. La investigación ganó el premio Pulitzer del 2014 en la categoría de Periodismo de Investigación.

<http://goo.gl/ZPZ9HE>



## CARTOGRAFÍA DIGITAL

[Recursos para ubicar los hechos y personajes en el lugar exacto]

### OpenStreetMap

<http://www.openstreetmap.org/>

Proyecto colaborativo para crear mapas libres y editables. Se pueden usar datos geográficos capturados con dispositivos GPS móviles y otras fuentes.

### My Maps

<https://www.google.com/maps/d/u/0/>

Herramienta para crear mapas a través del Google Maps. Es fácil de usar, compartir e insertar gracias al código que proporciona. El único requisito es tener una cuenta de Google.

### MapBox

<http://mapbox.com/tour/>

Sitio web que permite crear mapas personalizados de manera sencilla. Usa software libre.

### InfoAmazonia

<http://infoamazonia.org/>

Sitio web que cuenta con mapas ambientales de los nueve países de la región amazónica.

### Geocommons

<http://geocommons.com/>

Programa gratuito para crear mapas con múltiples capas. Permite utilizar la información geolocalizada de otros usuarios y compartirla.

## Medicare Desenmascarado

Medicare Unmasked

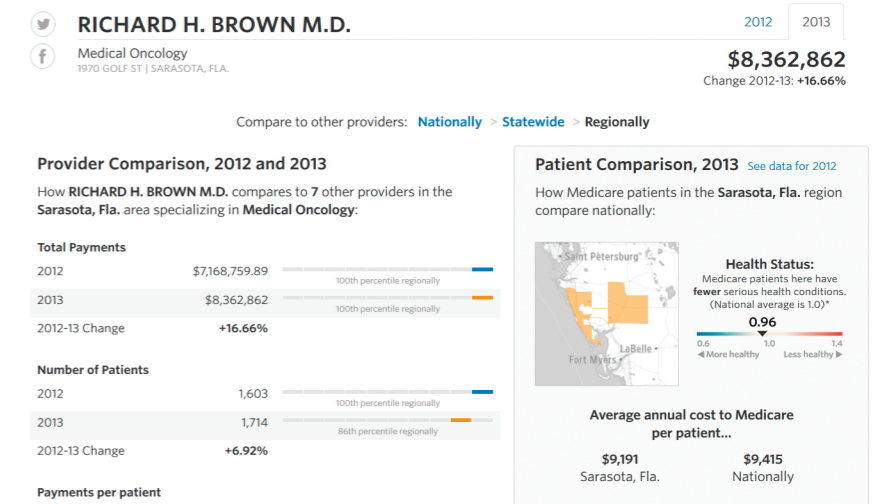
### MEDIO

## The Wall Street Journal (EE.UU.)

Fecha: Abril 2014

### REVELACIÓN:

Medicare, el programa de ayuda sanitaria a personas con más de 65 años y jóvenes con graves enfermedades, realizó pagos por 60 billones de dólares anuales a más de 880.000 médicos, servicios de ambulancia y laboratorios, varios de los cuales fueron casos de fraude, derroche y abusos.



### ANÁLISIS DE DATOS:

En mayo de 2014, tras una batalla de cinco años en tribunales, el WSJ consiguió que el Gobierno de EE.UU. liberara casi 10 millones de registros de contratos y desembolsos de Medicare a sus proveedores de servicios médicos, mantenidos en secreto desde 1979. Luego accedieron a una segunda base de datos comprada a la agencia CMS (Centers for Medicare and Medicaid Services), que incluía registros de reclamos de pago de proveedores en un periodo de 6 años. El cruce de información permitió construir una base de datos que sirvió para detectar casos de fraude, sobrecostos y abusos en un programa que gasta más de 60 billones de dólares anuales. Los periodistas y expertos en datos utilizaron el lenguaje de programación C# para convertir los registros en tablas relacionales y desarrollaron algoritmos para hacer conexiones con los datos. Luego importaron la información al gestor Microsoft SQL. Con esa data produjeron gráficos interactivos, rankings y una plataforma de búsqueda de los pagos a cada médico.

### IMPACTO:

Ganador del Pulitzer 2015 en la categoría Periodismo de Investigación.

<http://graphics.wsj.com/medicare-billing/>

# Aresep encarece gasolina y diésel para abaratar asfalto y gas

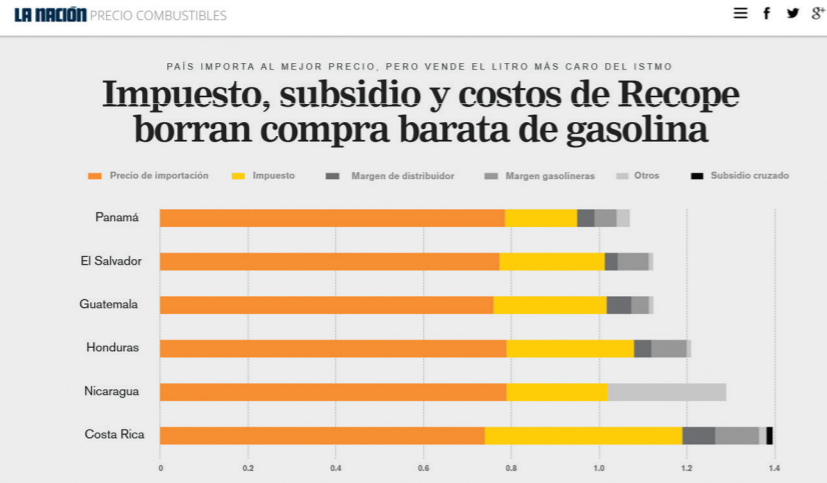
## MEDIO

### La Nación (Costa Rica)

Fecha: Diciembre 2014

## REVELACIÓN

En el 2008, la Autoridad Reguladora de los Servicios Públicos de Costa Rica (Aresep) aprobó en secreto una fórmula de cálculo precios que encarecía el diésel y la gasolina para abaratar las tarifas del gas y el asfalto. Este subsidio cruzado beneficiaba a las empresas cementeras en perjuicio de miles de usuarios.



## ANÁLISIS DE DATOS:

Las periodistas Hassel Fallas y Mercedes Agüero crearon en forma manual una base de datos en Excel con información extraída de 59 resoluciones de precios, ordinarias y extraordinarias, emitidas por la Aresep entre junio de 2009 y septiembre de 2014. Los componentes de la fórmula de precios se desagregaron para averiguar cómo se incluían los costos operativos de la Refinadora Costarricense de Petróleo (Recope) en la tarifa de los combustibles. Con ayuda de especialistas, se detectó que la Aresep había establecido un subsidio cruzado: asignaba a un producto un costo superior al real con el fin de reducir el precio de otro. En este caso, se aumentaba el costo del diésel y la gasolina para abaratar el del asfalto y el gas. El equipo tuvo que construir cinco versiones de la bases de datos hasta dar con la correcta.

## IMPACTO

La Aresep modificó la fórmula para calcular el precio de los combustibles y así eliminar los costos ocultos en la metodología que empleaba desde el 2008. La investigación fue finalista en los Data Journalism Awards 2015, de la GEN, en la categoría Investigación del año.

<http://goo.gl/aNriTd>

# La conquista del favor

*Courting Favor*

## MEDIO

### The New York Times (EE.UU.)

Fecha: Octubre 2014

## REVELACIÓN

Más de 20 corporaciones, entre las que figuran DirecTV, Pfizer, Coca Cola, Google y Citigroup hicieron regalos y contribuciones a fiscales generales en 12 estados para influenciar en sus decisiones.

## ANÁLISIS DE DATOS

Durante una investigación de nueve meses, el periodista de The New York Times Eric Lipton comprobó cómo ha penetrado el lobby corporativo en los fiscales generales para torcer sus decisiones en favor de los intereses de 21 compañías.

Lipton consiguió, a través de peticiones legales, 8.000 páginas de correos electrónicos de las cuentas públicas de los fiscales e hizo una base de datos en una hoja de cálculo con información extraída de estas correspondencias en las que se evidenciaban las relaciones de los funcionarios y las empresas. Luego, documentó los regalos y contribuciones que reciben los fiscales de las corporaciones que investigan. Esa información fue complementada con fotografías y viajes que el reportero hizo a conferencias académicas de fiscales patrocinadas por dichas empresas.

Para completar su informe, Lipton mostró que las contribuciones de las compañías a las asociaciones de fiscales demócratas y republicanos se cuadruplicaron en cuatro años.

## IMPACTO

Las publicaciones generaron investigaciones en cuatro estados y el Senado planteó un proyecto de ley que prohíbe que los funcionarios reciban regalos o contribuciones económicas.

*Courting Favor* fue galardonado con el Pulitzer 2015 en la categoría de Periodismo de Investigación y con el IRE Award en la categoría Print/Online-Large.

<http://goo.gl/oKMMKg>





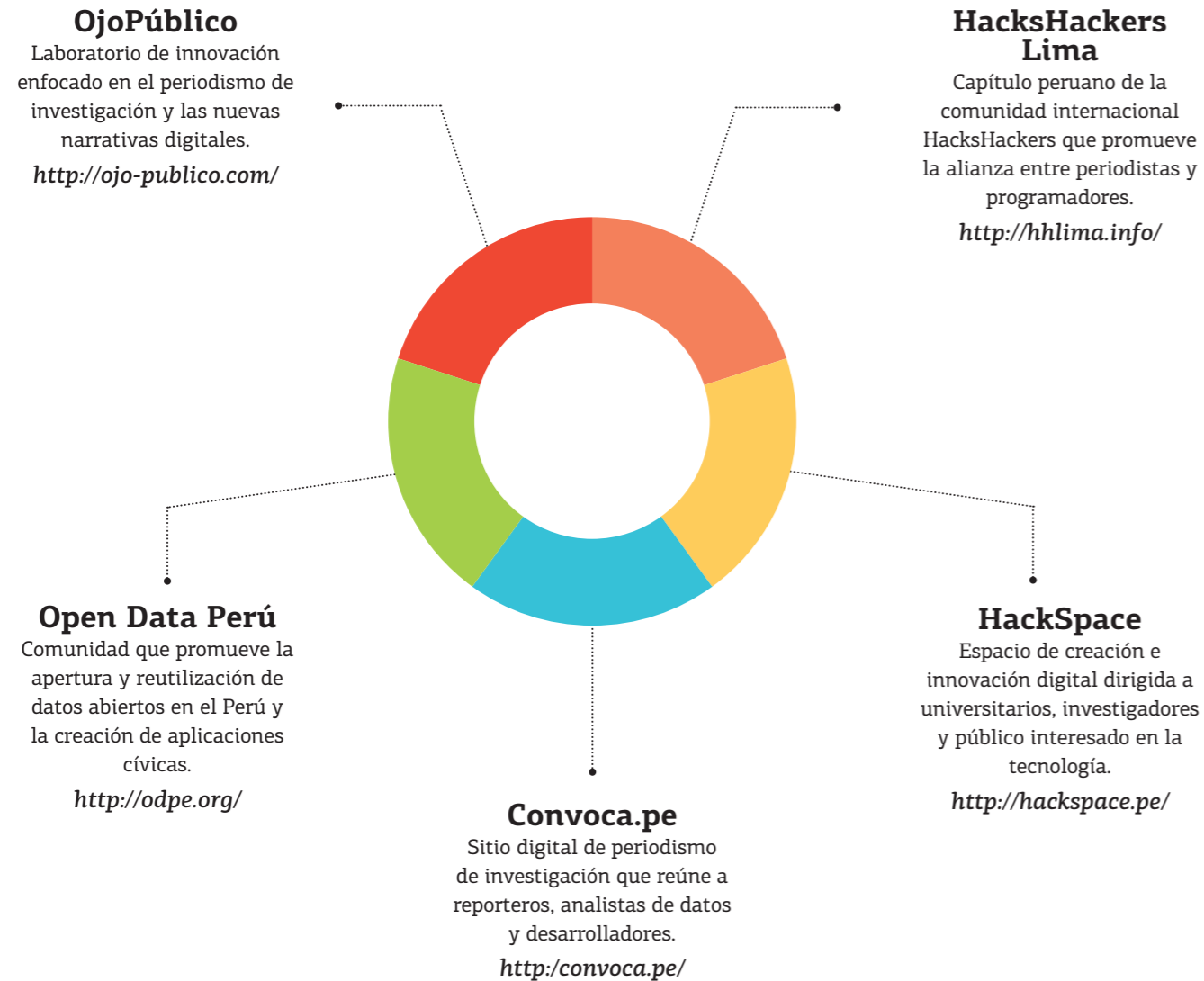
# EL CAMINO HACIA UNA CULTURA DE INNOVACIÓN

Los laboratorios digitales  
del periodismo de investigación  
en el Perú

3



## La comunidad de los datos en el Perú



**A**ntes de que el periodismo se amistara con la estadística, la periodista Liz Mineo detectó una forma de rastrear la corrupción con tablas numéricas. A fines de 1997, mientras el Perú era sacudido por frecuentes denuncias sobre malos manejos del gobierno de turno, Mineo puso atención en las obras públicas de prevención ante el Fenómeno del Niño, un proceso cíclico de alteraciones climáticas que suele desencadenar desastres en distintos puntos del país. La reportera se planteó una pregunta básica: ¿Cómo se había gastado el publicitado presupuesto público de cien millones de soles para esos trabajos? El problema era que el régimen de entonces no se caracterizaba por su transparencia. Tampoco existía la ley de acceso a la información. De hecho, el Instituto Nacional de Defensa Civil (Indeci), el organismo encargado de administrar el dinero y hacer las contrataciones, estaba controlado por militares ligados al temido asesor presidencial, Vladimiro Montesinos, quien había generado un ambiente casi policíaco en el Estado. Al hacer sus primeras averiguaciones, Mineo encontró que habían declarado en reserva toda información sobre obras públicas. Entonces se embarcó en una experiencia pionera de análisis forense de la información con fines periodísticos: construyó una base de datos para buscar patrones que permitieran descubrir la información oculta.

En una primera etapa, la reportera de El Comercio buscó fuentes internas. Tras un persistente trabajo de convencimiento, un funcionario del Indeci accedió a colaborar: le entregó un file de 80 páginas con información de las obras y los contratistas. Era el mejor insumo que podía obtener. “Me dio miedo que se desanimara si le pedía los

documentos en un disquete”, recuerda Mineo. El segundo paso fue trasladar toda la información a una tabla de Excel, que incluyó 293 obras en 21 departamentos del Perú, por un monto de 100 millones de soles, a cargo de 61 empresas. “Eran demasiados números para hacer cálculos a mano o con calculadora”, recuerda.<sup>1</sup>

El resultado fue un compendio de irregularidades: el primer filtro permitió detectar que muchas de las supuestas obras de prevención se concentraban en departamentos que no pertenecían al área vulnerable frente al Fenómeno del Niño. Cuando averiguó quiénes eran los accionistas de las empresas beneficiadas con los contratos, Mineo identificó que 12 de las 61 compañías eran de propiedad de compañeros de armas del jefe del Indeci, el general Homero Nureña. Además, una de las empresas estaba a nombre de su secretaria privada y otra, de su sobrino. Un tercio eran empresas creadas apenas meses antes de recibir los contratos y algunas incluso después de haberlos obtenido.

La tercera etapa fue la verificación. Mineo, con el apoyo de los corresponsales del diario, visitó los lugares de las inversiones y descubrió obras inconclusas e incluso algunas que nunca existieron. Así comprobó que el general Nureña había beneficiado a Cajamarca, su ciudad natal, pese a que ese departamento no sufría los embates de El Niño, y que mandó levantar allí una escuela primaria que luego bautizó con el nombre de su madre. De no haber realizado en persona ese recorrido, quizás los hallazgos no habrían sido tan contundentes.

La experiencia marcó un cambio en las capacidades operacionales de la periodista. En una época en que el Gobierno se reservaba la facultad de manejar la información pública, Liz Mineo combinó las técnicas periodísticas tradicionales con el uso incipiente de recursos propios de la mentalidad de datos para realizar una contundente investigación. El caso se convirtió en una serie de 15 artículos en



## 8

medios latinoamericanos tienen equipos periodísticos integrados por hackers para hacer análisis masivo de datos. En el Perú, **OjoPúblico** y **Convoca** trabajan con programadores.

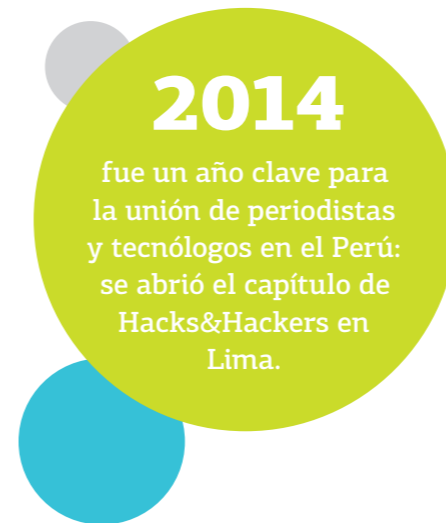
.....

El Comercio. Tiempo después, el general Nureña fue condenado a prisión por malversación de fondos.

Por mucho tiempo, esta experiencia pionera permaneció como un caso aislado en el panorama periodístico peruano. Replicarla era poco probable en un país con tan poco apego al registro de la realidad que había quien ponía en duda hasta la fecha de cumpleaños del presidente de la República.<sup>2</sup> Al menos hasta fines del siglo XX, los archivos públicos peruanos adolecían de amnesia y las agencias del gobierno administraban sus presupuestos con algo más parecido a la nigromancia que a la contabilidad. Los reporteros de investigación estuvieron concentrados en rastrear el narcotráfico, el terrorismo o la corrupción. Tenían que descubrir esas redes a partir de fuentes directas.

Solo a partir del 2001, con la publicación de la Ley de Transparencia y Acceso a la Información Pública, las diferentes entidades del Estado empezaron a digitalizar sus datos y a publicarlos en sus plataformas en Internet. Aun así, la informalidad de los mecanismos de recojo y actualización de información generaba datos poco fiables. Una muestra clamorosa fue detectada en el 2008 por los periodistas Gustavo Gorriti y Romina Mella, de IDL Reporteros (IDL-R), mientras indagaban en los registros de delitos de las comisarías en Lima Metropolitana.<sup>3</sup> Cuando solicitaron información sobre el número y tipo de delitos en las distintas jurisdicciones, advirtieron que casi todas presentaban resultados muy similares. Al indagar un poco más, descubrieron que los reportes eran llenados con un simple copy/paste, bajo la premisa de que en todos lados la situación era parecida. Los reporteros tuvieron que buscar otra manera de aproximarse al problema de criminalidad que investigaban.

Hasta la primera década del siglo XXI, el uso de bases de datos en el periodismo peruano se debió a esfuerzos individuales de reporteros de investigación para seguir adelante en sus pesquisas. En



1 Entrevista personal a Liz Mineo.

2 VALENZUELA, Cecilia. “Buscando la cuna de Fujimori”. En revista Caretas [Lima]. Ver: <http://www.caretas.com.pe/1475/fujimori/fujimori.htm>. [Visualizado: 25 de noviembre de 2015].

3 Tomado de la exposición de Romina Mella en inauguración de Chicas Poderosas Perú, 20/11/15.

el 2010, la periodista Milagros Salazar Herrera, de IDL-Reporteros, investigó la poderosa industria pesquera del Perú con herramientas digitales que le permitieron compilar, verificar y analizar en hojas de cálculo más de 47 mil reportes de desembarque de anchoveta, una especie en riesgo permanente de sobreexplotación. Los documentos correspondían a los volúmenes de pesca declarados por las empresas y, por otro lado, a los informes de peso registrados por los supervisores del Estado en los terminales del país, entre los años 2009 y 2010. El contraste al detalle permitió detectar un sistema fallido de auditoría que beneficiaba a la segunda industria pesquera más grande del mundo. Se había dejado de declarar un volumen de captura que representaba 100 millones de dólares en impuestos.

El reportaje de Salazar para IDL-R sería luego ampliado e incorporado a una cobertura global coordinada por el Consorcio Internacional de Periodismo de Investigación (ICIJ). En el 2012, dicho trabajo fue también uno de los ganadores del Premio Latinoamericano de Periodismo de Investigación, entregado por el Instituto de Prensa y Sociedad (IPYS).

Ese mismo año, la Unidad de Investigación del diario El Comercio publicó una serie de reportajes sobre las empresas y los grupos familiares que más se beneficiaban de las compras del millonario Programa de Asistencia Alimentaria (Pronaa) del Estado –que poco después sería desactivado y reemplazado por el programa Qaliwarma–. Se construyó en forma manual una base de datos de las contrataciones de dicho programa en los últimos diez años, en base a los reportes del Sistema Electrónico de Contrataciones y Adquisiciones del Estado (SEACE). Así se pudo identificar a las empresas que obtuvieron más contratos, sus propietarios y sus antecedentes. Poco después, se aplicó ese mismo método para investigar las compras de medicamentos por parte del Estado y se descubrió que un monopolio farmacéutico estaba imponiendo excesivos precios al sistema público de salud. La fuente estrella era una hoja de cálculo.

**“El periodismo de investigación está más vivo que nunca. Su apoyo en la tecnología ha logrado incrementar la calidad y el impacto de las historias”.**

.....  
**David Kaplan**, *director del Global Investigative Journalism Network*  
.....

Se iniciaba una etapa en que los periodistas exploraban nuevas herramientas para mejorar su capacidad de análisis y pasar del reporte a partir de un indicio particular a una investigación general a partir de evidencia masiva. El proceso tendría implicancias más allá de los resultados de las investigaciones: desde el enfoque para plantear investigaciones hasta el propio metalenguaje de la profesión.

## **PERIODISMO + TECNOLOGÍA**

Una mañana de junio de 2014, en las instalaciones de un instituto tecnológico de Lima, más de cincuenta periodistas y programadores se reunieron por primera vez para generar herramientas que permitieran dar un salto en las maneras de obtener y procesar información. La actividad, una hackatón que duró doce horas continuas, fue bautizada con la expresión canónica que ha guiado el mejor periodismo: “La ruta del dinero”. Se trataba de un ejercicio de vigilancia ciudadana y periodística sobre el uso de los fondos públicos. El encuentro se realizó en simultáneo con los de otros grupos en doce ciudades de América Latina, pertenecientes a la comunidad HacksHackers, que une a periodistas y programadores dispuestos a reinventar los medios de comunicación.

La expectativa estaba alimentada por recientes revelaciones periodísticas con soporte tecnológico que estaban logrando alto impacto en el contexto de las elecciones municipales y regionales de ese año. La más clara era una inédita alianza entre la ONG Transparencia, un hacker y dos periodistas del popular sitio web de noticias Utero.pe. El resultado se llamó Verita, un software que cruzó la información de las hojas de vida de los más de cien mil candidatos con varias bases de datos públicas, como los registros de sentencias por delitos civiles y penales. Los hallazgos fueron sorprendentes: 1.395 candidatos a alcaldes y gobernadores regionales poseían condenas; más de la mitad por no haber cumplido con su responsabilidad de pagar la pensión alimentaria de sus propios hijos.



## DOS TECNÓLOGOS APOYAN AL PERIODISMO



Detrás del seudónimo **Aniversario Perú** está un biólogo de 35 años y padre de dos niñas que estudia un doctorado en Europa y que en su tiempo libre se convierte en hacker cívico. Él se define como: “Un tipo que conoce algo de programación y que quiere hacer accesible la información a la gente”. El artífice de los software “Verita”, que sirve para **explorar las hojas de vida de los candidatos en las elecciones**, y “Manolo”, que **extrae información de los registros de visitas oficiales a entidades del Estado**, cree que **el buen periodismo se parece a la ciencia**. “En Biología o Física tienes que probar una idea y lo haces buscando datos y analizándolos para llegar a una conclusión. Es lo que hace también un buen reportero de investigación”, apunta.

Experiencia parecida tiene el hacker **Antonio Cucho**, quien en el 2014 fundó **Open Data Perú**, una de **las principales comunidades que promueven la liberación de información** de interés público y su conversión en herramientas informativas. Su misión no es fácil en un país poco transparente, pero Cucho ya ha logrado sumar a sus filas a **870 jóvenes programadores** y profesionales **interesados en que el Estado abra sus datos** en los dos últimos años. Los municipios de Miraflores, San Isidro y Lima han captado el mensaje y ya tienen portales con información reutilizable.

El proyecto confirmó el nuevo rasgo del periodismo de la era del Big Data: el sentido colaborativo entre expertos provenientes de diferentes campos del conocimiento. Lo particular de este proyecto es que sus protagonistas, los periodistas Marco Sifuentes y Ernesto Cabral, y el director de la ONG Transparencia, Gerardo Távara, nunca se han reunido en persona con el hacker que los ayudó. El artífice de la descarga y cruce masivo de toda la información no vive en Lima. Es un joven biólogo peruano que estudia un doctorado en Europa y que en sus horas libres se dedica a la programación. Hasta ahora prefiere identificarse con el seudónimo de Aniversario Perú.

Poco después, el panorama periodístico peruano se dinamizó con el ingreso de nuevos medios independientes, de características más parecidas a las startups que a las de medios tradicionales. Se trataba de organizaciones pequeñas, con equipos de periodistas altamente calificados y métodos de investigación dinámicos. El rasgo esencial, que marcaba la diferencia incluso con experiencias parecidas en otras partes del mundo, era que su propuesta se enfocó desde el inicio en el contenido de investigación de impacto realizado con apoyo de recursos digitales.

El primero de estos medios fue **OjoPúblico**, creado por los reporteros Óscar Castilla, David Hidalgo, Nelly Luna y Fabiola Torres, de larga experiencia en los medios más importantes del país, en sociedad con el programador Antonio Cucho, activista de los datos abiertos y fundador de la comunidad Open Data Perú.

**OjoPúblico** remeció a la opinión pública con el lanzamiento de Cuentas Juradas, la primera aplicación periodística que reveló la evolución del patrimonio de las autoridades locales que pretendieron reelegirse en el mismo proceso electoral del 2014. A través de un análisis masivo de la información declarada por estos aspirantes en sus declaraciones juradas y en sus hojas de vida presentadas ante el organismo electoral, el portal publicó una serie de investigaciones sobre las incongruencias, vacíos y otros aspectos irregulares relacionados a sus bienes e ingresos. El trabajo se hizo en alianza con

la ONG Suma Ciudadana, a cargo de una parte fundamental de los pedidos de información pública por la vía tradicional, y tuvo el apoyo de los miembros del HackSpace de la Universidad Nacional de Ingeniería, que asumieron la descarga y procesamiento de buena parte de la información necesaria para el cruce de datos.

“Si no me aliaba con los periodistas de **OjoPúblico**, todo un archivo gigante de declaraciones juradas de diez años de los alcaldes en Lima hubiera terminado como una ruma de papel inútil”, dijo Javier Casas, presidente de Suma Ciudadana, sobre Cuentas Juradas. Su organización había reunido decenas de declaraciones juradas de alcaldes solicitadas a la Contraloría desde el 2012, pero tuvieron que pasar casi dos años hasta que encontró a un equipo de periodistas dispuestos a procesar e investigar a fondo esos documentos para convertirlos en historias impactantes.

Poco después se produjo el lanzamiento del sitio digital Convoca, dirigido por la periodista Milagros Salazar Herrera, quien organizó un equipo de cinco jóvenes reporteros y dos desarrolladores. Su primer trabajo fue la construcción de un registro completo de más de mil doscientos procesos sancionadores abiertos por el Organismo de Evaluación y Fiscalización Ambiental (OEFA) a empresas del sector minero, hidrocarburos, electricidad y pesca que cometieron malas prácticas entre 2010 y 2014.

Esta base de datos construida por los propios periodistas, combinada con un trabajo de reporterismo de seis meses, generó la serie “Excesos sin Castigo: el rastro ambiental de las industrias extractivas”. El análisis reveló que las empresas mineras y petroleras más multadas por el órgano supervisor del Ministerio del Ambiente son también las más reincidentes, y que habían establecido un esquema legal para apelar ante el Poder Judicial, donde habían logrado congelar más de 30 millones de dólares en sanciones. Se trataba de otra revelación producto de un trabajo multidisciplinario, en el que los periodistas detectan el potencial de una historia y organizan la estrategia y los recursos ad hoc para buscar evidencias.

**“Creemos que el futuro del periodismo se descubrirá a través de montones y montones de experimentos”.**

.....

**Corey Ford**, *director ejecutivo de Matter*

.....

El resultado es un curioso fenómeno en el periodismo peruano: en un contexto en que varios de los medios tradicionales más importantes contrajeron recursos para afrontar una crisis de ingresos, los nuevos medios digitales se convirtieron en una fuente de contenido de calidad, alto impacto e innovación. Los reportajes desarrollados por **OjoPúblico**, Convoca y Utero.pe pasaron del online al offline gracias a convenios con diarios de circulación nacional, interesados tanto en la solidez de las investigaciones como en replicar la viralidad de los reportajes en las redes sociales.

Aunque el trabajo con bases de datos y la dinámica que requiere todavía es un tema pendiente en la industria periodística local, diarios como El Comercio y La República se han unido a la tendencia con el uso de algunas herramientas digitales gratuitas como Tableau, para la visualización de resultados deportivos o electorales, o Thinglink, para contenido de entretenimiento. La inquietud por aprovechar la tecnología para hacer mejor periodismo ya está en el panorama. El proceso, por fortuna, es irreversible.

# ESTUDIO DE CASO: Cuidados Intensivos

[News apps o la noticia que nunca muere]

**E**l encuentro entre periodismo y tecnología ofrece un abanico escalable de posibilidades para exponer los hallazgos ante nuevas audiencias. Es una idea aceptada que mientras las visualizaciones permiten entender una historia a partir de un gráfico, las aplicaciones permiten entender varias historias en una misma pieza periodística y dar al usuario la posibilidad de encontrar vías alternativas para consumir contenido. Es un proceso muy distante del tradicional sentido unilateral de la información, la diferencia entre escuchar un discurso y tener una conversación. A través de las aplicaciones ofrecemos a los lectores acceso libre a datos específicos de su interés, una experiencia personal de navegación y, en consecuencia, la posibilidad de comprender un tema de la manera que le resulte más cómoda.

Un claro ejemplo es Cuidados Intensivos, una aplicación periodística creada por **OjoPúblico** para investigar el sector privado de la salud en el Perú. En un primer nivel periodístico, la herramienta revela la penetración de los grandes grupos financieros del país en el negocio de clínicas y centros médicos, que en los años previos se expandieron con escasa supervisión del Estado. Pero su valor para el lector-ciudadano-paciente es que se trata del primer registro de clínicas y médicos que acumulan sanciones administrativas y procesos penales por casos de negligencia médica y malas prácticas de atención a los usuarios.

## El trabajo periodístico detrás de CI

.....

En la investigación sobre el sistema privado de salud en el Perú y la construcción de la aplicación periodística Cuidados Intensivos participaron cinco periodistas y un desarrollador. El equipo de reporteros trabajó los datos en Excel y usó el programa Open Refine para limpiarlos y cruzar la información de las distintas bases.

.....

Esta aplicación es el resultado de la construcción de bases de datos propias para comprender las dimensiones del sector. En una primera etapa se realizó 52 pedidos de acuerdo a la Ley de Transparencia y Acceso a la Información Pública. A eso se sumó la revisión de archivos documentales y descargas masivas de datos de los sitios web de 44 clínicas del país. Este volumen permitió diseñar una plataforma de búsqueda que muestra las fichas de 61.372 médicos colegiados, 9.920 establecimientos de salud y 21 empresas administradoras de fondos (entre Entidades Prestadoras de Salud, aseguradoras y las clínicas que ofrecen sus propios programas de atención médica). De este modo, la herramienta permite que cualquier usuario indague por su cuenta si el médico o la clínica a quienes confía su salud -e incluso su vida- están autorizados para prestar servicios, cuál es su especialidad, su nivel de eficacia y si ha recibido sanciones administrativas o demandas judiciales por malas prácticas.

Las envergadura del trabajo de recolección de datos y su procesamiento para obtener conclusiones relevantes permiten establecer algunas lecciones claras para cualquier periodista que desee afrontar un desafío de esta naturaleza:

### 1. La dimensión de la data se transforma

El punto de partida de Cuidados Intensivos fue el diseño de una estructura de bases de datos fundamentales para entender el sector y la identificación de todas las instituciones del Estado que tenían información necesaria. En esta primera fase, el equipo se propuso elaborar cuatro padrones: el primero, de los grupos corporativos con inversiones en el sector salud; el segundo, de todos los establecimientos médicos privados registrados en el país (desde ópticas y consultorios hasta clínicas especializadas); el tercero, de las compañías aseguradoras y administradoras de fondos de salud; y el cuarto, de los médicos colegiados en el Perú.

El principal reto fue comprobar las fechas de actualización de los datos. Cuando los reporteros de **OjoPúblico** solicitaron por pri-

mera vez el registro oficial de establecimientos médicos privados a la Superintendencia Nacional de Salud (Susalud), recibieron como respuesta que la información estaba publicada en el sitio web de la entidad. Allí, encontraron un listado de 2.500 servicios de salud particulares. Sin embargo, un mes después y mientras procesaban esa primera descarga, advirtieron que Susalud había creado un nuevo formulario para clasificar la información de los establecimientos privados. El nuevo padrón tenía más de 9.000 entidades registradas e incluía nuevos términos. El trabajo inicial quedó incompleto y desactualizado. El equipo tuvo que scrapear nuevamente todos los formularios en línea para convertirlos en un archivo Excel.

Un problema similar se produjo con el padrón de 60 mil médicos colegiados. En determinado momento, al tomar muestras aleatorias de verificación, detectaron que el Colegio Médico no actualizaba de manera regular las especialidades acreditadas de sus afiliados. En varios casos, los datos estaban incompletos. Se tuvo que subsanar el problema a mano, con búsquedas específicas del historial de los especialistas.

## 2. La información oficial es contradictoria

El equipo realizó un segundo proceso de recolección de información de carácter más sensible para elaborar las fichas de los establecimientos de salud, médicos, compañías aseguradoras y administradoras de fondos de salud. Solicitamos todas las resoluciones sancionadoras impuestas por el Instituto Nacional de la Competencia y de la Protección de la Propiedad Intelectual (Indecopi) contra empresas privadas de servicios de salud disponibles desde 1992 hasta mediados de 2015. Durante ese período, Indecopi había sido el único organismo del Estado que se encargó de fiscalizar y sancionar las malas prácticas del sector privado en perjuicio de los pacientes. Sin embargo, solo tenía resoluciones archivadas desde el 2011.

La primera respuesta del Indecopi fue que descargáramos los documentos en PDF almacenados en su sitio web, pero insistimos

### Las herramientas digitales del proyecto

.....

El trabajo de programación se realizó con el lenguaje Python, escogido por su eficiencia y performance. La estructura se hizo con el framework Django, que tiene un potente sistema de administración de contenidos, además de ser un sistema modular que permite que la aplicación sea escalable. Para el sistema de gestión de base de datos se eligió PostgreSQL, por su capacidad de almacenamiento de grandes cantidades de datos. Y el sistema de búsqueda se realizó en Elasticsearch, que posee un potente motor..

.....

en que nos entregaran los documentos en físico. Solo así pudimos comprobar que el organismo disponía de más información que no había procesado. Al hacer el contraste encontramos 30 resoluciones sancionadoras contra clínicas que no aparecían en su sitio web.

La aplicación Cuidados Intensivos procesó más resoluciones que el propio Indecopi para elaborar el ranking de clínicas con más sanciones administrativas. Además, con la lista completa, los periodistas pudieron verificar que la mayoría de clínicas no pagó las multas impuestas y las impugnó en el Poder Judicial.

## 3. Los términos técnicos esconden revelaciones

Durante la investigación, debimos empaparnos de los términos técnicos usados en el Estado para categorizar los establecimientos de salud. Sin el manejo de esta jerga especializada, hubiéramos perdido datos relevantes. Una muestra clara apareció al detectar un indicador que la Superintendencia Nacional de Salud (Susalud) denomina: nivel de riesgo operacional. El dato correspondiente era un porcentaje sin mayor explicación.

En la jerga oficial, este concepto se refiere al resultado de la supervisión realizada por Susalud a los servicios privados para medir su grado de cumplimiento de las normas de atención a los pacientes (condiciones y equipamiento de los servicios de emergencia, unidad de Cuidados Intensivos, farmacia, etc.). Los inspectores de Susalud consignaban como nivel de riesgo lo que en realidad correspondía al porcentaje de cumplimiento. De este modo, cuando en los reportes se decía que una clínica presentaba “Nivel de riesgo operacional: 6%”, lo que en realidad se revelaba era que el establecimiento no cumplía el 94% de las normas de atención. El impacto del dato cambiaba de manera radical.

Para comprender la terminología, los periodistas recurrieron a expertos que les ayudaron a explicarla en un lenguaje sencillo para los usuarios y que fue colocada en las fichas de los establecimientos de salud evaluados de una manera comprensible.



#### 4. Si la base de datos no existe, siempre hay formas de construirla

Uno de los mayores retos fue resolver la falta de información. El Perú carece de un registro oficial sobre negligencias médicas sancionadas. La información fue solicitada al Ministerio de Salud, el Colegio Médico, la Asociación de Clínicas Particulares y el Poder Judicial. Ninguna de estas instituciones disponía archivo alguno sobre el tema. Resolvimos construir una primera base a partir de las denuncias aparecidas en los medios de comunicación, que luego hubieran sido formalizadas ante el sistema de justicia.

Para ello, parte del equipo se sumergió en los archivos de tres de los periódicos más grandes del país: El Comercio, La República y Ojo. Revisamos el periodo de 1991 hasta mediados del 2015 y luego volcamos la información a una tabla de Excel con los siguientes campos: nombre de la víctima, clínica u hospital donde ocurrió la negligencia, médico o profesional de salud denunciado, breve descripción de los hechos, y año de la ocurrencia. Con este panorama, contrastamos los nombres de los médicos y de los establecimientos involucrados en los registros del Ministerio Público y el Poder Judicial. Solo los casos formalizados (con sentencia o en proceso abierto) fueron incluidos.

#### 5. Si la información cuesta, date el gusto de liberarla

Las fichas de las empresas que prestan servicios de salud tienen información que proviene de la Superintendencia Nacional de Registros Públicos (Sunarp) y del Poder Judicial. El pequeño detalle es que ambas entidades del Estado cobran una tasa por cada búsqueda. Para consultar cada partida registral de la Sunarp hay que abonar cuatro soles, y una sola empresa puede tener varias actas registrales. En el Poder Judicial, por cada reporte para conocer el estado de una demanda debe pagarse un sol. El equipo decidió asumir el costo para acceder a la información de unas 50 empresas en las que se concentró la investigación y liberó los datos en Cuidados Intensivos.

## WORKSHOPS PARA REPORTEROS

*[Organizaciones aliadas para combinar periodismo y tecnología ]*

### Investigative Reporters & Editors, IRE

<http://www.ire.org/>

Organiza conferencias y cursos de capacitación para periodistas. Su sede está en la Escuela de Periodismo de la Universidad de Missouri. Dirige el programa NICAR, que promueve el uso de bases de datos para la investigación periodística.

### Global Investigative Journalism Network

<http://www.globalinvestigativejournalism.org/>

La Red Mundial de Periodismo de Investigación, creada en 2003 en Copenhague, organiza la Conferencia Global de Periodismo de Investigación cada dos años. Su próximo evento será en Johannesburgo, Sudáfrica, en el 2017.

### Centro Knight para el Periodismo en las Américas

<http://www.knightcenter.utexas.edu/>

Entrena en forma permanente a periodistas de América Latina y el Caribe en las últimas herramientas digitales a través de seminarios gratuitos en línea. Entre sus expositores están los periodistas líderes en el uso de bases de datos.

### Internacional Center for Journalists, ICFJ

<http://www.ijnet.org/>

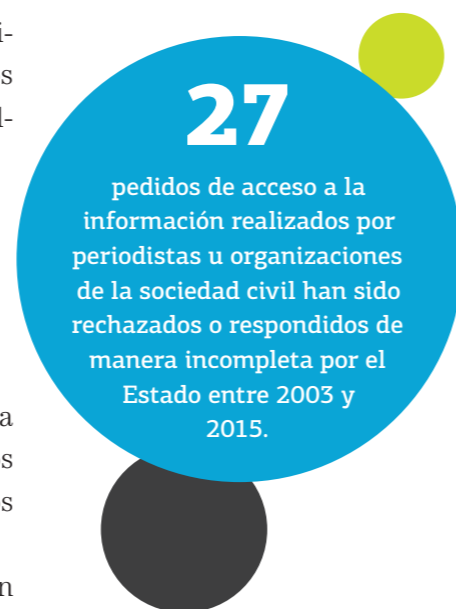
Esta organización ofrece entrenamiento en periodismo de investigación y herramientas digitales. Tiene un programa para desarrollo de proyectos periodísticos innovadores.

# Habemus o no habemus data

[Ley de Transparencia vs. Ley de Protección de Datos Personales]

**T**rece años después de que entrara en vigencia, la Ley de Transparencia y Acceso a la Información Pública en el Perú se ha convertido en una herramienta imprescindible para los periodistas de investigación. Sin embargo, hay varios retrocesos en esta materia. El principal escollo es el criterio que algunos funcionarios aplican a la información personal. En junio de 2015, el abogado Javier Casas solicitó a la Contraloría General de la República una copia de la declaración jurada de ingresos, bienes y rentas del presidente de la República, Ollanta Humala, con el argumento de que este documento es público y puede solicitarse de acuerdo a ley. Tras cuatro meses de insistentes reclamos, Casas recibió una carta en que se rechazaba su pedido con una dudosa explicación: las declaraciones juradas son información privada de los funcionarios públicos reguladas por la Ley de Protección de Datos Personales desde el 2011.

Casas, presidente de la ONG Suma Ciudadana y especialista en leyes de acceso a la información, consideró que la Contraloría había pasado por alto el precepto constitucional que establece como una obligación del Estado la difusión de las declaraciones juradas de los funcionarios para el control ciudadano. Pero el caso evidenció algo más: las fricciones entre la ley que promueve la apertura de los datos públicos y otra que los organismos estatales han empezado a utilizar a discreción para cerrar el acceso a los mismos.



Meses antes, Casas había entrevistado al jefe de la Autoridad Nacional de Protección de Datos Personales (ANPDP), José Quiroga León, quien aseguró lo siguiente: “Las declaraciones juradas de los funcionarios públicos no requieren consentimiento de los titulares para ser entregadas, porque no están dentro de lo que regula la norma de protección de datos personales”. El argumento de la Contraloría era entonces una interpretación particular para negar dicha información.

La paradoja de esta situación es que el Estado gasta 11 millones de soles al año en mantener una Secretaría de Gestión Pública para promover la apertura de la información y, al mismo tiempo, varios de sus ministerios y organismos niegan las solicitudes hechas por ciudadanos y periodistas.

Una revisión de los informes del Centro Liber, el Instituto de Prensa y Sociedad (IPYS) y la Defensoría del Pueblo permite identificar 17 entidades públicas que entre el 2003 y 2015 se negaron a respuesta a pedidos de información pública, lo hicieron en forma parcial o fuera del plazo límite. La mayoría de estos pedidos de información tenía fines periodísticos.

## ¿Podemos leer los correos de un ministro?

A mediados de 2014, los grupos de hackers Anonymous y Lulz Security fueron los artífices de la más grande filtración de correos electrónicos de un alto funcionario del Estado en el Perú. Sus miembros vulneraron la cuenta del entonces primer ministro, René Cornejo, e hicieron públicos 6.482 mensajes. La prensa bautizó el caso como los ‘Cornejoleaks’.

El volumen de lo filtrado tuvo gran impacto político y mediático, porque no solo reveló los presuntos lobbies secretos de los miembros del gabinete a favor de distintas corporaciones, sino que abrió un debate jurídico sobre el carácter público de las comunicaciones oficiales de las autoridades.

Entre los correos hackeados al entonces premier había una cadena de mensajes entre los ministros de Energía y Minas, Eleodoro Mayorga, y de Ambiente, Manuel Pulgar Vidal, sobre una norma aprobada para beneficiar en forma directa a compañías del sector hidrocarburos. Cuando los periodistas empezaron a reportar este hecho, el Centro Liber, una organización sin fines de lucro que promueve la transparencia en el Estado, pidió al despacho de Mayorga copia de los “correos electrónicos recibidos por el ministro a su cuenta de correo oficial o a cualquier otra que le haya sido creada en el ministerio, con sus respectivas respuestas, en las cuales se haya comunicado con cualquier persona para tratar asuntos relacionados con el Nuevo Reglamento Nacional de Hidrocarburos o sus similares”. La solicitud se hizo basada en la Ley de Transparencia y Acceso a la Información Pública, pero fue rechazada. El ministerio argumentó que ese pedido violaba el secreto de las comunicaciones del ministro.

Entonces el presidente del Centro Liber, el ex procurador anti-corrupción Julio Arbizu, decidió presentar un hábeas data con el argumento de que el contenido de la conversación del ministro de Energía y Minas era de interés público y no violaba de forma alguna su privacidad y el secreto de sus comunicaciones.

El caso quedó en manos del juez Hugo Velásquez Zavaleta, del Quinto Juzgado Constitucional de Lima, quien declaró fundada la demanda de hábeas data casi un año después, cuando Mayorga ya había dejado el cargo. “La información pública que puede solicitarse y que la administración estatal está obligada a entregar, puede constar en cualquier forma de expresión, ya sea gráfica, sonora, visual, electromagnética o que obre en cualquier otro soporte material”, indica el fallo de junio de 2015. La decisión consideró que “con la evolución de la tecnología, la comunicación ya no se efectúa solo a través del soporte de papel, sino a través de otros medios, como los correos electrónicos”.

El juez Velásquez se basó en dos principios: el de publicidad y el de máxima divulgación. El primero se encuentra regulado en el

**“La reutilización de datos hechos públicos por el Estado permite al ciudadano el ejercicio del derecho de acceso a la información”.**

.....  
**Miguel Morachimo**, *director de la ONG Hiperderecho.*  
.....

artículo 3 de la Ley de Transparencia y Acceso a la Información Pública, que establece que: “toda la información que posea el Estado se presume pública, salvo las disposiciones expresamente previstas”. El segundo principio fue desarrollado por la Corte Interamericana de Derechos Humanos y recogido en la jurisprudencia del Tribunal Constitucional: “la publicidad en la actuación de los poderes públicos constituye la regla general, y el secreto, cuando cuente con cobertura constitucional, la excepción”.

El Ministerio de Energía y Minas apeló el fallo pese a que la sucesora de Mayorga, la ministra Rosa Ortiz Ríos, dijo en una entrevista en televisión que entregaría la información pública hallada en el correo de su antecesor.

El Centro Liber también presentó recursos de hábeas data para solicitar correos de los ministros de Agricultura, Milton Von Hesse, y de Economía, Luis Castilla, en los que hubieran tratado asuntos directamente relacionados a su gestión. Al cerrarse este manual, ambos procesos seguían pendientes de fallo judicial.

En el debate sobre estos casos quedó claro que el periodista tiene en la Ley de Transparencia una herramienta poderosa para acceder a información sobre temas sensibles que atañen tanto a la gestión de funcionarios específicos como a las políticas públicas. No es una llave inmediata a la evidencia, pero debería ser uno de los criterios esenciales en la mentalidad del periodista de investigación.

## El dilema existencial: ¿Cuándo un dato privado es de interés público?

**E**n julio de 2011, la Presidencia del Consejo de Ministros (PCM) promulgó la Ley de Protección de Datos Personales con el fin de garantizar el adecuado tratamiento de la información privada de cualquier ciudadano. En teoría, es una norma positiva, pues evita la divulgación de datos sensibles (sobre salud, patrimonio personal, etc.), muy abundantes en la era digital. La norma no regula ni restringe el uso de datos públicos, pero admite zonas grises que han castigado el uso de bases de datos de acceso público que contienen nombres y datos de personas.

Un claro ejemplo es el caso del portal Datos Perú. En octubre de 2014, la Autoridad Nacional de Protección de Datos Personales (APDP) multó con S/. 228.000 al sitio por replicar las normas legales, nombramientos y sanciones administrativas contra funcionarios y empleados del Estado publicadas originalmente en el boletín de normas legales que viene adjunto al diario oficial El Peruano, una información que de por sí es pública aunque pocos se den el trabajo de revisarla.

Dos personas habían solicitado a los administradores del sitio web que eliminaran la información sobre sus casos, pero ellos se negaron a hacerlo por tratarse de datos recogidos de un documento público. Según la APDP, Datos Perú incumplió la Ley de Protección de Datos Personales por no tener el consentimiento de las personas para publicar esa información, pese a que el mismo contenido aparecía también en los portales de El Peruano y del Ministerio de Justicia.

**“Se necesita una autoridad autónoma que siga y fiscalice la entrega de información pública, porque los funcionarios encargados están limitados por las órdenes de sus superiores”.**

.....  
**Roberto Pereira, Centro Liber.**  
.....

Esta fue la primera sanción que efectuó la APDP y generó controversias, en principio, por la evidente contradicción de criterios para determinar lo público y lo privado cuando la misma información se divulga en un portal del Estado o en un sitio particular. A eso se sumaba una preocupación mayor: las posibles repercusiones de ese criterio para la práctica periodística y la vigilancia ciudadana. ¿Puede un funcionario decretar el secreto sobre lo que el propio Estado está obligado a informar? Y en un sentido más esencial: ¿Dónde empieza y dónde termina el interés público?

En plena época en que el mundo se encamina hacia una cultura de datos abiertos, los periodistas de investigación tienen varios obstáculos por superar. El primer paso es conocer todas las herramientas disponibles para obtener la información y para procesarla luego de manera innovadora. Hasta hace un tiempo, se trataba de un proceso más cercano a la artesanía y la intuición. La poderosa unión de periodismo y tecnología ha enriquecido los métodos y estándares de la profesión. Más que un set de nuevos instrumentos, tenemos un recurso estratégico. Es, como está dicho, un bagaje diverso y práctico, parecido a una navaja suiza.

**22**

indicadores tiene el Plan de Acción de Gobierno Abierto 2012-2014, pero no han sido cumplidos en su totalidad.



## LA LETRA PEQUEÑA: CÓMO HACER UN PEDIDO DE INFORMACIÓN EFICAZ



La Ley de Transparencia y Acceso a la Información Pública establece que los organismos públicos tienen como **plazo siete días hábiles**, después de haber recibido una solicitud, para emitir una respuesta. En caso de demora, tienen **cinco días de prórroga** para entregarla, previa comunicación al solicitante. Según el Manual Ciudadano para el Acceso a la Información Pública, elaborado por el Consejo de la Prensa Peruana, si un organismo público incumple la ley, como en varios de los casos presentados, **los ciudadanos tienen derecho a presentar un hábeas data**. Este recurso legal no solo procede cuando la institución pública rechaza un pedido de información, sino también cuando la que entrega es ambigua o cuando no emite respuesta alguna en los siguientes diez días hábiles posteriores al pedido recibido. El reportero también puede presentar un hábeas data cuando la entidad negó el acceso a determinada información luego de que se presentó una apelación dirigida al funcionario responsable con la indicación de que sea revisada por su superior jerárquico. En estos casos, sin necesidad de estar representado por un abogado, **el periodista tiene 60 días hábiles para presentar un hábeas data** ante el juez civil o el juez mixto de la zona donde vive o del lugar donde se ubica la institución pública que negó su pedido.

### SOLICITUD DE ACCESO A LA INFORMACIÓN PÚBLICA

Nombre de la Institución

#### I. Funcionario encargado de entregar la información:

#### II. Datos del solicitante:

Nombres y apellidos o razón social:

Documento de identidad y número:

Dirección:

Correo electrónico:

Número telefónico:

#### III. Información solicitada:

#### IV. Dependencia a la que se solicita la información:

#### V. Forma de entrega de la información

Nombres y apellidos:

Fecha:

Firma:

Observaciones: \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Aunque la ley no determina un formato oficial de solicitud, este modelo ayuda a pedir información pública.

## **SOBRE LOS AUTORES**

---

### **David Hidalgo**

*Director Periodístico  
de OjoPúblico*

---

Es autor del libro “Sombras de un rescate”, sobre la última acción armada del grupo terrorista MRTA. En 2006 ganó el premio nacional de Derechos Humanos y Periodismo. Becario del Edward R. Murrow Program for Journalists del Departamento de Estado de EE.UU. Integró el equipo peruano ganador de los Data Journalism Awards 2015.

---

### **Fabiola Torres L.**

*Editora de Análisis de Datos  
de OjoPúblico*

---

Periodista de investigación especializada en temas de salud, poder corporativo y gestión pública. Miembro del Investigative Reporters and Editors (IRE). Fue becaria de la Fundación Kiplinger de la U. de Ohio, y de la Global Investigative Journalism Network (GIJN). Integró el equipo peruano ganador de los Data Journalism Awards 2015.

---



OjoPúblico



Consejo de la Prensa Peruana

Con el apoyo de:



ISBN: 978-612-47141-0-8



9 786124 714108