

MANUAL DE PERIODISMO DE DATOS IBEROAMERICANO

Editado por Felipe Perry y Miguel Paz



Un proyecto de Fundación Poderomedia en alianza con la Escuela de Periodismo de la Universidad Alberto Hurtado y con el apoyo de HIVOS y ICFJ.

Manual de Periodismo de Datos Iberoamericano

Editado por Felipe Perry y Miguel Paz

Este libro ha sido posible gracias a una alianza de Fundación Poderomedia con la Escuela de Periodismo de la Universidad Alberto Hurtado y el apoyo de Hivos y el ICFJ -Centro Internacional para Periodistas-.

Un proyecto de:



Con el apoyo de:



Introducción

Este libro comenzó a gestarse entre cafés y conversaciones de varios periodistas y otros profesionales de los medios en el año 2012. La idea inicial fue hacer un paralelo latinoamericano del “Data Journalism Handbook”, el libro de bolsillo de periodismo de datos orientado casi totalmente al público anglosajón. Pero nuestro continente es distinto y con el paso del tiempo y la observación de la realidad del periodismo en general y del periodismo de datos en la región, el objetivo dibujado cambió: cómo unificar en un sólo texto nuestra diversidad, sin perderla, para crear una herramienta que sea útil para los nuevos periodistas y para quienes están entrando en el mundo de contar historias apoyadas en datos.

Dar voz a los mismos protagonistas fue una opción desde el principio del proyecto y el resultado fue alentador: los periodistas, diseñadores y programadores de habla hispana y portuguesa tenían ganas de contar sus experiencias y lo que pasa en sus países. La red de colaboradores se activó y el Manual fue tomando vida propia de la mano del trabajo comprometido de los integrantes de nuestra comunidad.

La primera y gran diferencia en el ejercicio del periodismo de datos en nuestros países respecto de EE.UU. y Europa es el acceso a los datos. En países donde los principios de transparencia están más arraigados, los periodistas obtienen datos actualizados y en formatos adecuados como parte de una práctica cotidiana. Nosotros no. En la región hay estados que aún no se imponen a si mismos el punto de partida que significa discutir y formular leyes de transparencia y de acceso a la información pública. Aún así, los periodistas de esos países generan historias y revelan a los ciudadanos información pesquisada desde miles de documentos, muchas veces dispersos o que no fueron concebidos con el fin de aportar datos útiles a la población. El trabajo de recolectar, procesar y entregar los datos en escenarios adversos ha producido un aprendizaje, no exento de ingenio, que este libro pretende rescatar.

De la reflexión surgida en la edición de los artículos que ahora entregamos, nos parece clave para las posibilidades futuras de un periodismo de datos de calidad el hecho de que aún transitamos, en lo respecta al acceso a la información pública, por un territorio en formación. Esa situación, de ir haciendo un camino, nos debe permitir estar alertas para no caer en el facilismo al momento de procesar datos y crear formas de visualización. La rigurosidad, la obsesión por los detalles e imponernos la decisión de abordar temas de impacto social, son elementos que permitirán la generación de historias relevantes, bien contadas y documentadas. Historias que signifiquen un aporte a las sociedades que las inspiran.

El Manual de Periodismo de Datos Iberoamericano no habría sido posible sin el apoyo y el compromiso de las organizaciones involucradas: HIVOS, International Center for Journalists (ICFJ) y la Escuela de Periodismo de la Universidad Alberto Hurtado de Chile. Sobre todo agradecemos a las personas que voluntariamente colaboraron en la creación de los textos, entregando su tiempo y conocimientos sin otro fin que el de hacer crecer el periodismo de datos en Iberoamérica.

Periodismo de Datos: historia y momento actual

Ana María Ávila

Inicios del periodismo de datos

No le creas a nadie, gritaba el director del diario cada vez que salía de la redacción y me veía pasar por su oficina. Era mi primer trabajo y temía mi ingenuidad al escuchar las declaraciones de los funcionarios públicos. Los tiempos han cambiado. Hoy, además de portar siempre preguntas e hipótesis que conducen a investigaciones; hay herramientas como las bases de datos que no dejan lugar a dudas. Son precisas y comprobables.

En la última organización donde participé había especialistas en estadística, bases de datos, programación e informática. El periodismo de investigación se vivía como un proceso colaborativo que se enriquecía con hojas de cálculo, mapas satelitales y visualizaciones capaces de viralizar una nota y explicarla con un vistazo.

Clay Shirky, profesor de la Universidad de Nueva York, ve al periodismo actual como un ecosistema de esfuerzos de investigación que reúne a diversas disciplinas en una red no jerárquica de colaboradores para responder a las necesidades de rendición de cuentas de los ciudadanos.

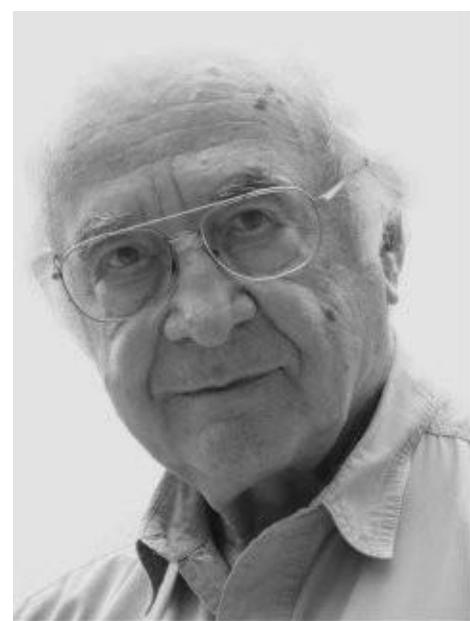
Sin duda el acontecimiento que permitió el arranque de este periodismo fue la llegada de internet, la cual derrumbó el monopolio de la producción de noticias en manos de los medios impresos, la radio y televisión. Sin embargo esta historia tiene muy viejas raíces.

En el primer tercio del siglo XX yacía latente la preocupación de hacer del periodismo, un oficio medible y comprobable. Max Weber en Alemania y Robert E. Park en Estados Unidos plantearon el sentido complementario entre el trabajo del periodista y el científico social.

De igual manera, el periodista Walter Lippmann tenía la esperanza de poder proveer un entrenamiento profesional en los diarios para subir los estándares en las noticias. Le parecía que el mundo se había vuelto muy complejo y que el periodismo no podía seguir usando sus herramientas convencionales.

Una de las razones del problema, según Lippmann, no sólo radicaba en la falta de adaptación de los periodistas y sus medios a los tiempos modernos, sino a lo intrincado y denso del objeto de estudio. Para el neoyorquino, el periodismo sólo sería capaz de reportar la complejidad moderna si hacía uso de observatorios políticos que examinaran las relaciones humanas con instrumentos y métodos científicos (Schudson, 2010: 101). Además, hacía un llamado a organizaciones científicas, independientes y apartidistas para que se comprometieran en la agenda de la investigación política y social.

El siguiente paso fue en los años treinta cuando se incorporó en la Universidad de Wisconsin un programa de doctorado especializado en periodismo dentro del área de Ciencia Política y Sociología (Dader, 1993: 103). Sin embargo, el gran salto al periodismo apegado a metodologías científicas fue el dado por [Philip Meyer](#) en 1959 cuando, cansado de que los diarios estuvieran cargados de entrevistas a políticos y boletines de prensa oficiales, decidió que crearía una fuente propia elaborando una base de datos.



"Meyer110719a" by Philip Meyer - Own work. Licensed under Creative Commons Attribution-Share Alike 3.0 via Wikimedia Commons

En ese entonces trabajaba en The Miami Herald y eligió una queja ciudadana sobre lo costoso que resultaban los seguros escolares contra incendios y huracanes, los cuales eran administrados por un consejo de funcionarios elegidos por la comunidad. Meyer investigó el financiamiento de las campañas de los miembros del consejo, así como las aportaciones que recibieron. Con un directorio de empresas consiguió los nombres de los ejecutivos aseguradores y, con lápiz y papel en mano, demostró que 65 por ciento de las aportaciones de la campaña del presidente tenían origen en las compañías de seguros. Otros dos miembros del consejo conseguían más de la mitad de sus recursos de la misma fuente.

Meyer decidió especializarse y gracias a una beca Nieman se dedicó a estudiar métodos de investigación social en la Universidad de Harvard. En el verano de 1967, luego de las marchas y disturbios causados en Detroit, los diarios publicaban que la mayoría de los manifestantes eran personas con baja escolaridad, que provenían de sectores marginales y grupos minoritarios. Meyer, que trabajaba para The Detroit Free Press, [descubrió que los manifestantes tenían estudios universitarios y eran de clase media](#). Después de ganar el Pulitzer con este trabajo, nació junto con sus métodos de investigación sociológica el periodismo de precisión, precursor de lo que hoy se conoce como periodismo de bases de datos o de datos.

Tres años más tarde The Washington Post publicaba un reportaje de precisión sobre las formas de reclutamiento de jóvenes para la guerra de Vietnam. En 1971 en la Universidad de Oregon, el profesor Everette Dennis incorporó a una clase el término de periodismo de precisión para definir así a los reportajes que usaban el método científico.

Al año siguiente, el periodista Gene Roberts, director ejecutivo del Philadelphia Inquirer, contrató a un periodista para que se dedicara de tiempo completo a analizar los datos del censo. Era la primera vez que un periodista no salía de la redacción para trabajar su fuente. Durante los años en que Roberts encabezó el diario, de 1972 a 1990, el periódico ganó 18 Premios Pulitzer.

En 1975 nació la organización [Investigative Reporters and Editors \(IRE\)](#) cuya aportación al periodismo de investigación y de bases de datos ha sido fundamental para fomentar trabajos a profundidad y con rigor científico. Actualmente, es la asociación que reúne a más periodistas de investigación en Estados Unidos. Todos los años cuenta con una serie de cursos y talleres para capacitar a periodistas en las metodologías de las bases de datos (antes conocido como periodismo asistido por computadora o CAR, por sus siglas en inglés), ahora incorporando la visualización de

datos, mapeo y nuevas tecnologías. La IRE alcanzó notoriedad tras el asesinato de uno de sus fundadores en 1976. Dono Bolles, reportero del Arizona Republic, en Phoenix, murió cuando investigaba temas de corrupción política y gangsterismo. A partir de ahí la Universidad de Missouri abrió un espacio a la organización para garantizarle estabilidad a la organización.

A finales de los años 70 y principio de los 80, los periódicos Philadelphia Inquirer, The Dallas Morning News, The New York Times, Los Angeles Times, entre otros, ya practicaban el periodismo de precisión y ganaban premios con ello. Sin embargo, no fue sino hasta 1992 que las noticias sobre periodismo de precisión llegaron al mundo de habla hispana con la publicación en El País de España de un texto sobre el tema bajo la autoría de Pedro Gómez y José Luis Dader (Esquivel Hernández, 1996: 158). Al año siguiente, éste último tradujo al castellano y publicó el libro de Meyer “Periodismo de precisión, nuevas fronteras de la investigación periodística”.

Los países latinoamericanos se encontraban embebidos en la lucha por la libertad de expresión, pues muchos atravesaban por dictaduras que impedían cualquier ejercicio libre de la prensa. De ahí que especializaciones en bases de datos y transparencia sonaban imposibles ante las dificultades de acceso a la información.

El periodismo de datos hoy

Con la llegada de internet, el acceso a fuentes de información se volvió inmediato y gratuito, lo que propició un salto exponencial en el uso de la tecnología para el periodismo. Entre 2005 y 2007 comenzaron a surgir herramientas digitales y tecnologías a precios accesibles para el manejo de datos. También en 2007 la [Knight Foundation](#) premió el proyecto Everyblock que permitía a los usuarios conocer información sobre su barrio gracias a una programación derivada de una base de datos (Ferrerez, 2012: 121). En 2009 el St. Petersburg Times recibió el Premio Pulitzer por el proyecto [Politifact](#). En ese año el jurado calificador dijo que el trabajo “demostraba que los periodistas junto con el poder de internet habían podido separar los hechos de la retórica de las campañas electorales para iluminar a los votantes”. Jack McElroy, editor del Knoxville News Sentinel, dijo “las bases de datos en línea se están convirtiendo rápidamente en una herramienta importante para el periodismo *watchdog* en la era digital. Al identificar a Politifact como el mejor reporte del año, se va a acelerar esta tendencia”.

Gracias al interés que tenían tres periodistas en la convergencia entre periodismo y tecnología nació el movimiento [Hacks and Hackers](#) en 2009. Aron Pilhofer, de The New York Times; Rich Gordon, de Northwestern University, y Burton Herman, ex corresponsal de AP, crearon una comunidad internacional para fomentar encuentros entre periodistas y expertos en tecnología para intercambiar ideas y encontrar colaboración entre ambos mundos.

En 2010, el Centro Europeo de Periodismo (EJC, por sus siglas en inglés) organizó en Ámsterdam la [primera conferencia sobre periodismo de datos](#). Al año siguiente, en el MozFest celebrado en Londres, nació la iniciativa de escribir el primer [Manual de Periodismo de Datos](#) hoy disponible en línea de manera gratuita. El texto es en sí mismo una muestra del periodismo colaborativo, pues periodistas de diversos medios aportan sus herramientas y las comparten. Cuenta con la participación de profesionales de la BBC, Chicago Tribune, Australian Broadcasting Corporation, La Nación (Argentina), The Washington Post, The Texas Tribune, Pro Publica, The New York Times, The Guardian, entre otros.

En 2011, Knight Foundation premió 16 proyectos relacionados con la innovación en el ámbito del periodismo de datos y fueron destinados 4.7 millones de dólares para el desarrollo de los mismos. Hoy en día, [Centro Knight para el Periodismo en las Américas](#), [Poynter Institute](#) e IRE ofrecen permanentemente cursos en línea y presenciales para capacitar a periodistas en bases de datos,

visualizaciones y herramientas digitales para el periodismo en línea.

Periodismo de datos en América Latina y España

La llegada del periodismo de datos a España y Latinoamérica fue hasta los años 90 con la publicación del libro de Meyer en castellano. Desgraciadamente, nunca vino acompañada del apoyo institucional de los medios para operar de manera articulada investigaciones con bases de datos. Mientras que en Estados Unidos importantes redacciones fueron capaces de incorporar equipos de programadores, especialistas en estadística y visualización de datos, en el mundo hispanoparlante los esfuerzos fueron aislados y, en muchas ocasiones, bajo la iniciativa de organizaciones de la sociedad civil.

Es importante señalar que un gran incentivo para el periodismo de datos en Estados Unidos fue la Ley de la Libertad de la Información (FOIA, por sus siglas en inglés) que fue firmada por el presidente Lyndon Johnson –pese a sus resistencias— el 4 de julio de 1966 y entró en efecto al año siguiente. En España, por ejemplo, la ley de transparencia entró en vigor apenas en 2013, lo cual tiene efectos en la producción de periodismo de datos.

En México tampoco hay medios que hayan incorporado a su redacción un equipo para hacer periodismo de datos. Si bien existe una ley de transparencia que entró en vigor en 2003, los esfuerzos por usar las bases de datos han sido aislados. Es imposible decir que haya una intención dentro de los medios de comunicación para coordinar trabajos junto con programadores, reporteros especializados en el análisis de datos y visualizadores. Han sido los periodistas en procesos individuales de profesionalización, o bien organizaciones de la sociedad civil como [Fundar](#) y [Artículo 19](#), en México y [Fundación Civio](#) y [Medialab](#), en España las que han usado las herramientas de transparencia de la información y han aplicado el periodismo de datos.

Asimismo, en México hay organizaciones como [Social Tic](#) y [Escuela de Datos](#) que han impartido capacitaciones a periodistas para el manejo de bases de datos, así como también han creado proyectos colaborativos conjuntos. También existe el capítulo México para Hacks and Hackers, el cual ha podido convocar a periodistas y programadores para crear Apps y descifrar bases de datos.

Es importante señalar que para haber esfuerzos institucionales en el uso y manejo de bases de datos, no es suficiente con tener leyes que lo permitan, claro que son un gran aliciente, pero no basta. Los periodistas mexicanos han tenido que enfrentar los ataques del crimen organizado y en muchas ocasiones sortear la violencia ha sido prioritario. Además, la transición interna de los medios a la democracia ha sido lenta, a pesar de que el país tuvo sus primeras elecciones libres hace 14 años. Todavía se siguen viendo prácticas autoritarias y relaciones opacas con el poder que impiden abrir paso a la transparencia y la profesionalización.

En las escuelas de periodismo el panorama es incipiente. Tanto en España como en México empieza a incorporarse una matrícula para dar formación sobre bases de datos, su interpretación y visualización, mientras que en los países anglosajones está presente desde los años setenta.

En otros países de América Latina el escenario no es tan precario. En Argentina el diario La Nación ha incorporado a su redacción un equipo de programadores y periodistas especializado en periodismo de datos. El periódico ya ha sido nominado a varios premios, además cuenta con un [datablog](#) que tiene entre sus asiduas colaboradoras a Sandra Crucianelli, quien es una pionera en el periodismo de datos y dedica gran parte de su tiempo a la formación de periodistas.

En Brasil, O Estado de S. Paulo tiene un coordinador, dos periodistas y un desarrollador que escribe código para las visualizaciones o las aplicaciones. Los trabajos además de publicarse en el diario, cuentan con un espacio especial en la página Web (Zanchelli y Crucianelli, 2012: 11).

El periodismo de precisión de ayer, el periodismo de datos de hoy

La administración en la abundancia es en parte una de las razones de su auge. Gracias al internet, a las iniciativas de transparencia y gobierno abierto en el mundo, la información al alcance de los periodistas es tan abundante que su sistematización, análisis e interpretación se ha vuelto crucial. De ahí viene el salto de lo que conocíamos como periodismo de precisión o *Computer Assisted Reporting* (CAR) a lo que hoy conocemos como periodismo de datos. Los elementos que lo caracterizan son: el acceso a grandes volúmenes de datos, su tratamiento y publicación. Para esto último se usan las nuevas tecnologías disponibles en línea (Crucianelli, 2012).

También debe añadirse que gran parte del mérito del periodismo de datos no es tanto encontrar una base de datos en concreto, sino el comprobar una relación peculiar entre variables de sucesos que podrían parecer aislados, o bien entre diferentes bases de datos. Es un principio en común con el periodismo de precisión cuya aportación más espectacular “ha consistido sin duda en la posibilidad de cruzar diferentes listados de información para detectar asociaciones entre personas, instituciones, cargos, circunstancias, etc, que aparecen aisladas en cada documento pero que tienen una presencia reiterada no detectada hasta ese momento. Puede afirmarse, en ese sentido, que la noticia bomba duerme en los archivos muy a menudo” (Dader, 1993: 108).

Por otro lado, la capacidad de extraer grandes cantidades de información y sistematizarla no es (aún) periodismo de datos. Es decir, conseguir los datasets o acceder a información que parecía confidencial no es un fin periodístico en sí mismo. El fin último dentro del periodismo es el interés público. Encontrar asociaciones que no se habían visto en cantidades abundantes de información disponibles en la Web para hacerlas públicas forma parte de la rendición de cuentas. Los poderes de un Estado, al verse obligados a transparentar sus actividades, permiten a los periodistas mantener y actualizar su tarea originaria de *watchdogs*.

Un ejemplo que debe mencionarse es el de Wikileaks. Después de las revelaciones de 2010, se dijo que la era del secretismo se había terminado, que la transparencia radical había llegado para quedarse. Ahora, a la distancia, luego de los diversos análisis sobre sus verdaderas aportaciones y la evaluación de si es correcto que la organización de Julian Assange y sus miembros se autodenomine como periodística, ¿qué se puede concluir? Es verdad que los más de 250 mil mensajes contenían información relevante sobre la guerra en Afganistán, las fiestas de Berlusconi, el detallado seguimiento a Sarkozy y los movimientos para bloquear a Irán, pero también es cierto que la información en sí misma se volvió periodística cuando Wikileaks se acercó a The Guardian, The New York Times, Der Spiegel, Le Monde y El País. La organización tuvo que buscar a los periodistas para que tradujeran los documentos, corroborarían con sus fuentes la información y pudieran convertirla en historias. Claro que los documentos en sí mismos tienen valor, pues sin ellos hubiera sido imposible hacer las revelaciones, pero sin el trabajo de verificación, análisis y sistematización de los miles de documentos hubiera sido muy complicado que se convirtieran en historias de interés público. La mancuerna entre quienes fueron capaces de interceptar las comunicaciones en el ciberespacio y los periodistas fue crucial. El periodista británico John Lanchester lo describió con las siguientes palabras: “la información revelada por Wikileaks no tenía precedentes, pero no es periodismo. Los datos tienen que ser interpretados, estudiados y convertidos en una historia” (Roberts, 2011: 18).

Cuando Wikileaks trató de hacer pública información sin el apoyo de los diarios más importantes del mundo, la gente no les creyó. Assange llamó a conferencia de prensa en el Club Nacional de Prensa en Washington D.C. y mostró un video de un helicóptero que sobrevolaba en Bagdad disparando indiscriminadamente a civiles que sin provocación alguna fueron asesinados. La reacción de las audiencias fue de escepticismo y acusaron a Assange de haber editado el video. El caso de Wikileaks es un buen ejemplo para describir otro más de los rasgos característicos del periodismo de datos y es su naturaleza colaborativa. Como se ha dicho anteriormente, el periodista especializado en el manejo

de bases de datos no puede actuar en solitario. Necesita de un programador y/o un desarrollador que usa el código para las visualizaciones. Las redacciones que ya han adoptado el periodismo de datos trabajan en equipo y de manera horizontal, desde luego hay un coordinador del equipo y un editor, pero en un trabajo de periodismo de datos la retroalimentación entre los integrantes del equipo determina el éxito de la historia.

Al referirse a la necesidad de la presencia de programadores y periodistas en la redacción, citados en “Integrando el periodismo de datos en la sala de redacción”, dos expertos dicen: “estar en la sala de redacción realmente importa” (Paul Overberg, editor de base de datos del Usa Today) y “las organizaciones de noticias se basan todas en geografía –en la proximidad a la mesa de redacción. Si estas cerca, es fácil sugerir reportajes y convertirte en parte del proceso” (Simon Rogers, ex editor del Datablog de The Guardian). Al ubicar a desarrolladores y periodistas en la misma sala de redacción se facilita el flujo de ideas para reportajes, las cuales surgen tanto de los desarrolladores y los expertos en datos como de los periodistas (Zanchelli y Crucianelli, 2012: 3, 5).

Se necesita el periodismo de datos

Después de la crisis que trajo internet para los medios tradicionales ha quedado claro que el papel del periodista sigue vigente y que, más que nunca, su rol es fundamental. El periodismo de hoy requiere de profesionales que, ante la marea de información contenida en la red, puedan curar la información, verificarla y aplicar técnicas rigurosas para presentarla ante la ciudadanía. Una manera de hacerlo es a través del periodismo de datos. Sin embargo, la rutina en la que llegan a estar inmersos los periodistas vuelve su tarea mecánica y, en ocasiones, los reduce a transmisores de la información oral a formatos impresos o audiovisuales.

Se necesita de un impulso institucional que nazca de las direcciones de los medios, así como ha sucedido en algunas redacciones de Argentina, Brasil, Colombia y Costa Rica. Además, las universidades latinoamericanas necesitan integrar a sus estudiantes en la era digital, o bien incorporar e impulsar los logros que han alcanzado periodistas y programadores al reunirse para trabajar juntos como en las múltiples ediciones latinoamericanas de Hacks and Hackers.

El periodismo de datos es un micro universo donde se respetan las diversas disciplinas, se escuchan las propuestas de todos y se construye en comunidad. El periodismo de datos no sólo trabaja para la democracia, sino que debe funcionar de manera colaborativa y bajo un régimen democrático interno, sino es poco efectivo. Sólo así se pueden desarrollar proyectos en beneficio del interés público. Así como un día la Universidad de Missouri decidió abrazar el proyecto de la IRE, de igual manera se requiere que tanto los periodistas, como la academia y los desarrolladores empujen esfuerzos para consolidar un periodismo de datos tan necesario hoy en día para la transparencia y la rendición de cuentas.

Ana María Ávila

Periodista. Colaboradora de varios medios de comunicación en México. Participó en investigaciones periodísticas para The Center for Public Integrity y el International Consortium of International Journalists. Es maestra en Ciencias Sociales y cursó una estancia académica en periodismo digital en la Universidad de Maryland, Estados Unidos. Fue fundadora de la maestría de periodismo del Centro de Investigación y Docencia Económicas, CIDE y fue co-ganadora del premio de periodismo ambiental de la Universidad de Columbia, Estados Unidos 2010.

Manual de reducción de riesgo digital y móvil para periodistas y blogueros

Jorge Luis Sierra

Los reporteros y editores de investigación necesitan incluir los planes de seguridad como parte ineludible de su quehacer. Al investigar temas que otros desean mantener ocultos, los periodistas enfrentan la agresión de fuentes de naturaleza muy diversa. Las formas de protección son distintas.

Mientras una amenaza potencial se enfrenta a través de medidas preventivas, las amenazas reales se enfrentan usualmente con medidas correctivas, de emergencia. Los periodistas deben desarrollar planes de reducción de riesgo a través de un fortalecimiento de sus capacidades profesionales, el ejercicio ético de su profesión y el seguimiento puntual de protocolos de seguridad.

Seguridad digital

Junto al hecho de que internet y los equipos digitales se han vuelto herramientas indispensables del periodista, la seguridad digital y móvil (celulares y tabletas) se han transformado también en un tema fundamental. Aquí te presentamos una serie de protocolos de auto protección a manera de guía de seguridad digital fuera y dentro de la redacción, así como en la comunicación entre reporteros y editores, reporteros y fuentes y el almacenamiento de información.

Medidas básicas de seguridad digital general:

- Usa un programa antivirus que se actualice automáticamente.
- Activa los firewalls o cortafuegos en tu computadora.
- Actualiza permanentemente los sistemas operativos, programas y navegadores que tienen tus equipos.
- Infórmate siempre de los nuevos virus y programas maliciosos.
- Examina con cuidado la dirección electrónica de quien te envía un mensaje.
- Nunca abras un archivo que no esperabas. Si decides abrirlo, examínalo antes con tu programa antivirus.
- Nunca le des clic a vínculos activos que te envían fuentes desconocidas o de baja confianza.
- Usa sólo contraseñas seguras (más de 15 caracteres, con mayúsculas, minúsculas, números y signos de puntuación).

Medidas básicas de seguridad digital en la redacción:

- Cierra tu computadora cuando te alejes de ella o protege la pantalla con una contraseña fuerte.
- Cierra tu navegador y desconecta tu equipo de internet si no navegas o si te alejas de tu equipo.
- Respalda y organiza de manera permanente tu información.
- Usa carpetas encriptadas para guardar información confidencial o encripta el disco duro entero.
- Si usas un equipo de la empresa en tu trabajo, evita usarlo para fines personales, en los equipos y redes de la empresa nada es privado.
- Si estás en la red de la empresa, sigue las instrucciones de seguridad del departamento de IT (Tecnologías de la Información).
- Acuerda con los directivos de tu empresa los medios de comunicación segura con fuentes confidenciales.

Medidas básicas de seguridad digital fuera de la redacción:

- Sólo lleva el equipo necesario.
- Activa medidas de seguridad en tus dispositivos móviles en caso de pérdida.
- Nunca lleves tus archivos completos en tu computadora mientras cumples una asignación o realizas una cobertura.
- Lleva un dispositivo de memoria USB con el sistema de navegación Tor.
- Si no es necesario, desactiva los sistemas GPS de tu teléfono celular o tableta.
- Desactiva el Bluetooth de tu teléfono celular si no lo usas.
- Lleva baterías adicionales para tu computadora, teléfono celular o tableta. Evita usar equipos en las salas de prensa.
- De preferencia, evita cargar la batería de tus equipos en salas de prensa u otros lugares públicos porque [otros pueden acceder a tus equipos por esa vía](#).
- Evita acceder a redes inseguras de internet. De preferencia, accede a internet con un modem USB o con el recurso de Hotspot móvil de tu teléfono celular.
- Si consultas a fuentes confidenciales o intercambias datos o información confidencial con tu editor, usa formas encriptadas de comunicación.

Protocolo de protección de celulares y tabletas:

- Elaborar una política de contraseñas para el teléfono celular que incluya: PIN, patrón, contraseña para iniciar el equipo y acceder a la tarjeta de memoria.
- Evitar el almacenamiento de la información de contacto de fuentes confidenciales y de información personal en el teléfono que pueda estar en riesgo de pérdida o robo.
- Mantener asegurado el equipo móvil para reponer el aparato en caso de robo o extravío, así como borrar la información y bloquear el acceso a control remoto.
- Usar programas antivirus y mantener actualizadas las aplicaciones.
- Llevar consigo siempre baterías de repuesto y mantener siempre cargada la batería del dispositivo móvil.
- Descargar las aplicaciones de navegación anónima en el móvil (Orbot y Orweb). Existen buenos [tutoriales](#) y se puede [descargar Orbot](#) y [Orweb](#) en la tienda Google Play.
- Descargar la aplicación [Ostel](#) para encriptar el audio de las llamadas telefónicas.

Protocolo de protección de contraseñas:

- Elaborar una política de contraseñas seguras de arranque, pantalla protectora, así como las cuentas de administrador y usuario de las computadoras en riesgo.
- Las contraseñas deben ser mayores de 15 caracteres y combinar letras mayúsculas, minúsculas, números y signos.
- Usar una contraseña distinta para cada cuenta de correo electrónico, redes sociales y sitios web.
- Cambiar las contraseñas frecuentemente.
- Guardar en un programa de almacenamiento protegido (como [Keepass](#)) las contraseñas anotadas en papeles, libretas, o archivos sin encriptación [en la computadora](#). Las computadoras MacBook tienen el sistema de “Acceso de Llaveros” con la que puedes administrar las contraseñas y guardarlas en forma encriptada.
- Crear passphrases en lugar de Passwords. Las [passphrases](#) son conjuntos de cuatro o cinco palabras o contraseñas que funcionan juntas para abrir documentos, aplicaciones, programas o redes sociales.

- Medir la fortaleza de las contraseñas en sitios de Internet. ([Passwordmeter](#))
- Usar el programa Truecrypt para usar una contraseña en el arranque de la computadora y encriptar el disco duro interno.
- Revisar un tutorial para la [protección de equipos](#).

Protocolo de almacenamiento seguro

Los periodistas suelen almacenar grandes cantidades de archivos electrónicos sin separar los archivos con información confidencial de los que no tienen datos que requieran protección especial. Ese almacenamiento desorganizado de los archivos puede ser un factor de riesgo si perdemos el control de la computadora o del equipo móvil.

Quizá uno de los ejemplos más extremos de esa situación es el [secuestro de dos periodistas](#) en Reynosa, Tamaulipas, que cubrían el enfrentamiento entre los carteles del Golfo y de los Zetas en la primavera de 2010. Uno de los reporteros llevaba en su computadora fotografías tomadas en una ceremonia militar que había ocurrido días atrás en la ciudad de México. Al revisar las computadoras de los periodistas secuestrados, los secuestradores interpretaron esas fotografías como una muestra de que los reporteros trabajaban para los Zetas, el grupo criminal que fue integrado por ex miembros del ejército mexicano. De alguna forma, el reportero pudo demostrar su condición profesional y con eso evitar el asesinato de él y su compañero.

El ejemplo nos ofrece varias enseñanzas que los periodistas que cubren situaciones de alto riesgo tienen que recuperar. La primera es que hay que separar los archivos con información sensible de los que no la tienen. La información sensible puede abarcar desde archivos de Word con nombres, teléfonos o correos electrónicos de fuentes de información hasta notas escritas con datos acerca de la corrupción de funcionarios públicos.

Los documentos que el periodista considere confidenciales deben guardarse en carpetas encriptadas con Truecrypt, un programa gratuito disponible en internet que permite la creación de drives encriptados que sólo pueden abrirse con una contraseña. Truecrypt permite la creación de carpetas encriptadas ocultas en otro documento como una fotografía, un documento de Power Point o un documento de Word.

Los periodistas deben viajar con computadoras y celulares “limpios” de información. En caso de pérdida, robo o confiscación, el periodista puede asegurar que ningún dato confidencial está en riesgo. Eso significa que antes de salir a realizar la cobertura, los periodistas deben mover las carpetas encriptadas a otras computadoras seguras y viajar con computadoras vacías.

Si crean archivos sensibles durante una cobertura, éstos deben ser almacenados en carpetas creadas con Truecrypt y ocultar en una fotografía u otro archivo similar.

Los periodistas también pueden crear carpetas con Truecrypt y [almacenarlas en Dropbox](#) o [Google Drive](#). De esta manera, los archivos siempre estarán disponibles y protegidos con una encriptación doble.

Protocolo de protección de documentos

En el contexto de situaciones de alto riesgo, se ha vuelto indispensable que los periodistas desarrollen habilidades para proteger su información, sobre todo aquella que incluye datos sensibles o confidenciales.

Una de las formas de protección de nuestros documentos es la [codificación o encriptación](#), que implica la conversión de un documento legible a uno ilegible para quien no tiene la contraseña para decodificarlo. Si por alguna razón perdemos la memoria USB donde teníamos nuestro documento, o si perdemos la computadora, el teléfono o la tableta con ese documento, las personas no podrán leer el

documento si no tienen la contraseña de decodificación.

Navegación anónima en internet

La navegación anónima consiste en disfrazar el código IP (Internet Protocol) que el proveedor de servicios de internet asigna cuando un navegador (*browser*) se conecta con un servidor de internet. Este número único está asociado con los datos de identificación del cliente del proveedor de servicios de internet.

En el caso de los IP que son asignados a la navegación en una sala de redacción, el código está asociado a los datos de la empresa. Sin embargo, cuando el periodista navega desde su casa o desde su móvil o tableta, es posible que el IP esté entonces asociado con sus datos personales, por lo que cualquier persona interesada podría conocer la identidad de los periodistas a través del rastreo del IP. El IP también puede ser revelado en los encabezados (*headers*) de correos electrónicos inseguros.

Una de las herramientas disponibles para proteger la identidad es el navegador de [Tor](#). La plataforma Tor utiliza una red de servidores proxy que prestan su IP para enmascarar el IP original del usuario de internet. Cuando el periodista o reportero ciudadano está conectado a la red Tor y abre un nuevo sitio de internet, el servidor proxy usa su propio IP, en lugar del IP de la computadora del usuario, para pedir los datos de un nuevo sitio a otro servidor. De esta manera, la identidad del usuario queda protegida.

En teléfonos celulares con sistema operativo Android, la plataforma Tor puede ser usada a través de las aplicaciones Orbot y Orweb. Conviene enrutar el teléfono celular para aprovechar todas las ventajas de Tor en celulares, entre ellas, el uso anónimo de todas las aplicaciones que conectan al celular con internet.

La protección de la privacidad incluye también el uso del prefijo HTTPS en todos los sitios donde haya que escribir una contraseña y el cuidado para no abrir sitios web con contraseña en redes inalámbricas abiertas, sin contraseña de acceso.

Protocolos de navegación anónima:

- Utilizar sitios de internet que tengan el prefijo HTTPS. Ese prefijo encripta la interacción entre el navegador de internet y el servidor.
- Usar herramientas como https everywhere y Anonymox (Firefox) en el navegador web.
- [Usar la plataforma Tor](#) cada vez que estemos haciendo una investigación en línea sobre temas complicados.
- Usar Redes Privadas Virtuales (VPN por sus siglas en inglés) para navegar en internet. Las VPN se utilizan como un túnel para proteger la privacidad de la navegación en internet.
- Utilizar sólo el acceso encriptado a internet inalámbrico.
- No abrir cuentas que requieran contraseñas en redes de acceso inalámbrico abierto (sin encriptar) a internet.
- Usar el ambiente TOR con Orbot y Orweb en la navegación de internet en celulares y tabletas.

Protocolo de comunicación segura en correos electrónicos:

- Usar herramientas de codificación en la transmisión de mensajes vía correos electrónicos. El servicio de [Hushmail](#) encripta el contenido de los mensajes. Hushmail tiene la ventaja de que no pide datos personales para abrir una cuenta. Sin embargo, la compañía sí tiene acceso a nuestra contraseña.

- Otra forma de proteger el contenido de los correos electrónicos es el uso de una herramienta gratuita y sencilla de usar en internet: [Infoencrypt](#). Esta herramienta nos permite copiar y pegar un texto previamente escrito y encriptarlo mediante una contraseña. El texto encriptado puede a su vez copiarse y pegarse en el cuerpo del mensaje electrónico. El reto es enviar la contraseña al destinatario. Infoencrypt recomienda siempre enviar las contraseñas por una vía distinta.
- Existen otras alternativas accesibles y sin costo para proteger la comunicación entre el periodista, sus fuentes o sus editores. El servicio de [Riseup](#) encripta el mensaje electrónico en el tránsito desde el remitente hasta el destinatario. Riseup no incluye el IP en sus encabezados ni guarda las contraseñas elegidas por los usuarios.
- Utilizar cuentas de correo electrónico obtenidas desde la plataforma de Tor para abrir blogs anónimos o cuentas de redes sociales también anónimas.
- Utilizar sólo cuentas de correo electrónico que ocultan el IP de los usuarios de internet en los encabezados del mensaje.
- Utilizar un procedimiento de doble autenticación (contraseña más código enviado al móvil) para abrir las cuentas de correo electrónico, mensajería instantánea o redes sociales.

Protocolo de seguridad del chat

Los periodistas y reporteros ciudadanos han descubierto que la mensajería instantánea es una de las formas más rápidas y accesibles para comunicarse con las fuentes, colegas o editores.

Hay herramientas que nos ayudan a encriptar las plataformas de chateo o mensajería instantánea. Esas herramientas como [Pidgin](#) en sistemas operativos Linux y Windows o [Adium](#) en iOS, usan un recurso llamado OTR (Off The Record) para encriptar servicios de mensajería instantánea como Google Chat, Facebook Chat, Yahoo Messenger o Microsoft Messenger (MSN Messenger).

Usa una combinación de chats inseguros y seguros: puedes usar por ejemplo Whatsapp y Google Talk encriptado para reducir la atención de espías potenciales a tus conversaciones electrónicas.

Seguridad en mensajes de texto

Una manera accesible de encriptar los mensajes de texto es [TextSecure](#), una aplicación para encriptar los mensajes de texto en dispositivos móviles con Android. Es gratuita y puedes bajarla en la tienda de Google Play.

Encriptación de llamadas telefónicas

Guardian Project tiene una herramienta llamada [Ostel](#) para encriptar las comunicaciones telefónicas. Funciona Android, iPhone, Blackberry, Nokia, PC, Mac y Linux.

Jorge Luis Sierra (@latinointx)

Periodista mexicano especializado en temas de seguridad y defensa, y desarrollador de herramientas digitales para investigar el crimen y la corrupción. Es Director de los Knight International Journalism Fellowships y actual coordinador del proyecto de Seguridad Digital y Móvil para Blogueros y Periodistas, apoyado por las organizaciones Freedom House, el Centro Internacional para Periodistas (ICFJ) y la Fundación Knight. Sierra ha escrito varios libros y capítulos sobre seguridad y defensa, entre los que se destaca: *El Enemigo Interno: Fuerzas Armadas y Contrainsurgencia en México*.

Miles de registros y sólo una historia

No le tengas miedo al “mostrito”

Apuntes para investigar a una corporación privada con un análisis masivo de base de datos y un reporteo persistente

Milagros Salazar H.

Hay algunos que le tienen fobia a las arañas, a viajar en avión, a los lugares cerrados, a los espacios abiertos; yo le tenía fobia al Excel. Ni siquiera podía usarlo para hacer una simple rendición de gastos con sumas y restas. Me alteraba ese conjunto de celdas y fórmulas que parecían atacarme cada vez que intentaba anotar algo en una hoja de cálculo. Lo odiaba. La cura solo llegó con un interés superior a este rechazo: descubrir lo que estaba oculto entre centenares de miles de desembarques de anchoveta en el Perú, la segunda nación pesquera más importante del mundo. Un interés que me llevó a realizar un análisis masivo de datos en el Excel, que fue fundamental para publicar durante más de un año una serie de reportajes que pusieron al descubierto que los grupos más poderosos de la industria pesquera en Perú subreportaban la captura de pescado a gran escala sin que el Estado ejerciera una fiscalización eficaz.

La serie investigativa empezó a publicarse en setiembre de 2011 en [IDL-Reporteros](#), un medio digital pequeño pero con un enorme compromiso para investigar temas cruciales en Perú bajo la dirección del valeroso y experimentado periodista Gustavo Gorriti. En 2012, la investigación obtuvo uno de los tres primeros galardones del Premio Latinoamericano de Periodismo de Investigación, organizado por el Instituto Prensa y Sociedad y Transparencia Internacional. A inicios de 2012, se publicó una ampliación de estos reportajes en alianza con el Consorcio Internacional de Periodistas de Investigación (ICIJ).

Describir los vaivenes del proceso de esta investigación con data sobre la pesca peruana, es una valiosa oportunidad para decir varias cosas sobre el ejercicio periodístico, sus métodos y sus nuevos desafíos.

Para empezar, no fue fácil decidir qué aspecto de la industria pesquera ameritaba una investigación a fondo. Cuando comencé las indagaciones, a fines de 2010, rápidamente me di cuenta que el terreno estaba plagado de varios hechos pantanosos que no se conocían en la prensa, o se conocían mal. Existía una convivencia llena de conflicto de intereses entre el sector privado y el Gobierno. La Sociedad Nacional de Pesquería, el gremio empresarial más poderoso de la industria, podía pagar consultores y destacarlos al Ministerio de la Producción, encargado de regular la actividad pesquera, para que se hicieran cargo de la redacción, modificación o defensa de algunas normas. O conseguir que se prioricen ciertos temas en beneficio de la industria y se dejen dormir otros más espinosos. ¿Qué tema elegir en un sector donde la irregularidad y la ilegalidad parecía lo común y no la excepción?, ¿cómo investigar una industria poderosa sin que las fuentes te lleven de las narices?

¿Hacia dónde lanzar la red?

Decidí investigar un tema central en la pesca con un método que me permitiera indagar lo que hacían todas las empresas y no solo una, lo que sucedía en todos los puertos del país, y no en uno, en todos los desembarques y no en unos cuantos. Me propuse investigar si las naves registraban la cantidad de anchoveta (el principal recurso pesquero del mar peruano) que realmente pescaban o estaban subreportando al Estado como denunciaban los pescadores, quienes son los que más saben de lo que sucede mar adentro y mar afuera.

La investigación se ubicaba en un contexto mayor y de gran importancia. Se suponía, según los empresarios, que la industria pesquera se desarrollaba en un ambiente limpio, desodorizado, lejos de las trampas, después que en 2008 se aprobara la norma más importante del sector en los últimos años:

la ley de cuotas de anchoveta por embarcación que le otorga el derecho de pesca de una cantidad determinada a cada nave para evitar que la flota salga al mar a acapararse el pescado como niños eufóricos lanzándose sobre la piñata en una fiesta infantil.

Las compañías más influyentes financiaron esta ley que les aseguró más del 60% de las cuotas de pesca de anchoveta y congelar por diez años el porcentaje de los derechos de pesca que debían pagar al Estado por cada tonelada capturada. Lograron un negocio millonario. Para darnos una idea: cada punto porcentual de la cuota llegó a valorizarse en 100 millones de dólares después de la aprobación de esta norma.

La revisión de los datos de los desembarques iba a permitir rastrear si las trampas en los puertos continuaron o no después de la aprobación de ley y en qué medida. Se trataba de comparar dos registros oficiales durante la inspección de los desembarques: la cantidad de anchoveta que declaraban los patronos de los barcos a los inspectores cuando llegaban a los muelles y la pesca pesada registrada en las balanzas que se ubican dentro de las fábricas, donde se procesaba la anchoveta para convertirla en harina principalmente para la exportación. Esta comparación se realizó a partir de un método validado y riguroso que se explicará más adelante, y que permitió demostrar un subreporte masivo y millonario de la anchoveta.

El principal desafío de la investigación fue la abundante data a la que se necesitaba acceder para luego limpiarla, analizarla y contrastarla. Perú es una nación pesquera que captura el 10% del volumen de todo lo que se pesca en el mundo y lo que principalmente captura es anchoveta: el 85% en promedio. Todo lo que pesca un país como Portugal en un año, puede pescarse en apenas semanas en un solo puerto de Perú. Se trataba de miles de actas de desembarques que debían obtenerse y revisarse para sacar a flote la verdad. Y en esas circunstancias, sólo tenía dos caminos: envejecer buscando y revisando los documentos año tras año (entregando mi vida a los pescados, el mar y sus tiburones) o procesar toda esa información con la ayuda de la computadora a partir de las bases de datos oficiales que contenían los registros de las descargas. Esto último era el mejor camino, no sólo porque hacía viable la investigación en términos de tiempo, sino porque permitía escapar de la información parcializada, y muchas veces falaz, que circulaba en la opinión pública sobre este lucrativo sector. La meta era acceder y procesar toda la torta, no solo una tajada direccionada.

Así ingresé al mundo de la “minería de datos”, el “periodismo asistido por computadora” o el “periodismo de datos”. Y así tuve que vencer mi fobia al Excel. Mi médico de cabecera para vencer este temor a lo desconocido fue la talentosa y premiada [Giannina Segnini](#), ex jefa de la unidad de investigación del diario La Nación de Costa Rica, hoy profesora de la Universidad de Columbia y la periodista latina pionera en el uso y análisis masivo de bases de datos para la investigación periodística. Giannina me asesoró en la investigación gracias a un programa de entrenamiento que realizó el Instituto Prensa y Sociedad (IPYS) en Lima. Con ella aprendí a reportear con brújula, con data, y comprobé lo que asegura con total conocimiento: “el periodismo basado en datos aumenta la credibilidad y la independencia. Por primera vez en la historia, el periodismo de investigación cuenta con herramientas poderosas para descubrir historias completas, no sólo las piezas que alguna fuente quiere revelar”.

El trabajo con bases de datos potencia la investigación periodística. Permite descubrir fenómenos, tendencias, que después hay que verificar con el reportero tradicional, en la calle, ese que nunca podrá ser reemplazado por ninguna computadora ni el programa más sofisticado. En la combinación del ejercicio fundamental de la verificación y las nuevas herramientas tecnológicas radica el actual desafío de un periodismo de investigación que dependa cada vez menos de las filtraciones de información de las fuentes.

¿Cómo fue el proceso?

Lo primero que tuve que hacer, por más básico que parezca, fue comprobar si el Estado tenía organizado en archivos de Excel (o en algún otro programa) la información de los desembarques de anchoveta de cada nave y conocer al detalle qué tipo de información contenían. Si no existía esa data organizada, la investigación no era viable. Como señala Mar Cabra, periodista española del Consorcio Internacional de Periodistas de Investigación (ICIJ), la clave no está necesariamente en las herramientas sino en tener una “[mentalidad de datos](#)” (data state of mind). “Es decir, pensar dónde puede haber datos y saber cómo hacer ese análisis sistemático”.

El trabajo con datos tiene diversas etapas (cada una con distinto grado de dificultad) que puede resumirse en lo siguiente: acceso, procesamiento, análisis y visualización de los resultados. En este caso, la primera gran dificultad estuvo en el acceso debido a que los registros de desembarque no estaban publicados para los ciudadanos en ninguna página web de las instituciones estatales. La data existía pero sólo podían revisarla de manera privilegiada las autoridades del Gobierno, las empresas fiscalizadas y las empresas inspectoras que a la vez eran consultoras de varias compañías del sector.

Por este motivo, solicitamos las bases de datos al Ministerio de la Producción vía la Ley de Transparencia y Acceso a la Información. No tuvimos éxito. Las autoridades negaron la información por considerar que los registros de desembarque por empresa estaban protegidos por el secreto estadístico a pesar que se trataba de la explotación del principal recurso del mar peruano. Esto obligó a buscar otras fuentes que tenían acceso legal a la data, mientras llevábamos al Poder Judicial el reclamo legítimo de obtener esta información de interés público. Conseguir aquellos archivos fue lo más difícil de todo. Hasta hoy, [el Ministerio no los entrega](#).

Mediante diversas fuentes, al inicio solo accedí a un grupo de las actas de inspección de los desembarques. Debido a que la información de los documentos estaba registrada a mano, fue imposible capturarla con un programa OCR (Reconocimiento óptico de caracteres). Esto me llevó a ingresar al Excel dato por dato. Después de varias semanas de insistencia, accedí a nuevas hojas de cálculo. Pero seguía teniendo algunas piezas, no todo el rompecabezas. Insistí y obtuve más archivos. Al final de tres meses de búsqueda, como sucede cuando uno chanca el fierro una y otra vez hasta que se dobla, accedí a la data completa y a las miles de actas de inspección impresas y digitalizadas en donde aparecían las cifras originales de las descargas. En total, más de 100 mil documentos.

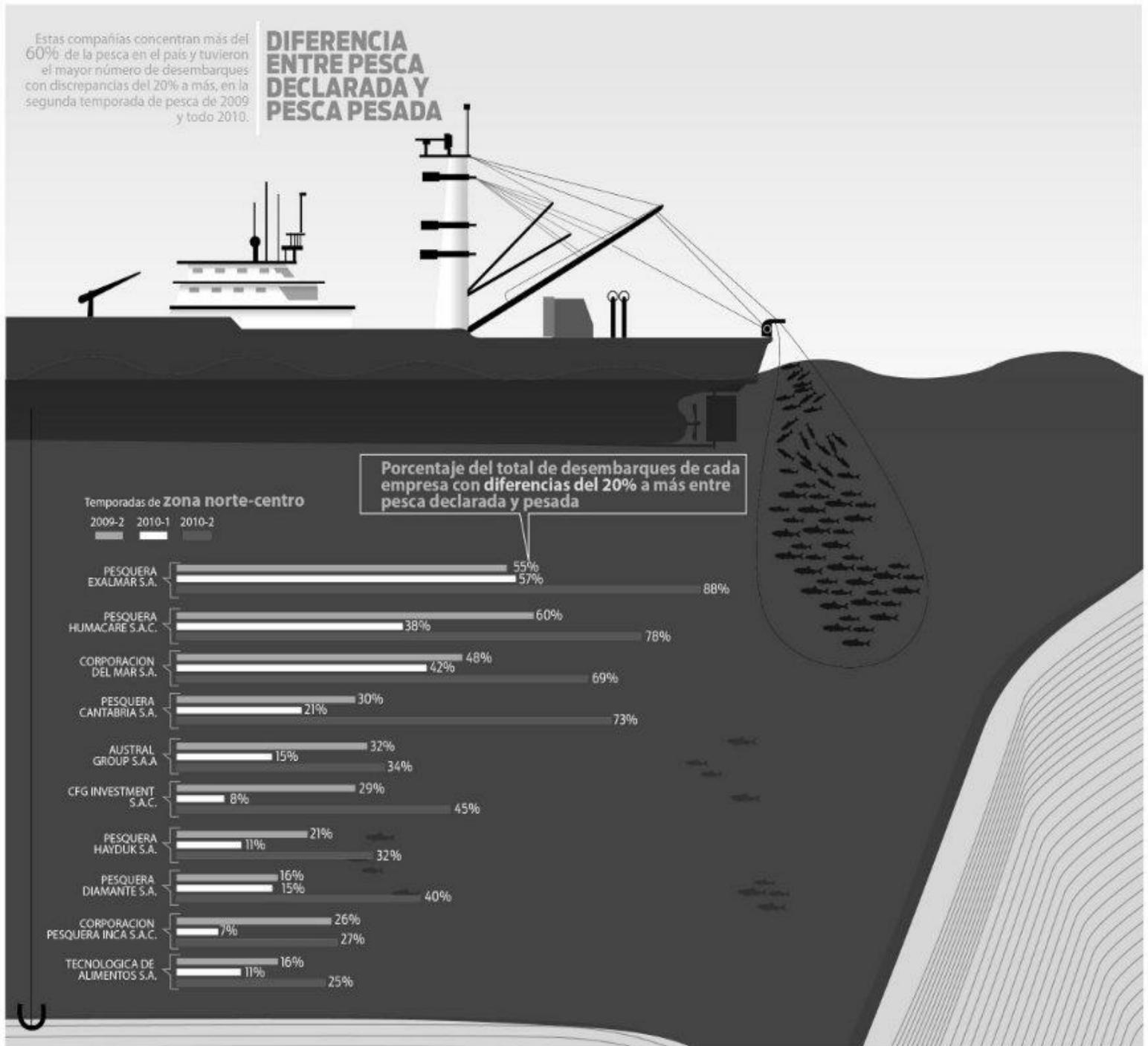
Lo que vino fue un procesamiento tedioso que me permitió entender después de varios desvelos que la data viene sucia y hay que detectar el patrón del error. Para descifrarlo, no solo se debe revisar los datos una y otra vez, analizarlos como quien se interroga a una persona para conocer sus fortalezas y debilidades, sino que es necesario tener suficiente conocimiento de cómo funciona el sector que se está investigando y cómo fue construida la data. Por ejemplo, en el proceso pude ver que había uno que otro desembarque con registro de pesca de más de mil toneladas cuando la nave anchovetera más grande del país tenía una capacidad de bodega de 900 toneladas en esos días. No era posible que un barco descargara más de esa cantidad. Saber eso, me permitió identificar errores de tipeo que se lograron corregir buceando en las actas de inspección originales. Después de comprobar la utilidad de este “mar de datos”, todo empezó a despejarse. Dejé de abrumarme con tantas cifras y fórmulas. Empecé a bucear.

El método sí importa

La metodología para analizar cualquier conjunto de datos debe tener un sustento técnico sólido y anclarse en la realidad. En esta investigación, como ya se dijo, comparé la cantidad de anchoveta que declaraban los patronos a los inspectores cuando llegaban a los muelles y la pesca pesada registrada en las balanzas de las fábricas harineras. Esto permitió descubrir que más de la mitad de los desembarques en los puertos del norte y centro del país, donde se descarga el 90% de la anchoveta capturada, tenían diferencias entre la pesca declarada y pesada que superaban el 10% y podían llegar

al 20%, 30% y hasta 50% en algunos casos. La balanza siempre registraba menos peso en cantidades sospechosas. Las diferencias evidenciaban un subregistro y en consecuencia una sobrepesca ilegal, debido a que la máxima discrepancia posible era del 10%, según las autoridades, expertos, empresarios, inspectores y pescadores consultados.

Las empresas que encabezaban las discrepancias eran las más poderosas. El análisis de los datos permitió saber quién era quién en la pesca industrial, desde la compañía más grande hasta la más pequeña.



En una primera etapa, analicé 45 mil desembarques que permitieron detectar un subreporte de 300 mil toneladas de anchoveta entre la primera temporada de pesca de 2009 y todo el 2010. Este volumen estaba valorizado en 100 millones de dólares si se transformaba en harina, sin contar lo que se dejaba de pagar al Estado en derechos de pesca y a los pescadores que ganan por tonelada pesada. Estas pérdidas también fueron calculadas y para ello se tuvo que desarrollar una metodología específica.

Después de la publicación de los primeros reportajes, se amplió el periodo de análisis de la data con el Consorcio Internacional de Periodistas de investigación (ICIJ) y recién entonces se abrió la posibilidad de trabajar en equipo. Se incluyó dos temporadas de pesca adicionales, una de ellas incluso tuvo que reconstruirse utilizando las cifras de centenares de actas de inspección con la ayuda de jóvenes ingenieros de sistemas peruanos encabezados por Miguel López, un programador-periodista nato. Con ellos, se desarrolló una forma de vincular las actas escaneadas con las cifras que eran incorporadas en las hojas de cálculo con el propósito de verificar cada dato nuevo. En forma simultánea, el ICIJ contrató un equipo en España para corroborar cada dato ingresado y analizado por el equipo peruano.

Para hacer periodismo de datos no se necesita trabajar en un gran medio de comunicación, ser parte de un equipo numeroso o usar un software sofisticado. Basta con un periodista y un programa básico como Excel como sucedió en la primera parte de esta investigación con IDL-Reporteros. Sin embargo, si se requiere construir y cruzar bases de datos cada vez más voluminosas y complejas, se necesita pensar en un equipo de investigación integrado no sólo por periodistas, sino también por ingenieros de sistemas.

Con la participación del ICIJ, se ratificaron los hallazgos iniciales de la investigación llegando a analizar más de 100 mil desembarques entre 2009 y parte de 2011, que permitieron concluir que después de la implementación de la ley de cuotas, se esfumaron entre las bodegas de los barcos y las balanzas, 630 mil toneladas de anchoveta, una cantidad que superaba toda la pesca que las flotas británicas llevaban a puerto en un año.



La metodología usada fue confiable por tres razones, a pesar que los empresarios involucrados intentaron desacreditarla: 1) El registro de la pesca declarada y pesada es oficial, y cada una tiene una ficha de inspección precisamente para poder comparar estas dos cifras como parte de la fiscalización de los desembarques que debe realizar el Estado. 2) Los empresarios pesqueros, los pescadores, los funcionarios del Estado, los expertos y los actores más importantes del sector fueron entrevistados antes de procesar la data para determinar a partir de qué porcentaje la diferencia entre pesca declarada y pesada era sospechosa. Todos ellos señalaron que como máximo podía haber 10% de discrepancia. Sin embargo, cuando luego volvimos a consultar a los empresarios con los resultados en la mano, éstos empezaron a sufrir de una repentina amnesia. No imaginaron que podíamos acceder a los datos y

mucho menos descifrarlos. Ningún periodista realizó antes alguna investigación parecida en el Perú. 3) El Ministerio de la Producción mandó a elaborar en 2009 a una de las empresas inspectoras, CERPER, un análisis de la pesca declarada y pesada debido a las denuncias reiteradas de los pescadores de “robo en la balanza”. Su conclusión era que las discrepancias detectadas en estos dos registros, revelaban indicios razonables de manipulación en el software de las balanzas. Esto fue determinante para realizar una auditoría que concluyó que el 31% de las balanzas inspeccionadas tenían indicios graves de [adulteración](#).

El reporte de cada día

Para corroborar el fenómeno del subreporte que evidenciaba la data, fue necesario entrevistar a autoridades, inspectores, auditores, empresarios, ex superintendentes de plantas, calibradores de balanzas, programadores de sistemas, entre otros. Fue clave conversar con el creador del primer programa informático que fue instalado en los tableros de control de las balanzas porque contó cómo podía alterarse el peso de las descargas en un puerto de norte del país, incluso desde la oficina de un gerente de una pesquera en Lima.

Después de más de seis meses de investigación, quedaba claro que no se trataban de hechos aislados y que no solo incurrían en la pesca negra los pescadores artesanales y de menor escala. La responsabilidad de la mayor parte del subregistro recaía sobre todo en la gran industria. Y aunque la investigación fue encapsulada al inicio por los medios tradicionales en Perú, se logró de manera paulatina con la publicación de reportaje tras reportaje, que se dieran cambios en la inspección de las descargas y que la industria pesquera empezara a ser vigilada por la opinión pública. La alianza con el ICIJ también fue importante porque la investigación ampliada sobre el “Saqueo de los Mares” en el Pacífico Sur, permitió que los hallazgos fueran recogidos por importantes medios internacionales: Le Monde de Francia, El Mundo de España, Internacional Herald Tribune, [entre otros](#). La presión también vino desde afuera.

Después de la publicación de la serie investigativa, las autoridades realizaron auditorías a las balanzas en cada temporada de pesca, se evaluó la compra de un nuevo software de pesaje para evitar la manipulación y el Estado contrató a inspectores propios para no depender de las empresas supervisoras financiadas por las compañías pesqueras y que, además, eran contratadas para certificar la harina de pescado que exportaban las empresas fiscalizadas.

La clave de este proceso investigativo fue la perseverancia y vencer el temor a lo desconocido, al “mostrito” como le llamaba Giannina Segnini al servidor que alimentaba día a día con bases de datos en su Unidad del diario La Nación para producir información nueva. La investigación de la pesca peruana responde a las nuevas tendencias del periodismo de investigación que se están desarrollando con grandes resultados en varios medios del mundo en tiempos en que hay una apertura cada vez mayor de datos en el ciberespacio.

Un nuevo camino

Ingresar a este universo de datos es posible con nociones básicas del Excel, una computadora y una voluntad férrea. Por eso, si Giannina Segnini y el ingeniero de sistemas principal de su ex equipo de La Nación, Rigoberto Carvajal, tuvieron la generosidad de mostrarme los beneficios de estas herramientas a pesar de mi rechazo a la informática, me tocaba reproducir este gesto con mis alumnos de periodismo en las universidades de Lima donde enseñé y en los talleres que dicté a nivel nacional a periodistas, estudiantes y profesores. Más de la mitad de los proyectos de investigación de mis alumnos más jóvenes incluye el trabajo con datos y el uso de apenas una computadora portátil. Los resultados son sorprendentes. Hoy, seducida totalmente por estas posibilidades, impulso un proyecto periodístico independiente en Perú que permita integrar, en esta primera etapa, el trabajo con

periodistas de investigación, ingenieros de sistemas, desarrolladores web y diseñadores multimedia, con miras a generar alianzas con apuestas similares en América Latina y el resto del mundo.

Investigar, emprender, experimentar e innovar en equipo es el camino que hemos decidido recorrer varios de los que amamos este oficio. La mejor arma siempre será querer saber y brindarle lo mejor al público. Richard Gingras, jefe de noticias de Google asegura que en estos tiempos en que los medios “ya no pueden decir: créannos porque somos quienes somos (...), el reportaje de investigación del mañana será el resultado de un persistente trabajo con minería de datos y cruce de información”. [Y esto no solo implica un cambio tecnológico, sino cultural.](#)

Dar cuenta al público cómo llegamos a tal o cual conclusión y responder con la misma transparencia que exigimos los periodistas a las autoridades y a las grandes corporaciones privadas, es ya un mandato. Al respecto, Giannina logra dar una vez más en el clavo: “cada vez que los periodistas ponemos en manos de la audiencia un conjunto completo de datos, ésta puede ver la costura de nuestras historias, las decisiones que tomamos y la jerarquía de los conceptos que aplicamos”. Eso permitirá conseguir lo que está en el fondo de todo lo antes dicho: cultivar un periodismo comprometido con la Humanidad.

Milagros Salazar Herrera

Periodista de investigación y docente universitaria. Trabaja para IDL-Reporteros, en Lima, es corresponsal de la agencia de noticias Inter Press Service (IPS) y miembro del Consorcio Internacional de Periodistas de Investigación, con sede en Washington. Es profesora de la Universidad de Lima y de la Universidad Ruiz de Montoya, donde enseña periodismo de investigación y de datos. Actualmente, impulsa un proyecto periodístico independiente en Perú que integra el trabajo de reporteros de investigación, ingenieros de sistemas, desarrolladores web y diseñadores, con miras a generar alianzas con otras iniciativas en otros países del mundo.

Qué es el *scraping* y cómo hacerlo bien

Rigoberto Carvajal

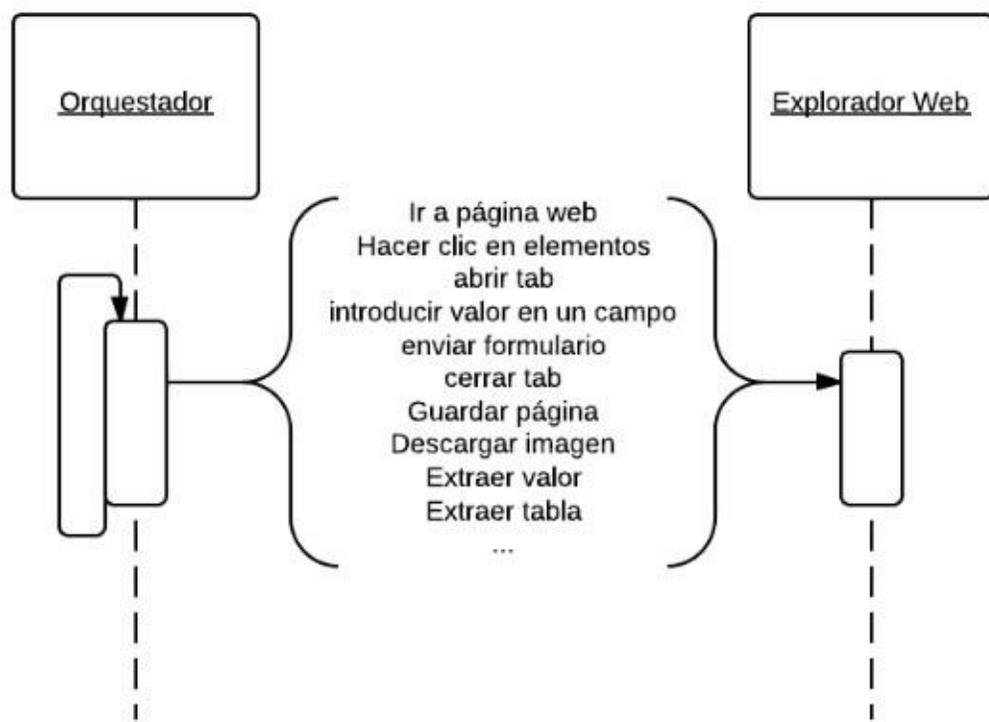
Se conoce como *scraping* al arte de controlar de manera automatizada sitios web con el propósito de extraer datos publicados que son de nuestro interés. Este artículo es para personas familiarizadas con el *scraping* y en él encontrarán una serie de consejos y herramientas, trucos y diseños muy útiles cuando se trata de obtener muchos datos.

Esquema orquestador + browser

El mejor modelo para hacer *scraping* es en el que existe un programa “orquestador” que puede estar en cualquier lenguaje de programación y que controla a un explorador web. Es muy importante esto porque aunque uno pueda programar solicitudes http y parsear un html, hay sitios donde la interacción tiene que ser muy parecida a la que tiene un usuario con el navegador, sobre todo por la presencia de elementos AJAX y manipulación con Javascript, donde los elementos se consultan en la base de datos y se pintan sin que haya un cambio de página o refrescamiento. Objetos flash y applets de Java también pueden complicar la creación de *scrapers*, o la manera como están programados sitios en .Net, que guardan el estado de los objetos en la sesión web.

Orquestador: Programa en lenguaje de alto nivel con el que se puede aprovechar muchos recursos de la computadora y que envía comandos al explorador web para guiar la navegación deseada.

Explorador web: Es como una terminal “tonta”, no metemos lógica aquí, sino que es el componente capaz de manejar todos los elementos de una página web y sigue las instrucciones que el orquestador le da por medio de comandos.



Identificación de elementos

Cuando estamos procesando consultas en un motor, es muy importante que logremos identificar elementos que nos indiquen el estado final de la consulta, según sea exitosa, sin resultados o si hubo un error, si no hacemos esto podríamos estar tratando de aplicar acciones sobre los resultados sin que el sitio nos haya devuelto la respuesta completa todavía.

Continuar proceso interrumpido

Muchas razones pueden interrumpir un proceso de scraping, por ejemplo: una desconexión de internet, reinicio voluntario u obligatorio de la computadora, o que un error no contemplado bote la aplicación, etc. No hay que comenzar desde cero a *scrapear* todo el sitio. Se debería llevar un registro en una base de datos con los registros procesados, de manera que cuando se vuelva a ejecutar la aplicación pueda continuar donde quedó, sin depender de nosotros.

Observación de proceso

A veces podemos programar *scrapers* que no usen un explorador web visible sino que manejan las consultas y la lógica de extracción de datos sólo mediante código, lo que pasa con estos casos es que son más propensos a no capturar todos los datos, porque siempre aparecen casos y cosas que no habíamos considerado en la programación, por ejemplo es muy frecuente que algunos registros desplieguen un dato y hasta toda una pestaña de datos pero en otros no. Por lo que es una buena práctica observar la mayor cantidad de ejemplos posibles de los registros que vayamos a *scrapear* para estar más seguros que estamos considerando todos los escenarios posibles.

Captchas

A veces creemos que cuando un sitio tiene *captcha* es imposible de *scrapear*, pero no es así, actualmente existen librerías y muchos servicios web que resuelven los captchas y devuelven el texto que contienen las imágenes, y pueden ofrecer este servicio porque han desarrollado software para el reconocimiento de imágenes en combinación con inteligencia artificial o con personas, las cuales brindan la respuesta correcta cuando el sistema se equivoca y con esto el sistema sigue aprendiendo, manejando un rango entre 1 y 15 segundos la resolución del captcha.

Uno muy popular se llama [DeathByCaptcha](#), tiene un excelente API y actualmente maneja un 92% de certeza y un tiempo promedio de 8 segundos en la respuesta.

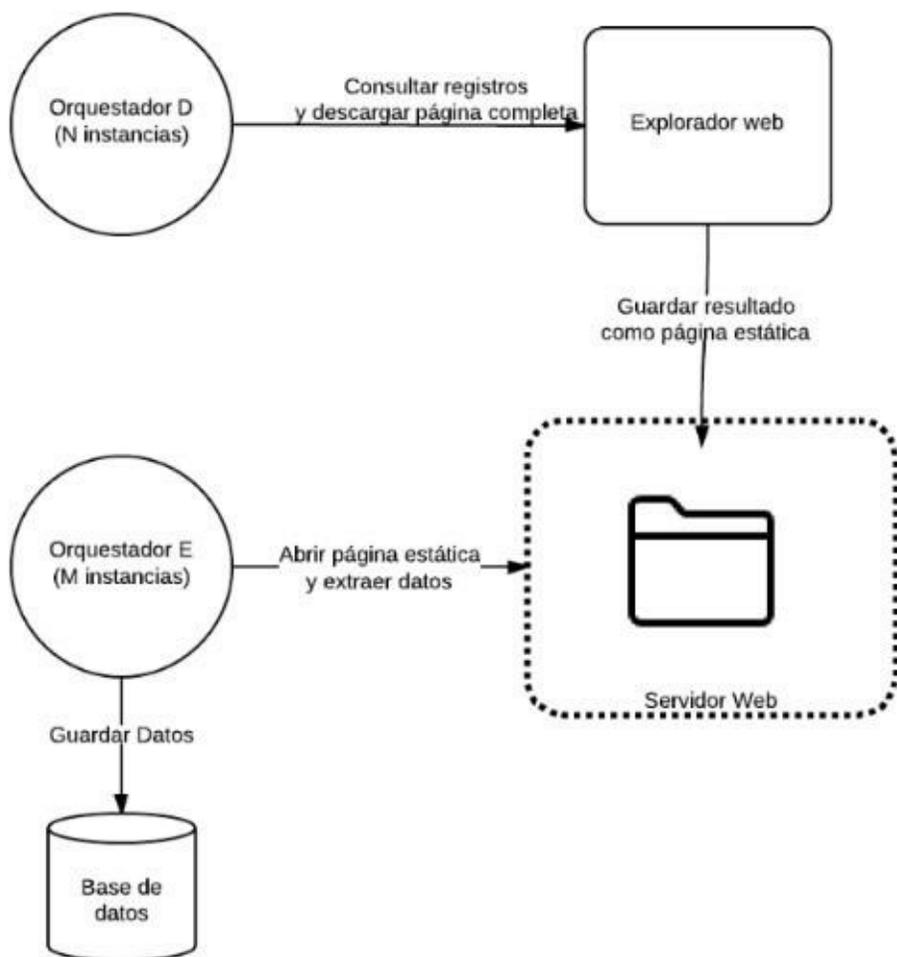
División de trabajo y procesamiento en paralelo

Seguramente conocerás la frase “Divide y vencerás”, pues esta es una de las frases más aplicadas en la programación, así mismo al *scraping*. Si ya hemos experimentado con esto sabemos que *scrapear* un sitio puede tomar bastante tiempo de ejecución, la clave para mejorar esos tiempos está en dividir etapas del trabajo en diferentes programas y que trabajen de manera paralela, entendiéndose esto como varias instancias de los mismos programas ejecutándose al mismo tiempo en una o varias computadoras.

La manera más sencilla de *scrapear* es repitiendo la secuencia:

- Consulto el registro X
- Extraigo los datos
- Los guardo en un archivo.

Esto funciona, pero no tanto cuando lo que queremos es *scrapear* más de 20.000 registros, a menos que estemos dispuestos a esperar todo ese tiempo. El siguiente esquema muestra un patrón de diseño que funciona muy bien para conjuntos de datos grandes y cuando los datos lo permitan:



En este modelo se divide el trabajo en 2 partes:

- Consulta y guardado de página completa: el robot orquestador D se encarga de hacer las consultas en los formularios web y cuando obtiene los resultados los guarda como una página web completa en un folder de documentos que sirve un servidor web.
- Extracción de datos y almacenamiento en base de datos: como las páginas web se están guardando en un servidor web, cualquier explorador web podrá visualizarlas como si fuera cualquier otro sitio, por lo que entonces aplicaremos nuevamente scraping a esas páginas que ahora correrán de manera local, con lo que se abrirán de inmediato y extraer sus datos va a ser más rápido.

La lógica de este diseño radica en los siguientes puntos:

- La extracción de datos es más eficiente trabajando con páginas locales que en Internet porque el programa no tiene que esperar por los resultados de una consulta ni repetir intentos para leer cada uno de los campos a extraer.
- Cuando descargo una página puedo identificar cual es el campo que se “pinta” en el browser como señal de éxito o de falta de resultados y sólo esperar en la programación por éste único elemento para dar la instrucción de guardar la página.
- Si se cuenta con buen ancho de banda y varias computadoras aunque tengan poca capacidad de procesamiento este modelo es el más adecuado porque puedo tener a una(s) descargando y otra(s) extrayendo datos, y entre todas esas instancias el trabajo se hará en menor tiempo.

- Si decido primero solo descargar puedo observar todos los posibles escenarios de resultados, por ejemplo: campos que solo se visualizan para ciertos valores y para otros no, podría identificar cuáles registros llevaron a una pantalla de error o de ausencia de datos sólo con el tamaño de las páginas HTML descargadas y descartarlos de una vez para no procesarlos.

La cantidad de descargadores debe ser congruente con nuestro ancho de banda; si usamos pocos podemos estar subutilizando los recursos, y si usamos demasiados habrá lucha por el recurso y tendremos a descargadores en “espera”. Hay que encontrar el balance, según el tamaño de los recursos que estemos descargando, el ancho de banda y la capacidad de procesamiento.

Para aplicar este modelo es indispensable utilizar un motor de base de datos multiusuario (como mssql, mysql o postgresql), pues éstos están hechos para controlar correctamente la concurrencia, dado que si pretendemos usar un archivo como un CSV o un archivo Excel habrá conflicto a la hora de que dos o más programas intenten escribir al mismo tiempo.

Segmentación

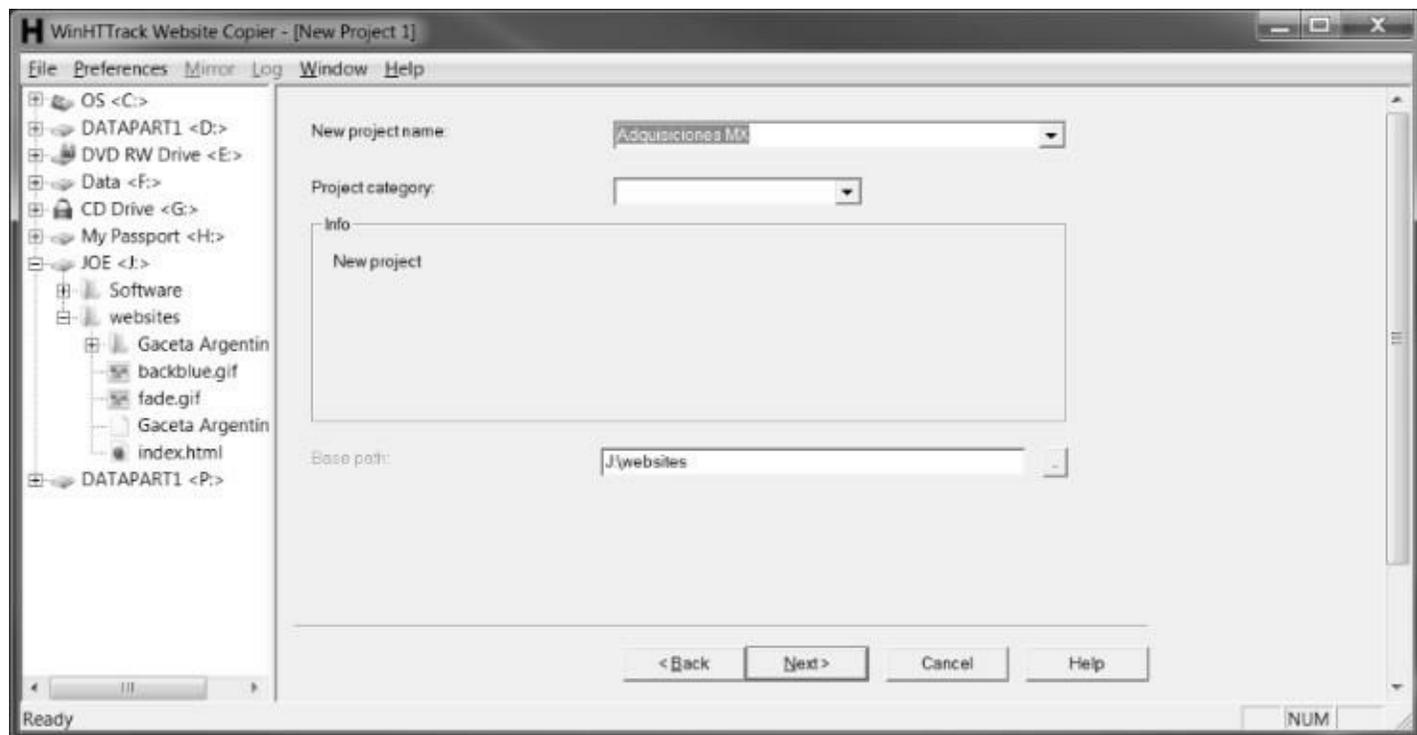
Para poder sacar provecho máximo del paralelismo es importante que repartamos la carga de trabajo entre varios “trabajadores”, sean estas instancias del programa en la misma computadora o programas corriendo en varias computadoras. Para esto podemos aplicar estrategias conocidas como la segmentación por rango y la segmentación por fórmulas.

Descargar un sitio entero

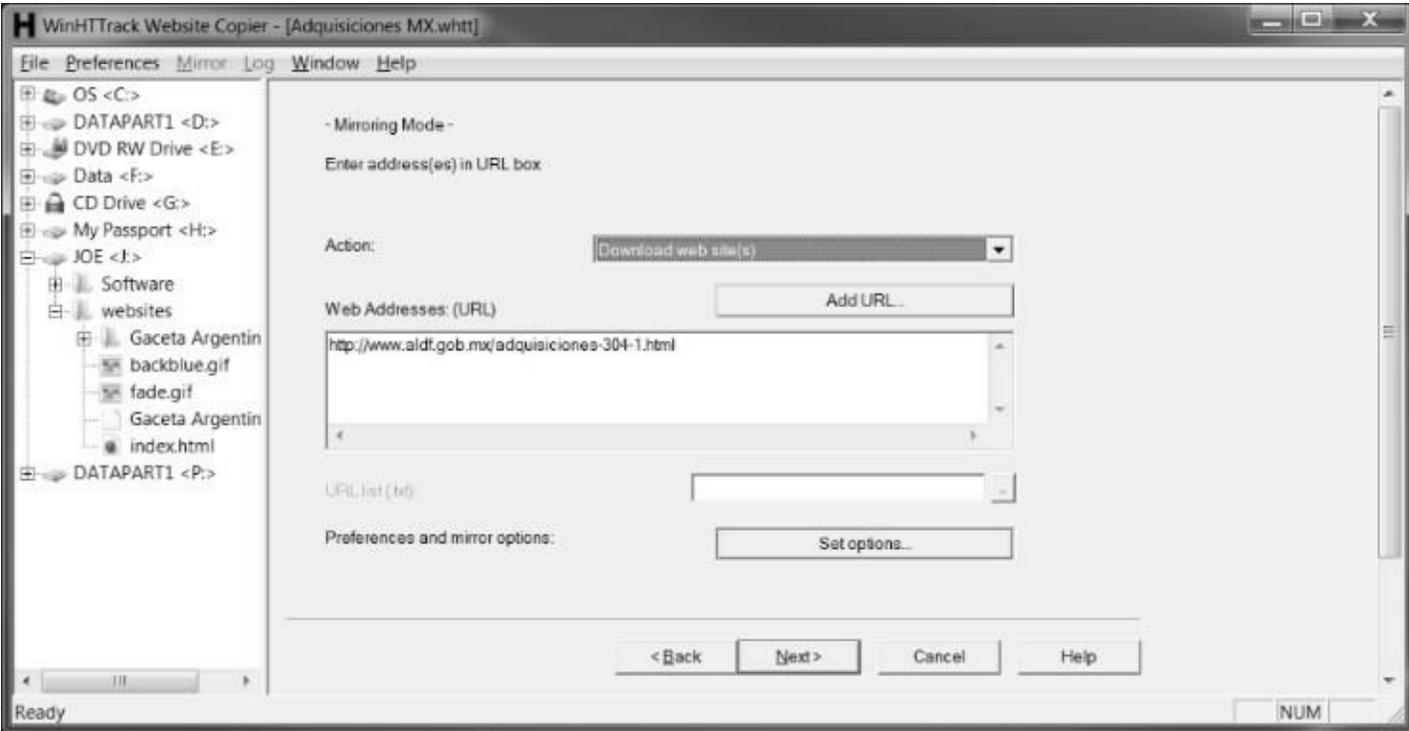
Una herramienta útil y fácil de utilizar es HTTrack, el cual se define como un copiador de sitios web, también sirve para este propósito el comando [wget de UNIX/Linux](#). Está disponible en varios sistemas operativos y acompañado de algunas opciones y parámetros nos permite descargar todos los elementos estáticos del sitio y los modifica para que funcione localmente. Es importante aclarar que no trae aquellos objetos dinámicos del sitio a menos que sean accesibles desde links dentro del mismo sitio.

Este es un ejemplo usando HTTrack con la página de adquisiciones de la Asamblea Legislativa del Distrito Federal (México). Podemos descargar todas las páginas y documentos de ese sitio.

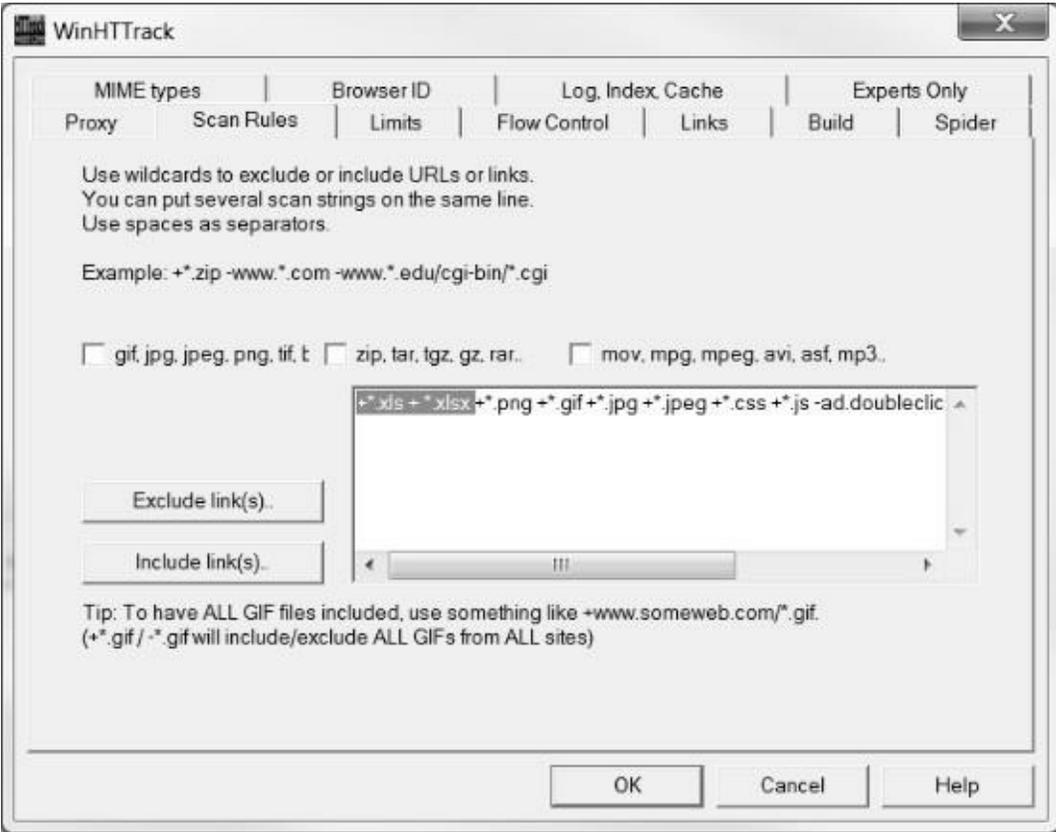
Crear un proyecto y definir la carpeta donde se va a descargar todo:



Definir las direcciones web para descargar:



Si hacemos clic en el botón de opciones se nos abrirá una pantalla donde podemos configurar nuestra descarga, por ejemplo definir los tipos de archivos que queremos incluir o excluir, de esta manera si sólo estamos interesados en descargar los documentos PDF de un sitio, podemos agregar una regla de la forma `+*.pdf`

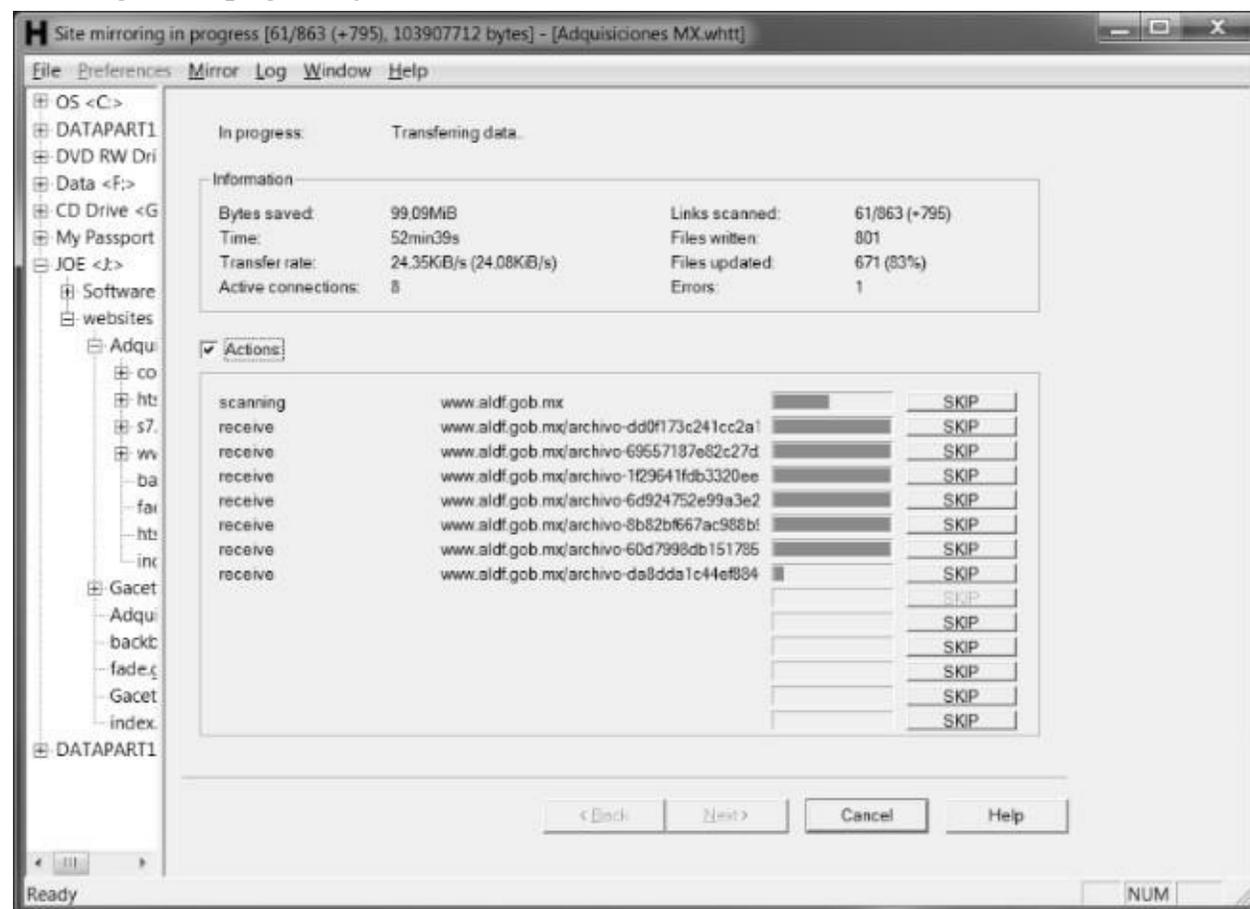


Por ejemplo yo he agregado para este ejemplo que descargue los archivos de Excel también `+*.xls +*.xlsx`

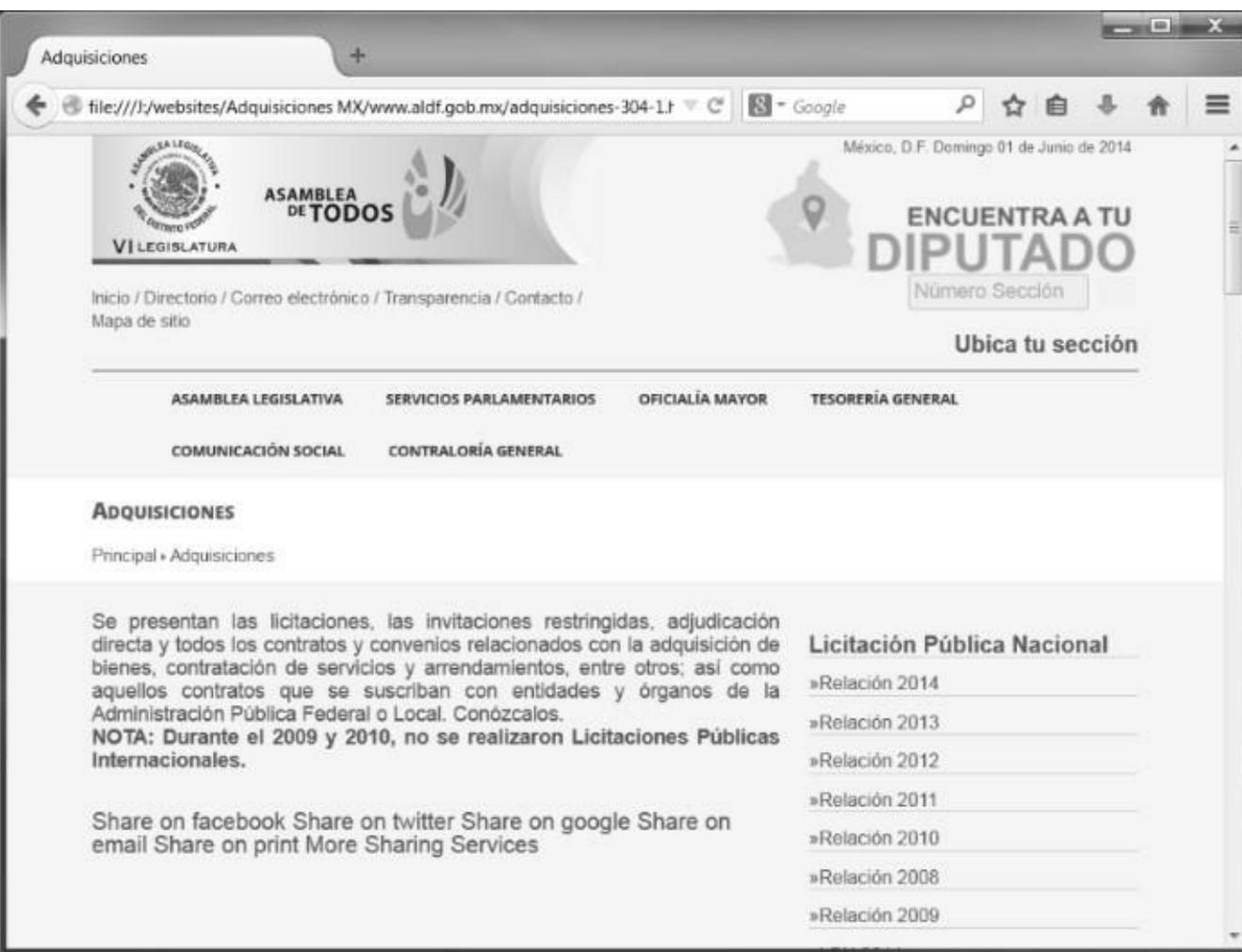
La aplicación “navegará” por todas las páginas del sitio y se meterá en todos los links de cada página de manera recursiva, y mantendrá una lista de links descargándose, el tamaño por defecto de esa lista es de 4 conexiones, pero si cuenta con buen ancho de banda y capacidad de procesamiento puede aumentar ese valor en la pestaña de “Control de Flujo en la opción: Número de conexiones”.

Si desea entender cada una de las opciones puede dirigirse al manual del sitio, el cual cuenta con un [paso a paso](#) muy útil.

Habiendo configurado todo podemos dar clic a finalizar e inmediatamente veremos la aplicación descargar las páginas y archivos web.



Cuando todo este proceso haya finalizado tendrás una copia del momento del sitio web funcional corriendo desde tu computadora.



Recomendación final

Existen muchas herramientas hoy para hacer *webscraping* y todas las puedes tener en tu caja de herramientas, porque todas pueden funcionar para diferentes casos, pero es importante que consideres estos consejos y trucos para escoger las herramientas adecuadas en un proyecto de *scraping*.

Finalmente comparto mi caja principal de herramientas para web scraping :

- [Talend Open Studio for Data Integration/ Big Data](#)
- [IMacros Enterprise Edition](#)
- [Microsoft Visual Studio 2010+ \(Entity Framework\)](#)
- [Microsoft SQL Server 2008 o superior](#)
- [Wget | HTTrack](#)
- [WAMP Server](#)
- JAVA, PHP y C# (Lenguajes de programación)

Rigoberto Carvajal (@rcarvajal85)

Ingeniero en Computación graduado del Instituto Tecnológico de Costa Rica, con más de 7 años de experiencia en producción de software y administración de datos. Es experto en data del International Consortium of Investigative Journalists. Trabajó como Jefe Analista de Datos de la unidad de Investigación en La Nación de Costa Rica. Está a cargo de tareas de administración de bases de datos, procesos de extracción, transformación y carga; análisis y visualización de información, ha sido un

caso exitoso de integración entre periodistas e ingenieros trabajando a diario en conjunto en el proceso de generación de noticias de interés público basadas en datos.

DocumentoMedia: reportajes a partir de documentos desclasificados

Carlos Basso

El 2013 se cumplieron 40 años del golpe de Estado perpetrado en Chile por Augusto Pinochet, sin que aún hayan cicatrizado las heridas que éste dejó en los chilenos.

En dicho contexto, en 2012, junto a un colega de la Escuela de Periodismo de la Universidad de Concepción, comenzamos a trabajar en el proyecto [DocumentoMedia](#), un centro de investigación periodística sin fines de lucro que utiliza como material básico para la construcción de sus reportajes documentación desclasificada y pública.

En función de ello hemos elaborado a la fecha cerca de 40 reportajes, así como un [libro](#) construido a partir de varios de ellos ([“La CIA en Chile”, Aguilar, 2013](#)), todo elaborado en su mayor parte gracias a las colecciones de documentos desclasificados a partir de 1999 por el Gobierno de William Clinton, las que hoy en día (con las sucesivas y posteriores desclasificaciones) abarcan cerca de 24 mil documentos sobre la historia chilena, desde 1950 hasta entrada la década de los noventa, y que corresponden a informaciones emanadas de los Departamentos de Estado y Defensa, así como de organismos como el National Security Council, la CIA y el FBI.

Dicha información se encuentra online, en servidores del Gobierno de Estados Unidos, y muy pocos medios la han trabajado. De tanto en tanto alguien publica algo particular al respecto, pero nadie se había dado el trabajo de revisar dicha documentación y menos de sistematizarla y ello no es tan complejo, pues todos los PDF escaneados cuentan con la metadata necesaria para –por ejemplo– agruparlos temáticamente.

Pese a la importancia que poseen estos documentos y el enorme volumen de los mismos (sólo el libro de la CIA está construido sobre unos 900 informes de esa y otras agencias) nadie ha sido capaz de resguardarla -ante una eventual caída de los sistemas, por ejemplo- ni mucho menos ordenarla y explorarla en su totalidad, pese a que es una riquísima fuente de información.

Lo mismo ocurre con la documentación -escasa, pero relevante- liberada por organismos británicos y alemanes, que alude a Chile, y con los documentos relativos a la operación Cóndor y a decenas de casos de detenidos desaparecidos chilenos que se encuentran en el llamado “Archivo del horror”, en Paraguay.

Al mismo tiempo, hay mucha otra documentación que se ha ido generando a partir de otras fuentes. Wikileaks, por ejemplo, ha filtrado muchos documentos (4.500) sobre la historia chilena más reciente, y en diversos sitios web se han ido publicando fallos en materia de Derechos Humanos del pasado reciente del país y del presente, pero no existe ningún sitio que preserve todo este material junto, ni mucho menos, que lo procese tanto periodísticamente como a nivel de metadata, a fin de que cualquier ciudadano pueda acceder a él e informarse.

En este sentido, nosotros nos apartamos un poco de la conceptualización tradicional del periodismo de investigación, aquella de buscar y publicar algo que alguien con poder ha mantenido oculto, pues debido a la enorme cantidad de datos que hoy existe en la red, en realidad ya no es necesario ocultar algo, pues es tal la cantidad de información, que esta permanece oculta a plena vista. Vaya contradicción.

Hay mucho que se puede recopilar, organizar y sistematizar de modo de generar una enorme base de datos abierta a todos, que sea la base de lo que pretendemos sea DocumentoMedia:

- Un centro de investigación periodística sin fines de lucro, basado en documentación, que

produzca reportajes de alto impacto y calidad, de libre disposición. Varios de los reportajes que hemos producido han sido reproducidos por importantes medios chilenos y esperamos que otros se sumen a ello.

- Un repositorio abierto y sistematizado de documentación chilena, con especial énfasis en Derechos Humanos.
- Un centro de periodismo de datos. Estimamos que DocumentoMedia es un proyecto que es perfecto para crear visualizaciones a partir de los datos que contendrá. En términos generales, nuestro proyecto aspira a crear un cambio positivo para Chile y el periodismo de investigación en general, demostrando que el periodismo de datos no sólo es el uso de números y hojas de cálculo, sino mucho más que eso: pretendemos ser el primer centro de investigación en generar data a partir de documentación. El concepto de DocumentoMedia se resume en su nombre, que en síntesis nos remite a crear medios a partir de documentación, combinando eso con periodismo de datos y con las viejas e imprescindibles técnicas de reporte.

De este modo, nuestra planificación de actividades considera, a futuro, la bajada y clasificación de toda la documentación que podamos, dentro de nuestras capacidades, creando además un app que nos permita extraer con mayor precisión los metadatos que se requieren para este proyecto.

Carlos Basso (@cbassop)

Periodista de la U. de Concepción, profesor asistente del Departamento de Comunicación Social de la misma, profesor de Periodismo de Investigación y Dirección y Supervisión Periodística. Autor de seis libros de investigación periodística y dos novelas. Actualmente cursa el Doctorado en Literatura Latinoamericana de la UdeC también.

Entrevista con una base de datos

Hassel Fallas

Una base de datos es como cualquier otra fuente a la que diariamente nos enfrentamos los periodistas. Es propensa a contarnos mentiras, escondernos información, darnos un panorama parcial de un fenómeno e inducir a equivocaciones.

Aunque esperamos mucho de los números, lo cierto es que son falibles y no poseen la verdad absoluta porque, simple y sencillamente, las bases de datos están confeccionadas por personas. En ellas puede haber errores involuntarios o deliberados y siempre es recomendable tenerlo presente antes de utilizarlas como cimiento para una investigación periodística.

Esa es la razón por la cual es fundamental aplicarle a cualquier banco de números la misma rigurosidad que a nuestras fuentes humanas o documentales: validar su contenido, verificando su autenticidad con terceros.

El reporte convencional no solo es ineludible sino una obligación en el periodismo de datos para evitar la publicación de conclusiones erróneas, que podrían terminar siendo una tragedia para la carrera del periodista y hasta de su medio.

La preparación

Teniendo claridad sobre esos puntos, lo siguiente por hacer previo a entrevistar una base de datos es conocerla a fondo. Indagar exhaustivamente sobre:

- ¿Quién recopiló los números y cuáles fueron sus propósitos?
- ¿Cómo fue la metodología de recolección?
- ¿Qué tan confiable es la persona o institución que la construyó?
- ¿El documento está completo, se excluyó alguna información o se trata de uno que sólo contiene algunos cruces de datos básicos para intentar satisfacer los requerimientos del periodista?
- ¿Qué intereses persigue la persona que entrega la data?
- O por el contrario, ¿qué puede estar tratando de ocultar la persona o institución que se niega a proporcionarla?
- ¿Ese registro de cifras provee toda la información necesaria para iniciar un proyecto, se debe buscar otras bases o incluso crear las propias?

Obsesiónese

Contestadas esas preguntas y si la mayoría de las respuestas son satisfactorias para el periodista, lo siguiente es pasar mucho tiempo examinando –a veces hasta la obsesión- a esas hojas de cálculo en Excel, Tableau, SPSS, SQL o cualquier programa informático que se utilice para el análisis de la información. Personalmente, recurro a los dos primeros porque he comprobado que, hasta ahora, sobran y bastan en la ejecución de la mayoría de los estudios de periodismo de datos que he realizado.

Sólo si invierte una buena cantidad de horas comprendiendo la estructura de una base de datos será posible entrevistarla adecuadamente y extraer las conclusiones significativas y jugosas que se convertirán en los pilares de un proyecto exitoso.

Hacerlo es vital para percatarse de inconsistencias como errores de digitación en las cifras y nombres repetidos o escritos de diferente manera aunque pertenezcan a una misma entidad. Esos son descuidos que derivarán en cálculos subestimados y alterarán los resultados de la investigación.

Un ejemplo de eso nos ocurrió durante un proyecto sobre [exclusión estudiantil](#) (deserción) en

secundarias públicas que efectuamos en la Unidad de Inteligencia de Datos de La Nación de Costa Rica. Cuando mi colega Amy Ross y yo inspeccionábamos la base de datos con la información de 643 colegios, uno de ellos resaltó como el que más disminuyó el fenómeno en todo el país.

Los números consignados en el registro oficial del Ministerio de Educación Pública decían que en esa institución la fuga de alumnos había pasado de un 68% de la matrícula en 2011 a 14% en 2013. Es decir el problema decreció en 53 puntos. El cambio era tan extremo que despertó sospechas. Cuando conversé con el director de ese centro educativo para contrastar los datos absolutos y relativos, él revisó sus expedientes y me confirmó la mala digitación de la cifra de abandono estudiantil del año pasado; la real alcanzó al 50% de sus alumnos.

Otro beneficio de explorar una base de datos a profundidad es advertir faltantes de números.

Una vez más en el proyecto de deserción colegial nos llamó la atención que en una de las instituciones grandes (más de 1.000 alumnos), la exclusión de estudiantes había pasado de 445 en 2012 a nada en 2013.

Evidentemente allí faltaba un dato. En efecto, el Ministerio de Educación Pública nos corroboró que “[por un error involuntario](#)” no se incluyó en esa celda a los 694 estudiantes que abandonaron el centro de enseñanza en 2013. Ese número era significativo; sin él habríamos pasado por alto que esa secundaria es una donde es más problemático el abandono escolar.

Es crítico ser meticuloso con esos detalles. Solo imagine lo que podría ocurrir si, usando el registro de cifras de criminalidad de su país, no advierte la ausencia de números de robos, asaltos o asesinatos en municipios clave. Todo su trabajo se iría a la basura porque arribaría a conclusiones falsas.

Interróguela

Una vez terminado el examen profundo a la base de datos sabrá con precisión si esta es capaz o no de resolverle, parcial o totalmente, las incógnitas que tiene sobre el tema a indagar. Conviene, entonces, listar las preguntas a las que buscará respuesta cuando analice el documento en Excel con la ayuda de filtros y tablas pivote.

Si no sabe cómo usar esas herramientas, le recomiendo estos tutoriales del [Centro para el Periodismo de Investigación](#) y del [Consortio Internacional de Periodistas de Investigación](#).

Supongamos que la base en cuestión es la de criminalidad que mencioné arriba. Yendo de lo general a lo específico, algunas consultas básicas que puede incluir esa entrevista a los datos son:

- ¿Cuál es la cantidad de crímenes totales que hubo en el país durante el año o años para los que cuenta con cifras?
- ¿Ha aumentado o descendido la criminalidad?
- ¿Cuáles son los tipos de crímenes más comunes y su frecuencia por año?, ¿han subido o decrecido?
- ¿Cuál es el municipio donde más ha incrementado la criminalidad, en general y por tipo de incidente?
- Por el contrario, ¿cuál es el municipio donde bajaron los indicadores de crimen?

Recuerde siempre para casos como este o de incidencia de enfermedades, calcular las tasas por cada 10 mil o 100 mil habitantes. Es la manera más fehaciente de corroborar si un fenómeno se ha disparado o aplacado en el tiempo. Para ello, debe tener el dato de la cantidad total de la población del país o de cada municipio para los años de interés.

Supongamos que en su jurisdicción hubo 40 delitos graves el año pasado y la cantidad de habitantes

total es de 50 mil La fórmula en Excel para calcular la tasa es la siguiente:

D2				f_x	$= (B2 * 10000) / C2$
	A	B	C	D	
1	Año	Crímenes totales	Población	Tasa por 10.000 hab.	
2	2013	40	50.000	8	

Utilizando ese ejemplo se concluye que hubo 8 crímenes por cada 10 mil habitantes durante 2013. Cabe preguntarse: ¿esa cantidad es más o menos que en 2004 cuando se registraron 25 crímenes en total?

Si la población del 2004 en la ciudad era de 30 mil personas, siguiendo la fórmula anterior, concluiríamos que la tasa de criminalidad se ha mantenido en 8 crímenes por cada 10 mil habitantes.

D3				f_x	$= (B3 * 10000) / C3$
	A	B	C	D	
1	Año	Crímenes totales	Población	Tasa por 10.000 hab.	
2	2013	40	50.000	8	
3	2004	25	30000	8	

Con ese dato comparativo usted podría plantearse:

- ¿Cuál fue el comportamiento de la criminalidad en mi municipio para cada uno de los años correspondiente a 2005 y hasta 2012?
- ¿Se mantuvo la tasa cerca de los 8 actos violentos por cada 10 mil habitantes o hubo variaciones entre años?
- ¿Esos cambios fueron abruptos o no?
- Si fueron abruptos los cambios: ¿por qué el combate a la criminalidad es un sube y baja de un año a otro?
- Si la tasa es estable, ¿por qué se mantiene así?
- ¿Cuántos policías por cada 10 mil habitantes hay en la ciudad?
- ¿Cuál es el presupuesto que las autoridades invierten anualmente en seguridad?
- ¿Es alta o baja la tasa de criminalidad de mi municipio respecto de los otros de la provincia o del país?

Como puede observar, una base de datos puede y debe ser entrevistada en varias ocasiones durante la investigación; igual que ocurre con cualquier otra fuente. Además, a menudo, algunas de las respuestas que le dará desencadenarán nuevas preguntas cuyas respuestas estarán inmersas en otras bases de datos o recurriendo a documentos y voceros oficiales.

Finalmente, nunca olvide reflexionar sobre la más crucial de las preguntas por hacer a una base de datos: ¿por qué es importante para la gente la historia que cuentan sus cifras?

Un periodista puede tener el mejor análisis de data y las mejores conclusiones, pero si olvida mostrar el lado humano de los números, su reportaje carecerá de significado.

Hassel Fallas (@HasselFallas)

Trabaja en la Unidad de Inteligencia de Datos del diario La Nación, Costa Rica. Ha participado en diversos proyectos de visualización de datos y exploración de narrativa multimedia. Máster en Periodismo Digital de la Universidad de Alcalá de Henares, España. Tiene además una especialización en el Centro de Periodismo Digital de la Universidad de Guadalajara y en el Instituto de Tecnología de

Costa Rica. Recibió el primer lugar en Desafío InnovaData 2013 y fue residente de ProPublica gracias a una beca Douglas Tweedale del ICFJ.

Periodismo de datos en las salas de redacción

Periodismo investigativo a partir de la inteligencia de datos

Giannina Segnini

En el pasado, los periodistas de investigación dependíamos de rastrear fuentes documentales o testimoniales, o de las filtraciones de “gargantas profundas”, para reconstruir o comprobar hechos de interés público. Esas fuentes eran reducidas o de difícil acceso, pero además, solo podían o querían guiarnos hacia pequeños pedazos de la realidad, limitados a un determinado tiempo o espacio. Con astucia, intuición y mucha persistencia, podíamos conectar muchos de esos pedazos de realidad inconexos y descubrir una gran historia.

La observación aguda de síntomas claros de corrupción, como el enriquecimiento acelerado de funcionarios públicos, la aprobación expedita de leyes que favorecen a sectores específicos, o licitaciones públicas construidas a la medida para que sólo una empresa pueda ganarlas, eran y siguen siendo puntos de partida clásicos para iniciar una investigación periodística.

Después de 16 años de liderar la [Unidad de Investigación de La Nación](#), en Costa Rica, entendí que esas fuentes tradicionales no eran suficientes y que investigar hechos, personas o empresas aislados tampoco lo era. Tras la caída de un político corrupto siempre habrá otro a la espera de su turno, así como una nueva empresa dispuesta a sobornarlos para lograr sus objetivos. Entendí que debía abordar problemas sistémicos, pero también que es falso aquel refrán que dice que “se debe ver el bosque y no los árboles”.

Si lograba consolidar un nuevo modelo de periodismo de datos para investigar, podría ver el bosque y los árboles al mismo tiempo y liberarnos de la dependencia de las filtraciones.

Una idea descabellada

A inicios del 2010, propuse en el diario la creación de un equipo multidisciplinario que potenciaría la experiencia de los tres periodistas de investigación que integrábamos la unidad con la de dos ingenieros en computación. La idea era descabellada en todos sus extremos: juntar en un mismo cuarto a tres veteranos periodistas de investigación con dos jóvenes ingenieros que hablaban en un idioma “extraterrestre”, con el fin de consolidar y analizar océanos de bases de datos públicas para investigar.

Además del reto de poner a convivir dos especies tan diferentes, la propuesta presentaba un gran desafío metodológico: en vez de definir una hipótesis para confirmarla o descartarla, siguiendo paso a paso el método científico, experimentaríamos a la libre, entrevistando grandes volúmenes de datos. El nuevo método también rompía el orden normal en el proceso de obtención de datos: buscaríamos los datos primero y haríamos las preguntas después.

Desde finales de los años 90, yo había recurrido a la Corte varias veces para demandar acceso a bases de datos que me habían negado sobre pagos de subsidios públicos. Los magistrados fallaron todos los casos a mi favor, así que contábamos con jurisprudencia, y, en general, una normativa que amparaba el acceso a bases de datos públicas.

Hasta el día de hoy, la mayor parte de los datos los obtenemos mediante solicitudes formales a las instituciones públicas que las custodian y otra parte es extraída de forma automatizada mediante robots que visitan páginas abiertas en la web y capturan y almacenan la información. Todos los datos que extraemos usando [iMacros](#), son de naturaleza pública y no forzamos la entrada a ningún sistema o base de datos protegido.

Nuestro modelo de “cero desperdicio de datos” (*Zero-waste data journalism*) considera cualquier bit de información como un nuevo insumo, aunque se trate de datos históricos, y busca reutilizar conjuntos de datos en diferentes historias periodísticas. En esencia, se trata de juntar simples bits o

datos inconexos, consolidarlos para crear conocimiento y analizarlos para generar inteligencia.

Seis meses después de haber iniciado con el nuevo modelo, habíamos limpiado y consolidado más bases de datos públicas que el mismo Gobierno. Información de registro de personas, propiedades, bienes muebles, empresas, embarcaciones, aeronaves, así como contratos públicos, pagos de subsidios, infracciones de tránsito y procesos judiciales, por ejemplo, fueron parte del primer catálogo de datos que alimentó nuestros servidores.

El mayor reto vendría después, cuando los datos estaban limpios y consolidados y estábamos listos para empezar a entrevistarlos. Formular preguntas inteligentes, combinando de la forma más efectiva todas las variables disponibles, es la tarea más difícil cuando se investiga con datos.

El bosque y los árboles

Una de las primeras historias bajo el nuevo modelo la publicamos a finales del 2010, durante las elecciones regionales de alcaldes en Costa Rica. Construimos una base de datos con los nombres de todos los candidatos a alcalde de todos los partidos políticos y la cruzamos con las condenas por delitos criminales, los morosos con el Estado, los que tenían sanciones administrativas o que habían sido inhabilitados para ocupar cargos públicos.

El ejercicio reveló que más de cinco candidatos habían sido condenados por delitos como secuestro extorsivo, fraude, estafa y robo y otros 27 habían sido inhabilitados para ocupar cargos públicos porque cometieron alguna irregularidad en cargos anteriores como funcionarios públicos.

El reportaje se publicó una semana antes de las elecciones con un [mapa interactivo](#) en el que los ciudadanos podían encontrar los antecedentes de cada candidato, así como su versión sobre los hechos.

Ese mismo año, un nuevo Gobierno asumió labores con la promesa de reforzar uno de sus programas estrella: un subsidio llamado Avancemos que entrega una beca mensual a más de 167 mil estudiantes de secundaria en condiciones de pobreza con el propósito de que no abandonen sus estudios. Obtuvimos la base de datos de todos los jóvenes beneficiarios y la completamos con los nombres de sus padres y madres. A esos padres, a su vez, les buscamos ingresos y posesiones como propiedades, empresas y vehículos.

Nuestra búsqueda partía de la hipótesis de que muchos de los jóvenes beneficiados no eran realmente pobres y que el beneficio se había entregado irregularmente. El primer cruce reveló una lista de 75 estudiantes cuyos padres recibían salarios de entre \$2.000 y \$9.000 al mes. Una revisión más detallada de estos casos, en expedientes físicos, entrevistas telefónicas y visitas a sus casas, comprobó que no se trataba de errores en la asignación del beneficio, sino de jóvenes que vivían en hogares pobres a cargo de mujeres solas, en su mayoría desempleadas. Al otro lado de esta realidad, se halló que los padres de estos niños recibían salarios de hasta \$9.000 mensuales pero mantenían un contacto mínimo o inexistente con sus hijos y ex compañeras, y les aportaban poco o ningún dinero al mes.

Tras complementar el análisis de datos con la verificación en terreno, encontramos a un ejército de jóvenes que había sido abandonado por sus padres y a un Estado que subsidiaba ese abandono. La [publicación](#) se acompañó con una aplicación interactiva, [“Los rostros del desamparo”](#), que integró los casos estudiados y las historias contadas en videos y audios tanto de las madres de los jóvenes becados como el descargo de sus padres.

Dos tipos de agenda

Por haber nacido en una Unidad de Investigación, buena parte de la agenda periodística del nuevo modelo tenía que ver con temas de transparencia y gasto público. Cuando en el país se debatía la aprobación de una nueva ley fiscal para crear o modificar impuestos, publicamos una historia basada en datos sobre cuánto pagaban de impuesto inmobiliario los ministros y diputados que promovían esa

reforma, sus empresas y sus cónyuges.

[La investigación](#) descubrió que la mitad de los ministros del Gobierno [no declaraban o pagaban impuestos inmobiliarios](#).

Uno de los casos más graves era el del entonces ministro de Hacienda y su esposa, la asesora de la Presidenta de la República, quienes no habían declarado el valor de sus propiedades durante 12 años y pagaban menos impuestos por sus casas. Información adicional comprobó que tampoco reportaban todos sus ingresos a la hora de tributar y ambos tuvieron que renunciar a sus cargos.

Aunque los temas de transparencia siguen ocupando un lugar importante en la agenda, conforme pasaron los meses empezamos a incursionar en reportajes sobre otros temas sociales de impacto nacional, como educación, salud y transporte público.

Para estudiar las disparidades en la educación secundaria, analizamos seis años de resultados de las pruebas de ingreso a la Universidad de Costa Rica (UCR), tanto de los estudiantes provenientes de colegios públicos como privados. El análisis reveló que los estudiantes que venían de colegios públicos tienen la mitad de probabilidades de ingresar a la universidad pública que aquellos provenientes de colegios privados.

Para el proyecto, utilizamos los datos para producir una diversidad de nuevos “productos” que complementaron los reportajes: [un ranking](#) con indicadores de calidad por colegio que permitía comparar los resultados de hasta siete centros educativos y compartirlos en redes sociales y [una calculadora](#) para que los estudiantes que estaban próximos a realizar la prueba ingresaran su carrera preferida y estimaran el resultado que debían obtener para ingresar, con base en un algoritmo que procesó los datos históricos.

Reconstruyendo Offshore Leaks

Desde que iniciamos con el nuevo modelo de datos, buena parte de los esfuerzos han estado concentrados en automatizar la extracción, limpieza y normalización de los datos mediante el uso de software libre, como [Open Refine](#) y [Talend Open Studio for Big Data](#), poderosa herramienta que permite integrar y organizar grandes volúmenes de datos en un ambiente gráfico, sin tener que escribir código.

Esas herramientas jugaron un papel fundamental cuando [The International Consortium of Investigative Journalists](#) (ICIJ) nos encomendó reconstruir más de 320 tablas y desarrollar una aplicación interactiva para el proyecto [Offshore Leaks](#), una de las mayores investigaciones periodísticas transnacionales de la historia.

La investigación emprendida por el Consorcio reveló cómo prominentes políticos, aristócratas y banqueros de diversos países usaron paraísos fiscales para crear compañías o fideicomisos, y con ello ocultar sus bienes o capitales, y en muchos casos evitar el pago de impuestos y ocultar actos de corrupción.

La Unidad de Investigación de La Nación fue escogida por ICIJ para procesar [dos grandes bases de datos](#), que medidas en gigabytes son 160 veces más grandes que la filtración de los documentos diplomáticos estadounidenses realizada por WikiLeaks en el 2010. Reconstruimos las conexiones de más de 105 mil compañías, fideicomisos y fondos secretos en paraísos fiscales, sus dueños y representantes y desarrollamos una [aplicación interactiva](#) que visualizó todas esas conexiones. Desde el lanzamiento de la aplicación en Junio del 2013, en Hamburgo, Alemania, la aplicación recibió más de dos millones de visitas en sólo dos días y hoy es utilizada por cientos de periodistas de investigación y otras organizaciones en el mundo para investigar paraísos fiscales.

El desarrollo de “Offshore Leaks”, es quizás el proyecto más ambicioso que hemos emprendido hasta

ahora, por la complejidad y abundancia de los datos, la trascendencia y el secretismo de su contenido y el impacto y tráfico mundial que generó su publicación.

Cuatro niveles de lectura

La experimentación con datos pronto nos llevó a entender que con sólo un conjunto de datos se pueden generar múltiples productos editoriales para diferentes tipos de audiencias. Nuestros proyectos más recientes contemplan la producción de al menos cuatro productos para diferentes niveles de lectura: los jóvenes o lectores esponja, que buscan la esencia de la historia en un video o animación de dos minutos; los lectores habituales, que leen la noticia impresa o en línea; los exploradores, que además de leer las notas interactúan con las aplicaciones interactivas y crean sus propias historias; y, finalmente, los topos meticulosos, que esperan la publicación completa de las bases de datos o documentos utilizados en el reportaje para generar sus propios cálculos o visualizaciones.

Una reciente investigación sobre el “Reciclaje en Costa Rica” demostró la efectividad de esos cuatro niveles y la utilidad del análisis de datos como brújula para dirigir el reporteo. Por primera vez en Costa Rica, el Censo Nacional preguntó a los ciudadanos si separaban su basura para reciclarla.

Comparamos todas las respuestas del Censo con una base de datos que construimos manualmente y que identificó a los gobiernos locales que reciclan y a los que no. El análisis general demostró que un 40% de los municipios desperdician el esfuerzo de sus habitantes al separar la basura. Simplemente recogen los desechos separados por más de 212 mil hogares y los vuelven a mezclar en sus camiones recolectores. Esos datos generales los publicamos en un video con animación que resumió los hallazgos en [las publicaciones impresas y en línea](#) y que luego se desglosan en una [aplicación interactiva](#), en la que todos los ciudadanos podían consultar cuánto recicla su gobierno local.

Nunca antes habíamos visto tan de cerca el poder de los datos como brújula para el reporteo de calle como en este reportaje. Tras realizar el análisis, emergió en el primer lugar de la lista de las regiones que más reciclan un pueblo cafetalero llamado Dota, donde según los datos, 8 de cada 10 personas separan sus desechos. Cuando la periodista [visitó el pueblo](#), se encontró con una [legión de formadores](#) que enseñan, casa por casa, las mejores prácticas para reciclar, ahora totalmente arraigadas en una población que, por sus prácticas verdes, recibió \$10 más por cada uno de los 50 mil quintales de café que exportó en el 2012.

Todas las bases de datos generadas para la investigación estaban disponibles para ser bajadas desde el sitio especial y eso permitió alimentar una [hackathon](#) sobre medio ambiente en la que programadores ensayaron otras propuestas de visualización de los datos de reciclaje.

Cuando se habla de periodismo de datos muchos imaginan a un grupo de nerds con sus ojos pegados a una pantalla y que renunciaron al reporteo de calle. Casi cuatro años después de iniciado este experimento, nuestro modelo de trabajo en equipo, junto con programadores, geógrafos y diseñadores, ha permitido crear un puente entre lo mejor del periodismo de investigación y la adaptación de nuevas tecnologías, convirtiendo datos inconexos en inteligencia para entender mejor cómo funciona nuestra sociedad.

De cara al futuro, figuran aún muchos retos para adoptar mejores tecnologías, ya en uso por los científicos, que nos permitirán analizar grandes volúmenes de datos en tiempo real y hasta anticipar comportamientos o fenómenos. La experimentación permanente y el trabajo en equipo son la clave para impulsar ese avance.

Giannina Segnini (@gianninasegnini)

Periodista. Durante más de 16 años lideró la Unidad de Investigaciones en La Nación de Costa Rica,

formando un equipo reconocido internacionalmente por su aporte al periodismo de datos. Ha trabajado en instituciones internacionales, realizado asesorías a medios y entrenado a cientos de profesionales en periodismo de investigación y periodismo de base de datos en América Latina, EE.UU., Europa y Asia. Es parte del gran jurado de la primera edición del Global Award on Data Journalism (Global Editors Network/Google).

Equipo de datos y buenas prácticas

Por Sandra Crucianelli

La imagen no es sólo una foto. Es un símbolo. Representa la comunión que es posible gestionar desde un equipo (inicialmente no integrado por periodistas) que trabajaba abriendo datos cerrados, hasta su posterior inserción en la sala de redacción donde se produjo un encuentro mágico: el de reporteros y editores de las ediciones impresa y digital, infografistas y diseñadores multimedia con programadores, expertos en data mining y analistas de sistemas.



A principios de 2012, el Knight International Journalism Fellowship Program, me encomendó –entre otras- la tarea de formar, no un equipo de datos porque el mismo ya estaba en funcionamiento, sino un “equipo periodístico de datos” en el diario La Nación de Argentina. Desde mi Bahía Blanca natal, hacia Buenos Aires y allí, en el corazón de una de las estructuras periodísticas más grandes de América Latina, han ocurrido una serie de procesos y cambios dignos de ser analizados.

Al principio de la historia, dentro de la gerencia del área Multimedia, -que lidera Momi Peralta Ramos-, se formó a partir del 2011, un grupo de profesionales que se mostró desde el principio interesado en transformar datos cerrados en datos abiertos; primero compartiéndolos con la audiencia a través de un catálogo en uno de los blogs que posee el periódico y también, interesando a reporteros como Diego Cabot, quien fue el primero en usar ese enorme caudal de información convertido en una base de datos, para escribir una serie de artículos sobre cómo el Gobierno argentino usa dinero público para subsidiar a empresas de transporte de pasajeros en todo el país. A él se sumaron luego tantos otros como Hugo Alconada Mon, Laura Serra, Maia Jastreblansky, Juan Pablo de Santis, Iván Ruíz, entre muchos otros más, tanto de la edición gráfica como de la versión digital del diario. Mientras desde mi *Fellowship* se realizaba un diagnóstico del recurso humano existente, Michael Zanchelli, miembro de ICFJ, investigó en Estados Unidos y en Inglaterra cómo funcionan los equipos que hacen periodismo de datos.

Sin embargo, a poco de andar el camino, me di cuenta que era necesario considerar el contexto latinoamericano y por eso ambos decidimos incluir en su estudio original las experiencias de los otros

dos equipos que ya estaban funcionando en países de habla hispana, concretamente [los equipos](#) de La Nación de Costa Rica, por entonces liderado por la periodista de investigación Giannina Segnini y el del diario O Estado, en Brasil, a cargo del periodista investigador José Roberto de Toledo. Hasta ese momento, había pocos puntos en común entre unos y otros equipos. Todos se caracterizaban por disponer entre sus filas a un minero de datos y programador, aunque no siempre funcionaban de la misma manera.

Dado el contexto, la estrategia que se aplicó en La Nación de Argentina fue a través de dos vías en simultáneo: una para sumar periodistas al equipo que procesaba los datos y la otra para insertar a ese equipo dentro de la sala de redacción. Sólo de este modo podríamos tener lo que se ve en la foto: lo que se conoce en inglés como *Data Journalism Team*.

La mayoría de los periodistas de habla hispana, hasta ese momento, sabía poco y nada sobre el periodismo de bases de datos. Muchos creían que se trataba de periodismo de precisión y otros lo consideraban un pariente cercano del periodismo asistido por computadora. En realidad, no se equivocaban ya que el periodismo de bases de datos es ante todo, periodismo de investigación, que incluye desde los principios postulados por Philip Meyer, hasta el uso de hojas de cálculo, pero con un ingrediente adicional: la incorporación del programador informático al equipo de noticias, sin cuya ayuda, no se lograría en ciertos casos, ni la extracción sistematizada de los datos, ni su gestión, mucho menos el diseño de lo que se conoce como *News Apps* o aplicaciones de noticias.

Para quien crea que formar un equipo de datos se basa en su entrenamiento sobre matemática y estadística, se equivoca. No se trata de training. El training es sólo una parte del proceso. Se trata de crear las condiciones en una infraestructura de noticias para que cada profesional interesado en ser parte de un equipo de esta naturaleza, pueda contar con los recursos en tiempo y forma para rescatar las historias escondidas detrás de los números.

Un equipo de datos es algo diferente a un equipo periodístico de datos. El primero es la piedra basal y condición necesaria, pero el segundo es el gran angular que hace posible la trascendencia de los datos, fluyendo hacia la información, para dar vida a noticias que habitualmente se encuentran atrapadas en celdas de cientos y cientos de hojas de cálculo. Buena parte de la tarea es abrir candados, metafóricamente hablando. En un país sin datos abiertos y sin ley nacional de acceso a la información pública, hubo que desarrollar una estrategia que tuvo como sello distintivo su flexibilidad, es decir; ningún punto del plan de acción inicial fue rígido. Cualquier pre concepto podría ser revisado una y otra vez y cualquier teoría ser sometida a revisión crítica hasta encontrar el camino más adecuado que nos condujo finalmente, a lo que se ve en la foto.

Dentro de las lecciones aprendidas puedo citar como positivas:

- La incorporación de un coordinador, experto en procesamiento de datos, a las reuniones con los editores, en las que se toman decisiones sobre las noticias del día y de la semana.
- Evangelización. Resultó vital el diálogo durante reuniones informales con reporteros, básicamente para mostrarles casos de éxito. Generalmente, los periodistas asocian al periodismo de datos con la matemática y la estadística, dos materias poco amigables dentro de la sala de redacción. Pero cuando toman cuenta de que a partir de grandes volúmenes de datos se pueden encontrar historias que generen alto impacto, esa percepción cambia y comienzan a preguntarse: ¿y esto, de qué se trata?
- Trabajar por proyecto. Las metas en el mediano y largo plazo funcionan mejor en el periodismo de datos, al no estar sometido a la presión diaria de la noticia caliente, en especial cuando no se pueden disponer de reporteros al 100 % de su tiempo. Para ello, los editores acuerdan con el

reportero interesado en trabajar con datos qué porción de su horario laboral semanal podrán dedicar a este tipo de trabajos periodísticos, sabiendo de antemano que el resultado que obtendrán, no es la misma nota que la de su competencia, sino un producto original, la mayor parte de las veces, una primicia. El trabajo por proyectos ahorra costos de producción en materia de recursos humanos y además permite integrar a un mayor número de periodistas del equipo que procesa datos. En este caso se apeló a reuniones de planificación semanales y training específico, adaptado al caso, cuando ello resultó necesario.

- Mayor visibilidad del proyecto de datos: ciertamente visualicé esta necesidad desde el comienzo de mi *Fellowship*. Planteada la necesidad, lo que se hizo fue ir más allá de dar un espacio destacado al Blog de Datos. Se estableció el tag “Nacion Data” y se diseñó un canal de datos, como producto único en su tipo, en el que se unificaron contenidos basados en datos. Fue incorporado a la barra de secciones de la página frontal de la edición digital a mediados de 2012.
- Consensuar buenas prácticas dentro de la sala de redacción.
- Desarrollar un programa de entrenamiento intensivo. A principios de 2012 no existía un programa global e integral de entrenamiento en periodismo de bases de datos. Algunas experiencias se basaban en análisis de casos y otras en condiciones de borde demasiado particulares que o bien funcionaban en determinado país o bien lo hacían sólo en grandes medios con recursos, pero dejaba abierta una gran brecha con relación a los medios más pequeños, como los de provincia. Para ello, se diseñó un programa con dos ejes transversales: uno general, que podría replicarse en cualquier país y otro particular, aplicado al caso argentino y fácilmente adaptable a cualquier otro país latinoamericano.

El diseño de este programa comenzó con una pregunta: ¿qué conocimientos básicos debe tener un periodista que desee sumarse al equipo de datos?

- Saber y entender, cómo funciona el Estado Nacional en al menos 3 puntos básicos: la estructura de la administración pública, los roles de las instituciones (incluyendo en especial a los entes de control) y la ruta del dinero entre el ciudadano que paga sus impuestos y el poder que administra los fondos públicos.
- Tener dominio sobre las técnicas de búsqueda y recuperación de datos en dos niveles: uno basándose en motores de búsqueda pero aplicando técnicas avanzadas de searching; otro accediendo a bases de datos nacionales e internacionales.
- Primer nivel en el procesamiento de datos: descarga desde la web, uso de conversores al formato Excel, siempre y cuando ello sea posible en la sala de redacción y operaciones básicas dentro de una hoja de cálculo (suma, resta, valor más frecuente, variación porcentual y gráficos simples como tortas, líneas quebradas y barras ; en otros casos, generar un canal de comunicación fluido a través de la figura de un coordinador de equipo, para aquellos casos en los que se requiera extracción automatizada de datos a través de software específico.
- Segundo nivel de procesamiento de datos: nociones básicas de matemática y estadística descriptiva y estadística inferencial. Test diagnóstico y aprendizaje por desafío.
- Tercer nivel de procesamiento de datos: definición de las variables disponibles a partir de los datos primarios. Análisis de posibles cruces y diseño de nuevas variables según cada caso en particular.
- Reflexión sobre las emociones que suelen prevalecer en el periodista investigador y que se trasladan al equipo de datos, cuando se abordan temas de extrema sensibilidad.

[El programa de entrenamiento intensivo](#) diseñado para La Nación involucró 40 horas totales de capacitación interna.

Reunió en un mismo espacio, en mismos días y horarios a todos los miembros del equipo, desde periodistas y editores, hasta analistas de datos, diseñadores, expertos en visualización y programadores.

La parte más importante del periodismo de datos es ... el periodismo. No hay periodismo de datos sin historia. Ahora, ¿qué tanto sabemos los periodistas sobre contar historias a partir de datos? Esta es la pregunta que hay que hacerse. Una mirada introspectiva que todo colega debería tener sobre sí mismo, sobre su propio conocimiento y su necesidad de aprendizaje constante. En esta disciplina, en algunos casos se requiere saber acerca de algunas técnicas básicas de programación (Python, Perl, Ruby, etc.) aunque soy partidaria de que este conocimiento quede en manos del programador.

Buenas prácticas en periodismo de datos

* Identificar correctamente los productos del periodismo de datos. Hay al menos cuatro productos diferentes, generalmente complementarios, que surgen de esta práctica:

- [Artículos basados en datos](#): Generalmente son artículos cortos, contruidos a partir de grandes volúmenes de datos como los que podrían estar contenidos en bases de datos o conjuntos de datos diseñados a partir de una o varias bases de datos.
- [Visualizaciones interactivas](#): En general son un complemento de los artículos basados en datos, pero pueden ser contenidos en sí mismos sin necesidad de que exista un artículo referencial: en algunos casos es suficiente un título, una bajada o subtítulo y una explicación a la audiencia sobre cómo leer los datos desde la visualización.
- Conjunto de datos abiertos (conocidos como set de datos): Los medios de comunicación no necesariamente se abocan a realizar trabajos periodísticos basados en datos. También pueden, como el caso de La Nación de Argentina, dedicar una gran parte de su esfuerzo a la apertura de datos públicos. En países donde no existen leyes de acceso a la información pública, donde éstas no funcionan como deberían, donde en los sitios web del Gobierno hay algunos datos pero otros no o están ocultos bajo distintas capas de información y donde los formatos para mostrar los datos vía web no son accesibles: ver un dato en la pantalla no es lo mismo que acceder al mismo. Para que un dato sea accesible se debe mostrar en un formato reutilizable. Ejemplo: un documento oficial subido a Flickr. En estas circunstancias medios y organizaciones pueden dedicarse a la búsqueda, extracción y apertura de datos públicos, tal como LN hace a través de su catálogo de datos. Esto es: los datos que los gobiernos o las organizaciones no abren (sea por las razones que sean), los medios de comunicación pueden incursionar en ese terreno con lo cual seguramente harán grandes aportes. Los datos pueden gestionarse o administrarse desde plataformas sencillas y gratuitas como [Google Drive](#).
- Aplicaciones de noticias: Conocidas en inglés como *News Apps*. Son el fruto de la labor de periodistas + programadores informáticos. Muchas veces el volumen de datos es tan grande que resulta imposible encontrar una noticia si no se diseña una aplicación que nos permita agrupar y analizar variables, por ejemplo, por localización geográfica, por fecha, por nombre de compañía, etc.
- Ejemplos:
 - [Gasto público en Bahía Blanca](#)
 - [Salarios de empleados públicos \(Texas Tribune\)](#)
 - [Pagos de compañías farmacéuticas a doctores \(ProPublica\)](#)

No es acerca de matemática que estamos hablando, pero ciertamente una buena práctica es tener conocimientos básicos en la materia. La meta en periodismo de datos: es minimizar el error de análisis al mínimo posible. Trabajar con error cero es imposible, pero minimizarlo debe ser una meta de calidad del Data Team. Esa fue una de las razones por las que deberían ver el filme "El Informante": la presión bajo la que trabajan los periodistas de investigación y las emociones que surgen de esa presión, es alta; por eso hay que mantener la mente fría, no apurarse, revisar al menos tres veces los cruces de datos y ante la más mínima duda consultar con otra fuente.

* El caso de la variable "valor del dinero"

El valor del dinero cambia a lo largo del tiempo. Por eso, una conclusión puede ser matemáticamente correcta pero fácticamente inaceptable por entrar en conflicto con la realidad. Por ejemplo, para una serie de datos relacionados con montos de dinero, donde los únicos datos disponibles son el valor final y el inicial, no podemos decir que el ritmo sea sostenido (lineal); tal vez es exponencial y no lo vemos porque no tenemos los datos internos dentro de ese lapso de tiempo.

Cuando la serie temporal es larga (2003 - 2011) en ese caso, se debe cruzar con el IPC.

Ejemplo: un funcionario reporta su patrimonio

2003 = patrimonio de 35.000 \$ En diciembre del 2003 el IPC era de 68,11

2011 = patrimonio de 981.093 \$ En diciembre del 2011 el IPC era de 135,67

El cálculo se hace de esta manera:

IPC 2003 = 68,11 --- 35.000 \$ (*patrimonio declarado al 2003*)

IPC 2011 = 135,67 --- $x = (135,67 \times 35.000) / 68,11 = 69.717$ \$ (*Patrimonio ideal*)

Pero el patrimonio real es 13 veces mayor al ideal. Esto quiere decir que si su crecimiento patrimonial hubiera guardado correlato con el IPC, su actual patrimonio debería ser de menos de 70 mil pesos y no casi de un millón de pesos como se reportó.

Me parece recomendable que vean este procedimiento porque es probable que tengan que toparse más de una vez con la variable "dinero", en la que de un año para el otro, se puede calcular la variación % y hablar en términos nominales porque a menos que haya hiperinflación, los valores no difieren mucho; pero en una serie temporal tan larga, cruzar con el IPC es un criterio de calidad en materia de análisis de datos.

* Verificación de datos con otras fuentes.

* Chequeo aleatorio de datos, sobre una muestra basada en el universo bajo estudio, al menos tres veces.

* Narrativa no demasiado extensa ni sobrecargada con información numérica.

* Remitir conclusiones a la visualización y mencionar el respaldo documental sobre el que se trabajó.

El periodismo de datos maneja un volumen grande de datos. Sin una adecuada visualización que acompañe el análisis de esos datos, sería imposible ver de manera sencilla las variables que se ponen en estudio. Las crónicas podrían ser extensas y pecar de aburridas si están plagadas de números que la gente puede no entender con facilidad si no son expuestos de manera adecuada. Por eso se sugieren crónicas cortas, con conclusiones contundentes y una visualización importante, dominante en el artículo, a la hora de contar una historia.

Los expertos en visualización interactiva son capaces de diseñar, a partir de una gran cantidad de datos, un gráfico que permite al lector decidir qué desea visualizar y cómo. La visualización de datos

explica de manera comprensible las relaciones entre gran cantidad de información que se genera por distintas vías.

El clásico ejemplo para el reportero es el presupuesto público. El Ministerio de Economía entrega un pesado reporte lleno de tablas y números que a simple vista no producen ninguna información interesante, mucho menos una noticia que no sea la anunciada en el parte oficial de prensa. Pero un presupuesto contiene variables, es decir elementos que son medidos a lo largo del tiempo. Y si se dispone de un presupuesto del año en curso, probablemente tenga a mano el del año anterior, que tiene... las mismas variables. Sin embargo, muchas veces, hasta la más sencilla de las tablas puede resultar muy aburrida.

La comparación es quizá una de las prácticas más usadas por el periodismo, pero cuando hay que procesar gran cantidad de información numérica, el asunto se complica. Por eso, se recurre a distintas formas de visualizar el conjunto de datos; de ese modo podemos analizar mejor ciertos procesos y “visualizar” cambios que de otro modo no habrían saltado a simple vista.

* Uso adecuado del hipertexto: compartir fuentes primarias y secundarias, no remitiendo a las fuentes web desde sus páginas de inicio sino a la URL permanente desde dónde se obtuvieron los datos utilizados.

* Visualización interactiva: usar una gama de recursos no demasiado extensa, pero sí lo suficientemente versátil como para que pueda dar solución a cada tipo de variables bajo estudio.

* Compartir datos abiertos con la audiencia a través de Google Spreadsheet.

* Mostrar a la audiencia la documentación de respaldo original completa a través de la plataforma [DocumentCloud](#).

* Mantener comunicación interna fluida incluyendo a todos los miembros del equipo involucrado y muy en especial, a los editores.

* Mantener abiertos los canales de comunicación con la audiencia a través de canales participativos mediante comentarios y difusión en redes sociales. Se sugiere no cerrar la vía de comentarios.

* Los sistemas de visualización interactiva deberían tener siempre la opción de descarga de los datos.

* No preocuparse por la competencia en la decisión de compartir documentos.

¿Qué requisitos deberían observarse para encuadrar los contenidos en esta disciplina?

Cuando las notas tengan alguno o varios de estos componentes:

- Datos abiertos a disposición de la audiencia (descargables o en DocumentCloud, Google Drive, Junar, etc.)
- Data Viz interactiva (Tableau Public, Google Fusion Tables, etc)
- Un mashup (que es un híbrido) como cuando se cruza data de Twitter posicionando esa información en Google Maps.
- Trabajo de *data mining* (minería de datos)
- *Scraping* de datos (extracción de datos automatizada)
- Uso de bases de datos propias o ajenas a partir de la cual se construye una propia.
- Trabajo de recuperación documental web: búsqueda en la internet profunda o invisible.
- Casos de acceso a la información, pero no cuando es uno aislado, sino cuando hay un conjunto de peticiones importante (más de diez por ejemplo y sistematizadas)
- Trabajo de campo propio: es decir cuando la data no está (caso anterior, pero en vez de basarse en peticiones se basa en análisis de webzines del Gobierno, hay que construirla y crear un Excel propio, sobre la base de un [análisis de información web](#).

Sandra Crucianelli (@spcrucianelli)

Periodista. Consultora y Trainer sobre contenidos digitales para los diarios Reforma de México, Mural de Guadalajara, La Prensa de Honduras, El Universal y Expreso de Ecuador, El Tiempo de Colombia, La Nación de Argentina, El Comercio de Lima, Perú y Folha de San Pablo, Brasil. Docente para el Knight Center de la Universidad de Texas; el Centro de Formación Digital de la Universidad de Guadalajara, México; CELAP, Centro Latinoamericano de Periodismo, Radio Nederland Training Center, PNUD Honduras, Instituto del Banco Mundial, Washington, USA; Fundación Violeta Chamorro, Nicaragua; Red de Periodistas de Provincias de Perú, Piura, Perú; Consorcio de Investigación Económica y Social, Perú. Fundación MEPI, DF, México. Autora del Libro "Herramientas Digitales para Periodistas" (2008, Knight Center, Universidad de Texas).

Especial de La Tercera: 40 Años del 11 de Septiembre de 1973

Francisco Capone

Desde principios del año 2013 se tenía claridad de que se debía construir algo potente en el sitio de LaTercera.com con motivo de la conmemoración de los 40 años del Golpe Militar en Chile. El trabajo fue liderado por Macarena Lescornez, editora de Medios Digitales de Copesa, quien congregó a un equipo de 9 profesionales de diversas áreas y con habilidades complementarias.

La gestación

El inicio del proyecto no fue fácil, puesto se presentaron diversas propuestas digitales para básicamente narrar lo que había ocurrido ese día de 1973. Finalmente se llegó a un consenso: [un gran interactivo multimedia](#). La duración del trabajo fue de un poco más de 3 meses y se contó en un principio con la asesoría de Mariana Santos, becaria del Centro Knight para el Periodismo en las Américas. Su labor fue la de proporcionar consejos sobre narrativa digital, mostrar trabajos en los que ella había participado para el medio The Guardian, y por sobre todo, convencer al equipo de que era capaz de realizar un gran proyecto.



Planificación

El proceso partió con la generación de una carta gantt con plazos bastante acotados. El primer hito fue la investigación y edición periodística. Para esto se construyó el guión tomando como base el trabajo periodístico realizado por el periodista Ascanio Cavallo para un especial del año 2003 de La Tercera en versión papel. Una vez construido el guión pasamos a la etapa de diseño del interactivo.

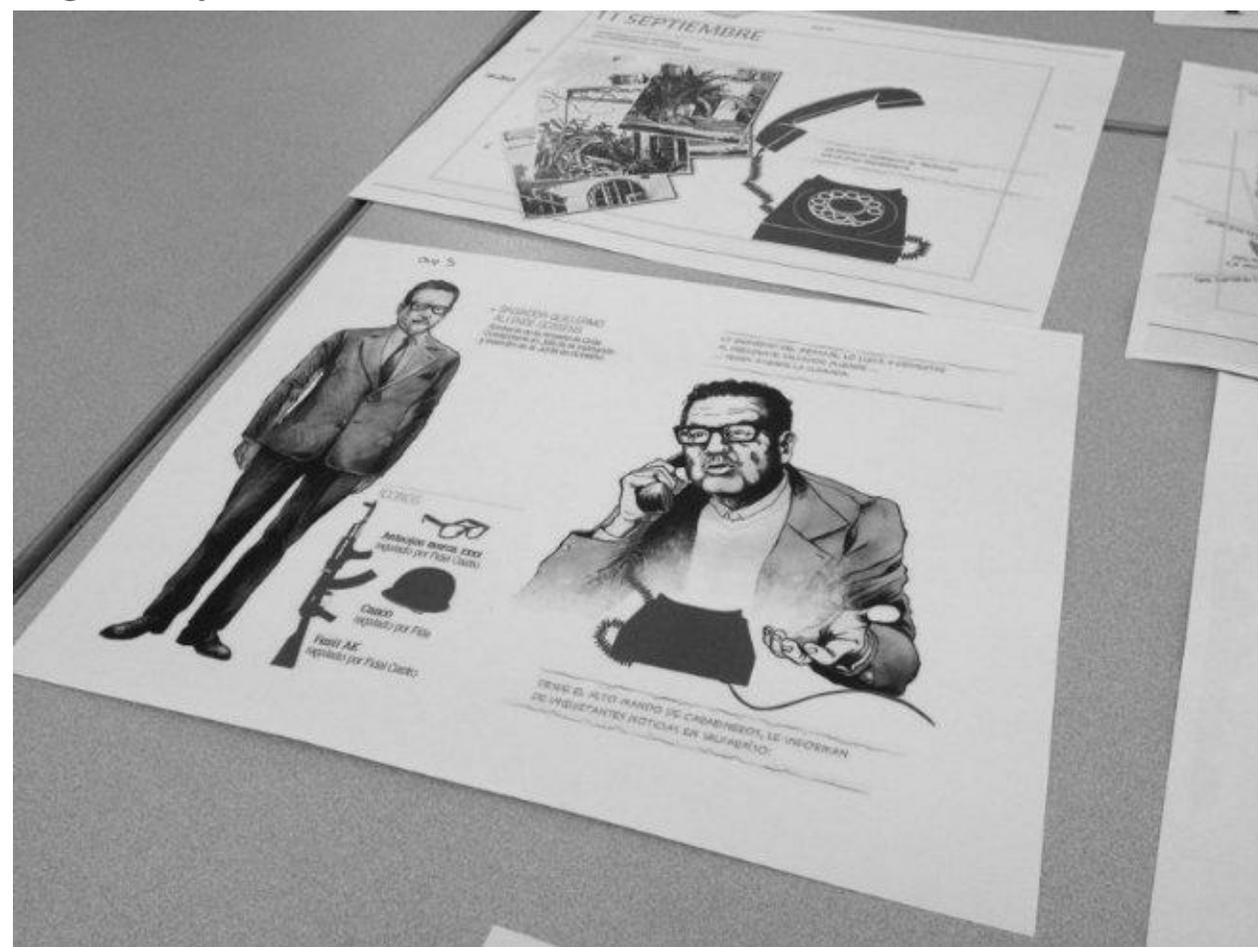


Diseño y desarrollo

Una vez que el diseño general estaba realizado se pasó a la etapa de creación gráfica y de recopilación de material multimedia. En esta labor fue muy importante el trabajo de Víctor Abarca, ilustrador de La Tercera, quien fue el encargado de crear la gran mayoría de las imágenes dibujándolas a mano para luego digitalizarlas y post producirlas. El trabajo de Víctor fue acompañado por el periodista Hugo Infante, quien fue fundamental en documentación y edición de imágenes. También se debe agregar la participación del equipo de LaTerceraTV con el apoyo y edición del material audiovisual que se utilizaría.



La siguiente etapa en el proyecto fue el diseño y maquetación gráfica web de las escenas dispuestas en el guión. El trabajo fue realizado en conjunto por los infografistas Paula Tala, Jesús Pérez y el diseñador gráfico Freddy Vásquez. Básicamente, Paula y Jesús construían escena por escena para que luego Freddy las llevara a formato web.

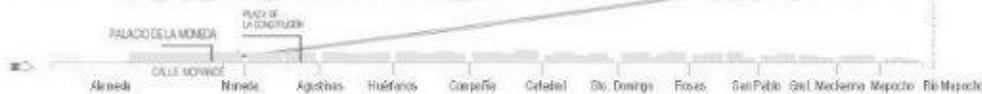
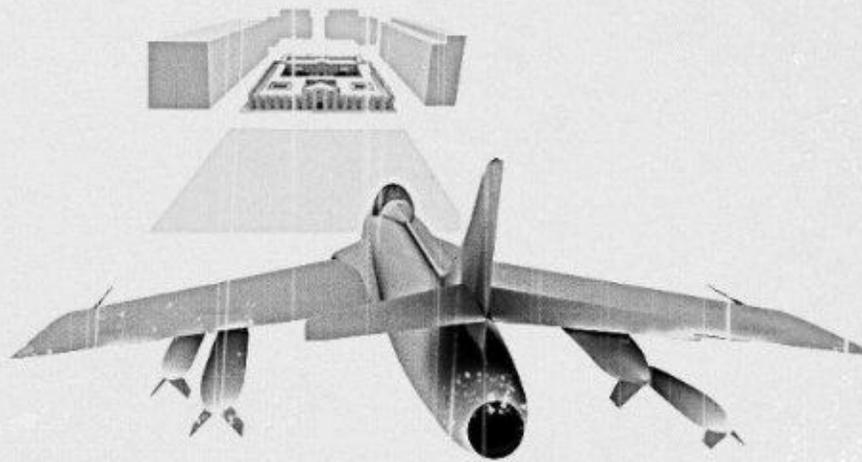


La fase final en la construcción del interactivo fue la programación y animación. Esta parte del trabajo fue una de las más complicadas, puesto que se contaba con más de cincuenta escenas donde cada elemento debía ser animado de forma relativa a la interacción del usuario con el mouse y el uso del scroll o barra de desplazamiento de un navegador web.

La técnica utilizada para lograr dicho objetivo se denomina [Parallax](#) y es una de las tendencias en la web. Su elección no fue azarosa sino que se basó en una investigación de otros trabajos periodísticos multimedia, como el realizado por el New York Times con su especial denominado [Snow Fall](#) y el trabajo realizado por The Guardian y su especial [Firestorm](#).



Un gran desafío fue la animación de la escena del bombardeo del Palacio de la Moneda, ya que para lograr el efecto deseado, se realizó una animación de más de ochenta cuadros, trabajando uno por uno.



Finalización, lanzamiento y logros

Una vez que el interactivo ya cumplía con la mayoría de las expectativas del equipo, se procedió a realizar distintas pruebas de funcionamiento y corrección de detalles o bugs encontrados. El lanzamiento de este trabajo fue programado para el día 9 de Septiembre del 2013, dos días antes de la fecha conmemorativa.

Ninguno de los integrantes del equipo imaginó el impacto y lo trascendental que sería este proyecto. Se superaron con creces las expectativas en cuanto a la cantidad de visitantes y el recibimiento del público en general.

11 SEPTIEMBRE
1973



Mañana Lunes no te pierdas en www.latercera.com
el **ESPECIAL INTERACTIVO** "11 de septiembre 1973"
a 40 años del golpe militar que cambio la historia.

40 años
GOPE MILITAR

LATERCERA

Sin duda, este proyecto marcó un antes y un después respecto a la presentación y visualización de un trabajo periodístico serio sobre una temática bastante controvertida, pero logrando matices lúdicos y llegada a un público heterogéneo.

Integrantes del proyecto:

- Andrés Azócar: Director Medios Digitales de Copesa
- Macarena Lescornez: Textos y edición general del proyecto
- Hugo Infante: Edición de imagen y documentación
- Víctor Abarca: Dirección de arte e ilustración
- Paula Tala: Diseño y maquetación
- Jesús Pérez: Diseño y animación
- Francisco Capone: Programación y desarrollo
- Freddy Vásquez: Maquetación y armado
- Israel Muñoz: Modelado 3D
- Pablo Gándara: Imagen y audio
- Juan Pablo Cortés: Imagen

Francisco Capone(@smokenight)

Ingeniero Informático de la Universidad Santa María (Chile), con más de diez años de experiencia en desarrollo, programación y gestión de programas computacionales. Actualmente ligado 100 por ciento al desarrollo web y miembro del área de Medios Digitales de Copesa.

Historia del periodismo de datos en Brasil

Marcelo Träsel

Los periodistas brasileños hoy pueden contar con material y herramientas necesarias para la práctica del Periodismo Guiado por Datos ([PGD](#)). Los servicios en línea como Google Drive, Infogr.am, DocumentCloud y CartoDB, sólo para nombrar unos pocos, permiten construir, organizar y analizar bases de datos, si se dispone sólo de un ordenador, conexión a Internet y habilidad con el inglés. En mayo de 2012, la Presidencia de Brasil promulgó la [Ley N° 12527](#), conocida como la Ley de Acceso a la Información, que obliga a todos los organismos públicos a ofrecer activamente datos administrativos y atender las solicitudes de información de cualquier ciudadano. Estos dos factores han alimentado el interés de la prensa nacional por la aplicación de técnicas computacionales en la producción de noticias.

Una indicación importante de este interés creciente fue la creación de un equipo dedicado al periodismo de datos en el periódico O Estado de S. Paulo, que se consagró como pionero en Brasil. En mayo de 2012, el equipo Estadão Dados puso en marcha el [Basômetro](#), una de las primeras aplicaciones periodísticas brasileñas. En agosto del mismo año, el periódico Folha de São Paulo pasó a albergar el weblog [FolhaSPDados](#), cuyo objetivo es la creación de vistas gráficas y mapas relacionados con los artículos publicados tanto en la versión impresa como web del diario. En Río de Janeiro, O Globo también ha de aplicar estas técnicas con cierta constancia. Además de estos tres principales periódicos nacionales, Gazeta do Povo, del Paraná, ha utilizado su experiencia en periodismo de investigación para producir grandes reportajes basados en datos. El periódico gaúcho Zero Hora, por su lado, se ha dedicado al tema de la transparencia pública a través de informes y del weblog [Livres Acesso](#), que se inició en 2012 para supervisar la aplicación de la Ley de Acceso a la Información en el país.

En el campo del periodismo independiente, el mejor ejemplo es el [InfoAmazônia](#), creada en 2012 por el Knight Fellow Gustavo Faleiros en colaboración con el periódico web O Eco y la Internews. En 2013, O Eco creó un Laboratorio de Innovación en Periodismo Ambiental, llamado [EcoLab](#). [La Agencia Pública](#) es otra iniciativa independiente que aplica las técnicas de PGD, aunque lo hace con moderación, sin embargo, fue el responsable de una importante contribución a través de una asociación con Wikileaks, elaborando informes a partir de la biblioteca PlusD (Public Library of US Diplomacy).

Estos ejemplos sugieren que estamos experimentando los primeros pasos en un intento de institucionalizar el ejercicio del periodismo de datos en las salas de prensa brasileñas. La base del éxito de PGD en el país, sin embargo, se puso en marcha en la década de 1990.

Durante el Gobierno de Fernando Collor de Mello, el periodista Mário Rosa, en ese momento en el Jornal do Brasil, utilizó el Sistema Integrado de Administración Financiera del Gobierno Federal (Siafi) para comprobar el sobreprecio en la compra de leche en polvo por Legión Brasileña de Asistencia (LBA), presidido por la entonces primera dama, Rosane Collor de Mello.

En 1990 Mário Rosa tuvo acceso al Siafi a través del senador Eduardo Suplicy, del Partido de los Trabajadores. Lúcio Vaz reporta el caso en el libro “A ética da malandragem” (La ética del engaño):

“Firmado por el periodista Mário Rosa, el informe era completo, con datos nunca antes vistos, tales como pedidos y compromisos bancarios (reservas hechas en el presupuesto de la Unión). Mário había descubierto el Sistema Integrado de Administración Financiera (Siafi), una expresión que se tornaría muy conocida de periodistas y políticos de renombre en los años siguientes. El acceso a este sistema que registra los gastos del Gobierno federal, le permite hacer una radiografía completa de todos los

pagos efectuados a los contratistas, proveedores, estados y municipios. Una mina de diamantes para los reporteros. El periodismo ganó una nueva e importante fuente de información, más técnico, casi científico.”

En aquel momento, el acceso a este tipo de base de datos del Gobierno era cerrado a los periodistas y los ciudadanos. El autor del artículo, Mário Rosa, sólo pudo llevar a cabo investigaciones en Siafi porque el entonces senador de la República Eduardo Suplicy le prestó la contraseña que tenía por el ejercicio de sus actividades parlamentarias. A partir de este y otros reportajes, el Gobierno Federal decidió permitir oficialmente el acceso de los periodistas al Siafi, por lo que es una de las primeras bases de datos públicas usadas por los periodistas en Brasil.

Ascânio Saleme, ahora jefe de redacción de O Globo, es otro reportero que utilizó la contraseña de un parlamentario para llevar a cabo investigaciones en Siafi en los años 90, en colaboración con el analista económico Gil Castelo Branco, director de la organización no gubernamental [Contas Abertas](#). Estos dos casos son, probablemente, los primeros ejemplos de PGD en la historia del periodismo brasileño.

A lo largo de la década de 1990, periodistas como Fernando Rodrigues y José Roberto de Toledo, del Folha de São Paulo, comenzaron a usar técnicas de Computer-Assisted-Reporting (CAR). Gracias a los cursos impartidos por tutores del Instituto Nacional de Periodismo Asistido por Computadora de los Estados Unidos, una subdivisión de la asociación Reporteros de Investigación y Editores ([IRE/NICAR](#)), estas técnicas se difundieron y luego se convirtieron en parte del plan de estudios del programa aprendiz de Folha de São Paulo.

En 1998, Fernando Rodrigues comenzó la construcción de la base de datos [Políticos do Brasil](#), publicado en la web y en libro, actualizado hasta hoy. En 2002, José Roberto de Toledo se convirtió en uno de los miembros fundadores y Vicepresidente de la Asociación Brasileña de Periodismo Investigativo ([Abraji](#)), entidad fundamental en la difusión de los conceptos y técnicas del CAR en Brasil, que ha formado ya a más de cuatro mil periodistas. Otra iniciativa de Toledo, la agencia PrimaPagina, también fue de gran importancia en la difusión del CAR, ya que varios periodistas de datos de hoy dieron sus primeros pasos en su redacción.

Abraji participó de un seminario patrocinado por el Centro Knight para el Periodismo en las Américas en diciembre de 2002, en cual los principales oradores fueron Brant Houston, autor de un manual de CAR y luego director de IRE, y Pedro Armendares, de la Organización de Periodistas Mexicanos de Investigación, que también era uno de los tutores de los cursos organizados por la IRE/NICAR en Folha de São Paulo.

Aunque su misión es el periodismo de investigación, la Abraji ha servido en la última década, principalmente a la difusión del CAR y a la defensa del acceso a la información gubernamental, como una de las entidades que integran el Foro por el Derecho de Acceso a la Información Pública, creado en 2003. Los cursos y conferencias organizados por ellos son dos factores clave para la aparición de gran cantidad de proyectos de PGD en la década de 2010. Otras dos organizaciones no-gubernamentales jugaron un papel importante en el establecimiento de estas prácticas en las salas de redacción: [Transparência Brasil](#) y [Contas Abertas](#).

La primera fue creada en 2000 con el objetivo de construir y mantener bases de datos sobre financiamiento de las campañas políticas, la vida pública y los procesos judiciales enfrentados por los legisladores a nivel municipal, estatal y federal, así como noticias sobre la corrupción publicadas en los principales periódicos brasileños y la eficacia de los jueces de la Corte Suprema. Contas Abertas, establecida en 2005, supervisa el proceso de la ejecución presupuestaria y financiera del Gobierno, a través del monitoreo del Siafi, y promueve la formación de periodistas para examinar el gasto público.

Las bases de datos mantenidas por [Transparência Brasil](#) y [Contas Abertas](#) permitieron la realización de varios informes de investigación durante la década.

Una indicación de la importancia que las bases de datos han ganado en el periodismo brasileño es la lista de ganadores del [Premio Esso](#) de Mejor Contribución a la prensa. Lo ganó en 2002 y 2006 Fernando Rodrigues, por el archivo de datos biográficos y declaraciones de rendimientos de los políticos brasileños y por el libro "Políticos do Brasil", respectivamente. [Transparência Brasil](#) ganó el mismo premio en 2006 y [Contas Abertas](#) en 2007. En 2010, el reportaje ganador de la categoría más importante del Premio Esso fue la serie [Diários Secretos](#), publicada por la [Gazeta do Povo](#), de Paraná. Para dilucidar los movimientos de la contratación de empleados en la Asamblea Legislativa de Paraná, los periodistas construyeron una base de datos con todos los nombramientos entre 2006 y 2010, a partir de boletines impresos. Cruzando los datos en Excel, fueron capaces de descubrir los casos de contratación de empleados fantasmas y nepotismo.

El reto ahora es consolidar la práctica de PGD en las redacciones de Brasil. Como dice Dan Ariely, el periodismo de datos es como sexo en la adolescencia: todo el mundo habla de ello todo el tiempo, pero pocos lo hacen. Pocos medios, los más estructurados, disfrutan de la profusión de cifras dadas a conocer por el sector público brasileño, y aún menos redacciones producen sus propias bases de datos. El principal obstáculo es la escasa calidad de la educación brasileña, con lo que los estudiantes salen de la escuela con un bajo nivel de habilidad en matemáticas. Por otro lado, la mayoría de los universitarios de periodismo parecen sufrir de aversión a los números. Entrenar a los reporteros para utilizar hojas de cálculo y otras herramientas no trae ningún beneficio, si no saben cuáles son las operaciones matemáticas apropiadas para cada tipo de análisis de datos. Otro obstáculo es la dificultad de encontrar una manera de hacer PGD atractivo desde el punto de vista de los negocios, sin lo cual ninguna empresa va a invertir en equipos especializados.

Un caso a destacar es Estadão Dados, cuyo reportero Lucas de Abreu Maia se especializó en estadística en los Estados Unidos y, en el último año ha producido en colaboración con sus colegas dos informes basados en modelos estadísticos sólidos: uno apunta a la falta de educación de los padres como la principal causa de [mortalidad infantil en Brasil](#); el otro demostró que el partido político PMDB, cortejado por todos los candidatos presidenciales por ser el más grande en militancia en Brasil, contribuyó muy poco a la [victoria de Dilma Rousseff](#) en las últimas elecciones. Este tipo de información es el ideal que debe ser perseguido por todas las salas de redacción: en lugar de guiarse por los datos publicados por el Gobierno y otras fuentes, producir nuevos datos y conocimientos originales.

Iniciativas de Periodismo de Datos en Brasil

- [Diários Secretos](#)
- [Basômetro](#)
- [InfoAmazônia](#)
- [EcoLab](#)
- [Transparência Brasil](#)
- [Contas Abertas](#)
- [Homens de bens da ALERJ](#)
- [Políticos do Brasil](#)

Secciones de datos en diarios brasileños

- [O Globo](#)
- [Estadão Dados](#)
- [Folha de São Paulo](#)
- [Zero Hora](#)
- [Gazeta do Povo](#)

*PGD es la denominación para Periodismo de Datos en Brasil.

Marcelo Träsel (@trasel)

Periodista, Mestro en Comunicación y Información de la Universidade Federal do Rio Grande do Sul (UFRGS). Candidato a Doctor de la Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) y profesor de la misma institución. Asesor fiscal de la Asociación Brasileña de Periodismo de Investigación (ABRAJI). Investigador interesado en el periodismo basado en datos y periodismo participativo.

Sitio web: <http://trasel.com.br>

Mantiene un [listado colaborativo](#) con recursos sobre periodismo de datos.

Transparencia en Puerto Rico: anatomía de una ilusión

Farasch López

Internet ha sido no sólo un vehículo del flujo de la información sino que también ha implicado un cambio en lo que se entiende como acceso adecuado. En un principio parecía bastar que se nos permitiera un medio de distribución ilimitada o irrestricta a través del cual fluyera una narración de la información. A eso se sumó el hambre del ojo por la evidencia, llegaron así los despliegues de documentos, fotos y videos.

Entonces las suspicacias se movieron a otras partes, ya no era el dato, ni la narración sino si alguien mediaba al articular esa narración para manipular la opinión, y si ese mediador a su vez escondía algo. El reportero ya no era un contador de la historia, sino el narrador de una historia probable, ahora el lector o la audiencia, como ha terminado por nombrarse a la masa ciudadana que atestigua el despliegue de "espectáculos" noticiosos, lo veía como un intérprete. A las notas se sumaron los documentos a los que se hacía referencia. La prensa también tuvo que despertar y responder al reclamo por la transparencia. Los medios ya no estábamos exentos y teníamos que rendir cuentas. El ciudadano también ejerció el derecho de fiscalizar el material que hacía o no accesible la fuente de noticias. En medio de ese clima se privilegió lo que podían contar los datos y hacerlos accesibles permitía que cada quien interpretara la historia.

En GFR Media, grupo de medios que aglutina los dos diarios de mayor circulación del país (El Nuevo Día y Primera Hora), se creó una unidad periodística con personal destacado para atender asuntos de fiscalización a profundidad. La Unidad Investigativa es el espacio para un periodismo de largo aliento que se enfrenta a diario con los bolsillos o con las complejidades de las leyes de acceso a información. La Unidad es una inversión que el medio asumió como cosa urgente y necesaria en contraste con el periodismo en caliente o del diario que batalla con otros retos como el formato, agilidad, la competencia por la primicia y la exclusiva, entre otros.

Ese grupo, compuesto por una editora de investigación (que es en realidad un rol estratégico y táctico la mayor parte del tiempo), dos investigadores, un documentalista y dos redactores, está complementado por un equipo legal externo que hace de consultor y da apoyo a las gestiones de solicitud de información. Se creó un protocolo como guía de los trámites. El documento y la estrategia surgen tras un profundo análisis legal del estado de derecho y una reconciliación de esto con la realidad en las agencias públicas. En cada entidad pública parece haber un contingente de funcionarios ignorantes o irrespetuosos del orden jurídico que bloquea o dificulta intencionalmente el suministro de los datos. El protocolo creado toma en cuenta esa ignorancia del derecho, por lo que incluye cartas de solicitud que cuentan ya con un desglose de los estatutos. También contempla la burocracia, por lo que es entregado en más de una instancia de la jerarquía al tiempo que desglosa de manera puntual los datos solicitados.

Este protocolo también establece las vías de segundas y terceras solicitudes, con la intención de agotar los distintos recursos y prácticamente ir validando un expediente que le permita al medio prevalecer en caso de una gestión ulterior en que se vea obligado a recurrir a los tribunales. A pesar de estas provisiones, son muchas las ocasiones en las que el medio se ha visto obligado a recurrir a los tribunales a través de remedios especiales para obtener documentos como información de nómina o contratistas, documentos sobre vacaciones, viajes, gastos oficiales, facturas de servicios, entre otros, de alcaldes y otros funcionarios públicos electos.

El gusto fetichista por los datos

En este contexto jurídico y de exigencias de la prensa y del ciudadano ha surgido el gusto fetichista

por el dato. A la batalla por documentos se sumó la contienda por los números, por analizarlo todo desde el punto de vista estadístico. Las redacciones crearon una nueva trinchera, la matemática, con la que pretenden contabilizar la gestión pública.

Ciertamente los datos estadísticos son muy elocuentes y reveladores. Pero a esto se le sumó un aspecto sexy que los convirtió en una presa deseable. Las computadoras de los reporteros, que antes se bastaban con el programa de diseño en el que se diagramaba el impreso o con la plataforma en la que se subían sus notas a la web y un procesador de texto, ahora están pobladas y asediadas por múltiples herramientas que les permitan contabilizar por igual la gestión gubernamental, un concierto, un evento, una legislación, etc.

La noticia se trató de contraer a una especie permanente de censo y casi cualquier cosa se redujo a la cifra y la coquetería de las gráficas. Un extraño analfabetismo analítico y crítico empezó a instalarse. Al poco tiempo casi todo debía narrarse en cifras y visualizarse en gráficas. La investigación perdió momentáneamente su seducción porque la gerencia editorial se enamoró del fetichismo del dato o el periodismo de datos, depende de quien lo cuente.

Llovieron los premios y los ejemplos de los grandes trabajos periodísticos que parecían un caleidoscopio de los proyectos de contadores y programadores. La transparencia era el argumento principal pero terminó por quedar casi encriptada en gráficas que eran verdaderas obras de arte, un parque de diversiones, pero no una historia. No tardó mucho en que alguien redescubriera que la pornografía del número no era periodismo en sí mismo. Pero igual, la lección principal estaba aprendida, había otras formas de narrar y los datos eran una vía franca hacia la transparencia. Además, en Puerto Rico habíamos aprendido que sistemáticamente el Gobierno evadía la responsabilidad de contar con bases de datos fiables, abiertas, disponibles y en formatos procesables para poder hacer análisis.

Primer frente de batalla: el dato doméstico

Sin orquestarlo, cada medio evaluó su entorno. Todos descubrimos nuestros respectivos retos. Algunos eran de índole económica en cuanto a la adquisición de recursos técnicos y humanos, otros tenían que ver con asuntos de gerencia y otros, los más, con que lo primero que había que denunciar era la falta de datos.

Nuestro primer frente de batalla fue el dato doméstico. Tratándose de una isla y además con una tendencia casi enfermiza a interesarse sólo por su inmediatez doméstica, el reto era poner en función todo el andamiaje legal para requerir la información relevante. Superada esta parte, de inmediato surgió un nuevo y más desafiante reto: la información se entregaba en hojas de papel, por lo que tomaba semanas que fuera recopilada en las agencias, otras semanas más en fotocopiar, y meses en que fuera procesada por reporteros e introducida en bases de datos que nos permitieran cuantificar y analizar la información. Además, requería de un esfuerzo de corroboración y contraste de los datos, así como del ejercicio básico de la investigación tradicional de leyes y reglamentos aplicables a las situaciones bajo análisis.

Una muestra de cada especie

Como jefa de investigación siempre he tenido claro que es necesaria una buena dosis crítica para replantear que el periodismo de datos no es solo estadística, ni solo interpretación y mucho menos es sólo números y acceso a bases de datos libres, sino que la superación del fetichismo implica un componente indispensable de análisis de largo aliento, que abarca casi una arqueología legal, de reglamentos, legislaciones, etc. y que el esfuerzo periodístico es inútil si claudica a la visión panorámica que pone en perspectiva y en su contexto cada asunto.

Para ilustrar los retos, los aciertos y desatinos que como equipo investigativo hemos experimentado he

elegido dos proyectos modelo en los que nos servimos de distintas herramientas y que dieron resultados de periodismo de datos con distintos niveles de alcance y grados de dificultad. El propósito es destacar lo que hace de cada proyecto un ejemplar único.

[Destape al manto legislativo, cuatro años de sombra](#)

Publicado impreso, en un especial de 20 páginas, el 5 de marzo de 2013

Este es un proyecto sin precedente en Puerto Rico. Por primera vez un medio se dio a la tarea de investigar a profundidad las finanzas de todos los legisladores y publicó sus hallazgos en el primer trimestre del cuatrienio con el propósito de que la ciudadanía tuviera elementos de juicio para establecer sus exigencias a la nueva Asamblea Legislativa.

Durante más de cinco meses, un grupo de seis personas (editora de investigación, investigadores, reporteros y documentalista) se dieron a la tarea de solicitar acceso a las planillas sobre ingreso que están obligados a entregar a la Comisión Estatal de Elecciones (CEE) todos los candidatos a puestos electivos.

La primera dificultad: se nos dio el acceso a los documentos pero el reglamento de la agencia no nos permitía fotocopiarlos. De modo que el equipo se dio a la tarea de transcribir cada número de los últimos cinco años de las planillas de los legisladores.

Tener esto era como no tener nada. Sabíamos lo que reportaban en ganancias y propiedades, pero eso no era suficiente para determinar si estos funcionarios mentían en sus planillas, reportaban sus verdaderos ingresos o evadían el pago de contribuciones. Era preciso investigar con más profundidad. Exigimos el acceso a los estados financieros presentados por los funcionarios en la Oficina de Ética Gubernamental (OEG), la agencia denegó el acceso aduciendo que era información confidencial.

Activamos al equipo de abogados para que analizara las leyes y reglamentos que regían el acceso a esta información en cada cuerpo legislativo. La Cámara de Representantes y el Senado tenían Códigos de Ética distintos con limitaciones diferentes. Iniciamos las gestiones en ley para solicitar que los propios legisladores hicieran entrega de los estados financieros que radican en la OEG. El andamiaje legal promovió la dilación de los procesos.

El Nuevo Día determinó que pediría a cada legislador el acceso a sus estados financieros y haría pública esta petición. En otra acción sin precedentes, en un mismo día el diario entregó en la mano las cartas de solicitud a cada legislador. Sólo unos pocos contestaron el llamado y en su mayoría pensando en el capital político que esa apariencia de transparencia le generaría en medio de la campaña electoral que alanzaba su punto culminante para entonces. Con estos documentos empezó una nueva etapa, la de contrastar los datos que los funcionarios sometían al fisco con los que sometían a la CEE y la que presentaban en supuestos estados financieros auditados en la OEG.

El resultado, cada quien tenía su escándalo. Algunos mostraban en sus planillas al fisco un estado de insolvencia casi total (antes de ocupar el puesto público), mientras su estado financiero presentaba a pseudo magnates, reclamaban dependientes que no vivían con ellos, tenían propiedades en ejecución por falta de pago, tenían historial de demandas de cobro, algunos no tenían su residencia principal en el distrito al que representaban, no tenían ahorros, tenían gastos alegres mientras no saldaban sus préstamos estudiantiles, casi ninguno tenía cuentas para el gasto futuro de los estudios de sus hijos, tenían negocios que no reportaban, eran empresarios y legisladores al mismo tiempo, algunos tenían deudas con Hacienda sin planes de pago o sin emitir pagos por años. En fin, obtuvimos el retrato del desastre de las finanzas privadas de quienes debían administrar las finanzas públicas de un país casi en la bancarrota. Probar todo esto requirió de corroboración de datos, visitas a los predios de los que eran dueños, días de inmersión en expedientes de registros de la propiedad, entrevistas a vecinos, revisión de expedientes judiciales, entre otras muchas gestiones.

El resultado: veinte páginas de publicación de escándalos, más el esfuerzo del medio por discutir con expertos en administración pública, ética gubernamental y procesamiento penal de altos funcionarios las recomendaciones que debía poner en práctica el estado para sanear la vergonzosa situación.

El lugar de este proyecto en la selva del periodismo de datos podría resumirse en:

- No hubo un despliegue gráfico aunque indiscutiblemente este fue un trabajo monumental de análisis de datos y cifras, pero no se contaba con los elementos tradicionales en los que descansa el periodismo de datos.
- Las llamadas bases de datos estaban sustituidas por archivos cuasi arqueológicos de propiedad inmueble, escrituras de compraventa, estados financieros, etc.
- Los datos que se analizaron eran extremadamente complejos y personalísimos, es decir, había una investigación por cada sujeto, puesto que los datos a corroborar no coincidían entre sí y eran muy particulares. Cada investigador tenía una arqueología distinta que hacer. Los archivos y registros estaban regados por toda la isla.
- Fue preciso el análisis comparado de las leyes y reglamentos que rigen los cuerpos legislativos y las agencias que los fiscalizan a través de los años, lo que requirió consultas especializadas.

Esto resultó en un trabajo faraónico en el que no había ni gráficas ni números, pero fue el proceso investigativo de datos más amplio que se ha hecho en medio alguno del país.

Desierto el Capitolio

Serie impresa publicada entre el 10, 11 y 12 de febrero de 2014

- [1ra parte](#)
- [2da parte](#)
- [3ra parte](#)

Este proyecto combina el típico esfuerzo del periodismo de datos en el que se contabilizaron las ausencias y comparencias de los senadores durante la primera sesión de 2013 y un análisis profundo de los Códigos de Ética y reglamentos aplicables la conducta de los senadores. Esta Legislatura enfrentaba temas medulares que desde hace mucho se venían postergando en el aspecto legislativo pero que ocupaban la atención pública constantemente. Ese era el caso de los estipendios y dietas que cobraban estos funcionarios, los que elevaban sus salarios por encima de un promedio de \$120 mil anuales, la divulgación de información financiera, los Códigos de Ética de cada cuerpo, entre otros.

Pero no se trataba de sólo contabilizar cuándo se presentaron y cuándo no a su trabajo, sino que la Unidad sumó a este propósito el análisis en profundidad sobre el desempeño de las comisiones, a qué sesiones, vistas públicas y demás convocatorias asistió cada senador y esto lo contrastó con sus expresiones públicas sobre los distintos temas. Tampoco había precedente de un informe periodístico que trascendiera el ejercicio cuantitativo de las comparencias.

Los resultados nuevamente nos llevaron a un escándalo. Senadores con continua presencia en los medios, no acudían a las vistas, se ausentaban sin presentar ningún tipo de excusa, la senadora con mejor asistencia era de la oposición, se votaban proyectos y nombramientos de altos funcionarios de gabinete por referéndum (llamadas telefónicas), no se presentaban a las vistas de nombramiento de altos funcionarios y a algunos ni se les veía por los pasillos del Capitolio ni en los debates de mayor interés público.

Se midieron cuantitativamente aspectos como asistencia, excusas, nivel de compromiso y desempeño de cada senador. Este trabajo también enfrentó el consabido escollo de la lucha por el acceso a la información y posteriormente el desafío de procesar los datos puesto que se nos entregaron documentos en papel. En este proyecto trabajaron unas ocho personas (editora de investigación, un jefe de diseño, dos infografistas, una investigadora, un reportero redactor y el equipo legal). El proyecto en la web recogió principalmente las notas y fue la apuesta impresa la que contuvo la mayor cantidad de elementos gráficos.

Farasch López Reyloz

Editora de Investigación Unidad Investigativa GFR Media, Puerto Rico. Ha trabajado en los diarios Primera Hora y El Nuevo Día. Tiene estudios formales de Bachillerato y Maestría en Filosofía y Literatura Comparada, Lenguas Clásicas, Universidad de Puerto Rico.

En la búsqueda: periodismo de datos en Ecuador

Darwin Borja Salguero

El periodismo ecuatoriano está en evolución. Poco a poco, varios medios de comunicación están abriendo las puertas al periodismo de datos. Los que están mirando hacia ese lado, en su mayoría, son los diarios y revistas en sus versiones digitales.

Una de ellas es Revista Vistazo. Esta publicación, con sede en Guayaquil, edita el especial de las [500 Mayores Empresas del Ecuador](#). Para realizar este ranking un grupo de periodistas especializados recopila los datos de varias fuentes oficiales como el Servicio de Rentas Internas y la Superintendencia de Compañías, además de los que entregan las empresas. En su edición impresa se analizan los datos y se determinan tendencias en las distintas actividades económicas. El cuadro principal es fuente de consulta para directivos y gerentes de empresas, inversionistas, académicos, entre otros. En esa tabla se detallan cifras de ventas, de empleados, entre otros.

Pero no sólo temas económicos se están abordando con datos. Las elecciones presidenciales y legislativas de 2013 y las seccionales del 2014 se convirtieron en la oportunidad para el periodismo de datos en Vistazo.com. En el 2013 se estableció una sección especial sobre los comicios. En esa ocasión se usó la herramienta [Thinglink](#) para que los lectores conocieran a los [binomios a la Presidencia](#). Además de presentar imágenes y biografías, se podía acceder a sus planes de trabajo. En 2014 se usó Tableau Public para mostrar cómo se dividieron las provincias para esas [elecciones](#) y cuántas personas debían acudir a las urnas en febrero pasado.

Otro ejemplo es El Comercio. Este diario, con matriz en Quito, creó la Unidad de Investigación y Datos. Ahora su página web tiene una sección llamada [Data](#) en la que publica notas periodísticas con infografías interactivas. Una de las herramientas que más utiliza es Tableau Public.

En la nota [“1999: el año en que más niños nacieron en Ecuador en las dos últimas décadas”](#) se analizan las cifras de nacimientos desde 1992 hasta 2012 a través de un gráfico de línea. Lo complementa con un texto en el que se cuentan historias de familias que tuvieron sus hijos en ese año y se explican más datos del Instituto Nacional de Estadísticas y Censos (INEC).

En cambio, El Telégrafo apuesta a infografías realizadas por su propio equipo de diseño. Se destaca la que realizaron previo al Mundial de Fútbol de Brasil. En Los [23 en cifras](#) se detallan las características de los jugadores de la selección ecuatoriana, su rendimiento en la Tricolor y datos personales.

Google Maps es usada por El Tiempo en [Crímenes en Cuenca](#). A través de esa herramienta se determinan las ubicaciones de las muertes violentas en la ciudad de Cuenca en los años 2012, 2013 y 2014, se describe cómo ocurrió el hecho y un enlace a la nota publicada en su página web.

Manos a la obra: recopilando datos en Ecuador

Ante millones de datos, el primer paso es realizar una previa identificación de cuáles son los datos que una institución puede tener. Este primer acercamiento se realiza a través de una búsqueda avanzada, de revisar artículos con temas similares para identificar posibles fuentes o con una revisión de cada uno de los enlaces de un sitio web. Una de las claves es revisar el mapa del sitio. A partir de esa perspectiva inicial se determina cuáles son los caminos para obtenerlos: recopilar los que están disponibles en Internet o hacer una solicitud a la institución. En esta última opción, las limitaciones son el tiempo de entrega de la información y si cumple con las expectativas periodísticas.

Para la búsqueda avanzada en Google o Yahoo hay que buscar en sitios “gob.ec”, “gov.ec”, “fin.ec”, “com.ec”. Lo recomendable no es poner límites sino recopilar la mayor cantidad de información con la que se pueda realizar un cruce de datos para determinar su validez.

En los sitios del Estado hay una sección llamada [Transparencia](#) en la que se recopilan archivos con información administrativa, financiera, operativa, contractual y de planificación institucional. Además incluye las declaraciones juramentadas de los funcionarios.

Estos son algunos sitios desde donde se pueden obtener datos relevantes en Ecuador:

- [Instituto Nacional de Estadísticas y Censos: censos y encuestas, ente rector de la estadística nacional.](#)
- [Comunidad Andina: estadísticas comparativas de la región conformada por Ecuador, Colombia, Perú y Bolivia.](#)
- [Sistema de Indicadores Sociales: encuestas y censo.](#)
- [Banco Central del Ecuador: cifras económicas y de comercio exterior.](#)
- [Banco Interamericano de Desarrollo: economía y participación política.](#)
- [Secretaría Nacional de Planificación: inversión pública.](#)
- [Sistema Nacional de Información: datos en información relevante para la planificación del desarrollo y las finanzas públicas.](#)
- [Superintendencia de Bancos y Seguros: cifras del sistema financiero.](#)
- [Centro de Gestión Gubernamental: decretos ejecutivos.](#)
- [Superintendencia de Compañías y Valores: balances financieros y accionistas de empresas creadas en Ecuador.](#)
- [Superintendencia de Economía Popular y Solidaria: cooperativas y asociaciones.](#)
- [Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación: consulta de personas con títulos universitarios.](#)
- [Superintendencia de Telecomunicaciones: internet, telefonía, televisión pagada y concesiones de frecuencias.](#)
- [Servicio de Rentas Internas: recaudación de impuestos.](#)
- [Ministerio de Finanzas: presupuesto.](#)
- [Consejo Nacional de Electricidad: generación, distribución y consumo de energía eléctrica.](#)
- [Ministerio de Salud: hospitales y centros de salud.](#)
- [Compras Públicas: contratos del Estado para obras, bienes y servicios.](#)
- [Asamblea Nacional y Derecho Ecuador: legislación.](#)
- [Derecho Ecuador, revista jurídica:](#)
- [Voto Transparente: resultados de elecciones.](#)
- [Consejo de la Judicatura: juicios.](#)

Al momento de recopilar los datos, es recomendable realizar capturas de pantalla u organizar los archivos originales como constancia de dónde se obtuvieron.

En Excel se organiza la información, que en muchos casos corresponden a varias fuentes. Por ejemplo, en Vistazo se realizó un reportaje sobre los alcaldes electos de los veinte cantones más grandes del país. Se hizo una búsqueda en que se obtuvo datos sobre la educación a través de sus hojas de vida publicadas en el Consejo Nacional Electoral, la consulta de títulos universitarios en la página del Senescyt, y los perfiles en LinkedIn o las páginas web de campaña. En este caso se realizó un cruce de datos para determinar cuáles eran los más actualizados. Sin embargo, en caso de duda es necesario acudir a la fuente.

Las herramientas que he usado para la visualización de los datos son:

- [Tableau Public](#)
- [Thinglink](#)
- [Infogr.am](#)
- [Timeline JS](#)
- Google Drive
- [Storify](#)
- [Mapbox](#)
- [Google Maps](#)

Darwin Borja Salguero (@darwindws)

Periodista ecuatoriano de 32 años. Es Licenciado en Comunicación Social de la Universidad Central del Ecuador. Ha colaborado en la redacción Quito de Revista Vistazo, en Diario El Tiempo de Cuenca, en CotopaxiTV en Latacunga. Actualmente labora en la redacción Guayaquil de Revista Vistazo, en la sección País. En los últimos años se especializó en investigación y periodismo de datos.

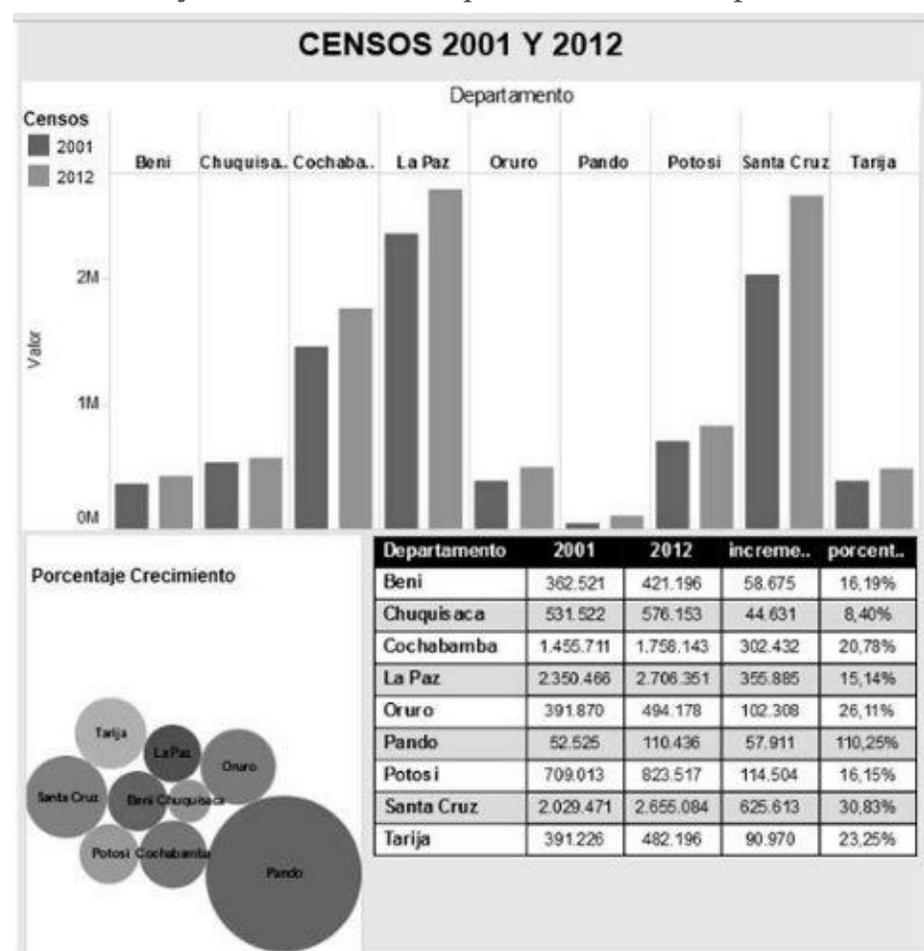
Periodismo de datos en Los Tiempos

Fabiola Chambi

Fue un impulso. Luego de asistir al primer [DataBootCamp](#) de Bolivia, que se realizó en La Paz en junio de 2013, nos interesamos más en los datos con valor e impacto para contar historias. Surgió entonces la idea de consolidar un proyecto de periodismo de datos en Los Tiempos.

Cuatro personas: el gerente de proyectos, dos periodistas y un programador, asumimos el desafío de explorar el apasionante mundo del “*open data*”, intercalando el tiempo con nuestras actividades diarias dentro del periódico. Comenzamos a capacitarnos, revisar los referentes en América Latina y diarios del mundo. Con paciencia y mucha motivación decidimos abrir el espacio en nuestro portal digital. LTDATA (Los Tiempos Data), es la primera incursión del periodismo de datos en Bolivia.

Dado el primer paso, publicamos en septiembre del año pasado nuestro primer trabajo con una visualización sobre la población mundial. Comenzamos empleando Google Maps y luego reforzamos los siguientes trabajos con Tableau Public para mostrar un estudio sobre los principales usos de Internet en Cochabamba; la distribución de escaños parlamentarios, y los datos del Censo de Población y Vivienda 2012, que recién fueron publicados este año.



Por otro lado, recurrimos a Tabula para la obtención de tablas a partir de archivos PDF complementando algunas cosas con Google Fusion Tables. Ahora estamos practicando con Carto DB para almacenar, visualizar y compartir aplicaciones con información geoespacial fácilmente; asimismo, Tiki Toki, para crear líneas de tiempo interactivas y otras de la misma naturaleza que vamos explorando.

Proyecciones

El esfuerzo del LT DATA en esta gestión se concentrará en las Elecciones Generales que se realizarán

el próximo mes de octubre. Queremos mostrar a nuestros lectores el proceso en tiempo real, para lo que nos estamos capacitando con nuevas herramientas y, sobre todo, buscando las historias que darán vida a los datos.

Nuestro objetivo es servir a quienes siguen diariamente nuestro medio a través de la web, facilitando su acceso a datos de interés y con visualizaciones atractivas y datos de descarga libre.

Ejemplo de este trabajo es la publicación que realizamos de los centros de empadronamiento establecidos por el Tribunal Supremo Electoral, que previamente hizo poca y mala difusión informativa. Entonces, el equipo del LT DATA elaboró un mapa interactivo con la herramienta [Umap](#) mostrando todos los puntos de registro habilitados en Cochabamba de una manera dinámica y accesible para la población. Este trabajo tuvo muy buena recepción y comentarios de nuestros seguidores.

Lo hecho hasta ahora apenas es el inicio de un largo camino. El proyecto involucra una constante capacitación y riesgos en cada propuesta por lo que nos mantenemos muy ligados a la ética y responsabilidad, pilares del periodismo, que se enriquecen con el uso de herramientas de visualización y análisis.

En Bolivia, todavía los datos tienen un valor limitado para el público en general, pero estamos conscientes de que el periodismo de datos no es el futuro sino el presente. Esta convicción hace que nos sintamos motivados para continuar con el trabajo y, lo fundamental, que busquemos que la ciudadanía se involucre y nos ayude también a contar las historias.

Equipo Los Tiempos Data:

- Mauricio Canelas- Gerente de Proyectos
- Fabiola Chambi- Periodista
- Michel Zelada- Periodista
- Raúl Vera- Programador

Fabiola Chambi (@fabiolachambi)

Periodista y Licenciada en Comunicación Social. Editora de Los Tiempos Digital y parte del equipo de Los Tiempos Data. Trabaja también para la revista Destinos y realizado clases de periodismo en varias instituciones.

Periodismo de datos: una práctica incipiente en México

Jesús Ibarra

La creciente penetración de nuevas tecnologías de la información, el surgimiento del movimiento Open Data, la fusión entre desarrolladores, activistas y periodistas, así como la competencia entre empresas periodísticas por nuevos nichos de audiencia se han conjugado para presentar como oferta una versión actualizada del periodismo de datos que surgió a finales de la década de 1960 con la obra de Philip Meyer. Esta nueva ola se genera principalmente en Estados Unidos, Reino Unido, Europa Occidental y Argentina, a nivel latinoamericano.

Sin embargo, a pesar de este reciente impulso, el periodismo de datos apenas se asoma en la sala de redacción de algunos diarios mexicanos de circulación y un puñado de sitios web noticiosos que operan recientemente como pequeñas y medianas empresas. Llama la atención que el principal grupo informativo, Reforma, no integra esta modalidad en sus ediciones diarias, y tampoco aparece de manera preponderante en su modelo de negocios. La Jornada, periódico de ala izquierda, y de gran tradición periodística gracias a su línea editorial combativa que suele arropar causas sociales, tampoco lo pone en práctica, y sólo apenas en julio relanzó su página web con un nuevo diseño, pero poco dinámico. El caso del semanario Proceso es una muestra tradicional de un medio enfocado sobre el periodismo de investigación en su versión impresa, pero que apuesta poco en realidad a explorar la narración de sus trabajos integrando el periodismo de datos.

Apenas hasta el año 2013 el periódico El Universal inició la elaboración de artículos de investigación haciendo un uso intensivo del manejo de bases de datos, visualizaciones, tablas dinámicas, mapas con información geolocalizada y acceso a información por medio de leyes de transparencia. A la fecha, el caso de [El Universal](#) sigue siendo el único diario nacional que mantiene esta práctica con éxito relativo y con un área creada al interior de la empresa para trabajar desde esta vertiente temas de investigación y proyectos con desarrolladores de software u organizaciones promotoras del Open Data.

Aparte de estos diarios podemos ubicar sitios web de menor tamaño, pero con una práctica periodística de calidad como [Reporte Índigo](#), [Animal Político](#) y [Sinembargo](#), que parecen animarse para abrazar al periodismo de datos en sus productos noticiosos.

En el noroeste de México lamentablemente el periodismo de datos es inexistente y apenas se asoma como asignatura en las aulas de clases de universidades locales, desde una perspectiva introductoria para una generación de pasantes de Periodismo y Ciencias de la Comunicación que tendrán como un reto más incursionar de manera decisiva en esta vertiente profesional.

En el estado de Sonora diarios de gran tradición y larga historia como El Imparcial, o más jóvenes como Expreso, son completamente ajenos a esta tendencia y con ello condenan a sus lectores a una oferta informativa que no integra y no potencia recursos para complementar sus historias y dar mayor profundidad con el uso de datos procesados integrados en la estructura narrativa de sus noticias.

Jesús Alberto Ibarra Félix

Se inició en el año 2003 como reportero investigador para el diario El Imparcial. Desde entonces ha continuado en periódicos como Cambio (2006) y Expreso (2009 y 2014). Es maestro de Ciberperiodismo en la Universidad de Sonora (2007-2014) y de Periodismo Asistido por Computadora en la Universidad Kino (2012-2014). Actualmente se desempeña como editor para el diario Expreso, en Hermosillo, Sonora, México. Candidato a Doctor en Ciencias Sociales por El Colegio de Sonora

(2010-2014).

Periodismo económico y datos: la historia que cuentan los números

Iván Weissman

Un miércoles de noviembre de 1989 Michael Bloomberg, en ese entonces CEO de la firma de análisis e información financiera que fundó y lleva su nombre, llamó por teléfono a un joven periodista del Wall Street Journal para pedirle un consejo:

"Qué se necesita para entrar al negocio de las noticias", le preguntó a Matt Winkler.

Su respuesta ya es parte de la historia del periodismo: "Tú y los cientos de empleados que trabajan contigo saben más acerca de la plata, mercados y contabilidad de empresas que cualquier grupo de periodistas. Ustedes muestran ese conocimiento diariamente en las pantallas de sus terminales en forma de números. Si de alguna forma ustedes pueden ponerle una prosa, un relato, a ese conocimiento, a esa información, y adjuntarle data a cada historia que publiquen, tendrán algo que no existe en ninguna parte del mundo", fue la respuesta de Winkler. Y así nació Bloomberg News, la empresa de medios más poderosa e influyente del mundo del dinero.

Winkler renunció al Wall Street Journal y se fue a trabajar con Bloomberg y fundó la división editorial de la empresa. Hace 24 años que es el director. Como cuenta en su libro "The Bloomberg Way" (el método Bloomberg), el objetivo era hacer el mundo de los negocios y finanzas más transparente y más democrático. Ya en ese entonces el terminal Bloomberg tenía la capacidad de analizar información y números, lo que le permitía a los operadores e inversionistas de Wall Street analizar cómo comprar los activos más baratos y vender los más caros.

¿Y por qué es tan importante lo que hizo Winkler y Bloomberg?

Porque la historia del dinero, en todas sus formas, es probablemente la gran historia de las últimas generaciones. Winkler una vez contó a un grupo de periodistas (entre los cuales estaba yo) que "Follow the money", o "¡Sigán la plata!", es probablemente el mejor consejo que le hayan dado a un periodista en la historia de la profesión. La frase se la dijo "Garganta Profunda" a Bob Woodward, reportero del Washington Post, que investigaba el caso Watergate. Su trabajo, realizado junto a su colega Carl Bernstein, provocó la renuncia de Richard Nixon, el único presidente norteamericano en la historia que se ha visto forzado a dimitir de su cargo.

El dinero -quién lo tiene, cómo se invierte, cómo se acumula y distribuye- es, nos guste o no, lo que mueve al mundo actual. El modelo de mercado es el imperante y lo más probable es que eso no va a cambiar. El mercado, bien regulado, ha demostrado ser una manera eficiente de crear riqueza, de sacar gente de la pobreza, y de un gran número de los logros extraordinarios de la civilización moderna. Por otra parte, la economía de mercado mal regulada y con poca transparencia, ha demostrado ser responsable de muchas de las grandes injusticias y desigualdades de nuestra sociedad. Bloomberg fue la primera gran apuesta a la transparencia. Ejemplos de cómo esta versión de "data reporting" ha revolucionado el mundo del periodismo financiero hay miles.

Una de las más celebradas de los últimos años es un reportaje que analizaba el mercado del oro, el petróleo y las tasas de interés de la Reserva Federal de Estados Unidos en medio de la crisis "subprime". Mirando a los datos y cifras permitió a Bloomberg reportar que lo que estaba sucediendo en esos mercados apuntaba a una intervención coordinada en el mercado de los principales bancos centrales del mundo. Al día siguiente que la nota se publicó, la Reserva Federal, el Banco Central Europeo, el Banco de Inglaterra y tres más anunciaron el recorte simultáneo de tasas de interés. Una acción sin precedentes y que tenía el objetivo de inyectar liquidez al sistema bancario para evitar un colapso financiero.

The Wall Street Journal, un medio con tanto o más prestigio que Bloomberg, se ganó un Premio

Pulitzer en 2007 por la "completa y creativa investigación" del escándalo de falsificación de fechas de *stock options*, que provocó la renuncia de decenas de altos directivos de más de 60 compañías de Estados Unidos. El reportaje "generó pesquisas, procesamientos de altos responsables y cambios masivos en empresas estadounidenses", explicó el Comité Pulitzer, al dar a conocer los motivos del galardón. Y todo comenzó con una idea muy simple de investigar ciertos informes de unos académicos que mirando a los datos observaron que las fechas en que se habían otorgado las opciones era improbable. Parecía que caían siempre en la fecha en que el precio de las acciones estaba a su nivel más bajo del año. Los académicos, con la cooperación de los periodistas, concluyeron que las fechas habían sido manipuladas para que los ejecutivos pudieran comprar las acciones a su precio más bajo y así poder maximizar sus ganancias al venderlas.

En El Mostrador Mercados hemos tratado de emular alguna de las mejores prácticas que aprendí en mis años en Bloomberg y eso nos ha ayudado a establecer credibilidad en el mercado. El año pasado una de las grandes historias financieras fue la agresiva expansión de CorpBanca, el banco controlado por el Grupo Saieh, uno de los más ricos y poderosos de Chile, al mercado colombiano. Gastó cerca de US\$ 2.500 millones comprando dos bancos locales. Fue una movida agresiva y que lo obligó a "apalancarse" de una manera que actores importantes del mercado cuestionaban. Pero no era algo que un inversionista retail hubiese podido saber fácilmente. Los Saieh también controlan Copesa, el grupo editorial que controla La Tercera y Pulso, así que esos medios no lo iban a reportar. El Mercurio y el Diario Financiero tampoco hicieron hincapié en el tema. Y ahí estábamos nosotros. Sabíamos que era verdad, pero no lo podíamos probar. Hasta que CorpBanca anuncia que emitió un bono de US\$ 800 millones para financiar su aventura colombiana.

El comunicado oficial era glorioso, pero mirando a los números de la operación y analizando la tasa que estaban pagando, pudimos comprobar que CorpBanca estaba pagando la tasa más alta de un banco chileno en los últimos tiempos por los US\$ 800 millones que emitió en Estados Unidos. Eso significaba que los inversionistas veían el riesgo de la operación y demandaban una tasa más alta que la de sus rivales para financiar la aventura al Grupo Saieh.

CorpBanca pagó una tasa de 3,24 % y el "spread" fue mayor al anticipado. En septiembre el BCI pagó una tasa de 3 % y en noviembre BancoEstado pagó 2 % por emitir deuda de características similares.

No necesitamos fuentes en off ni hacer un periodismo de investigación, sólo tuvimos que mirar a los números e interpretar la historia que escondían. Eso nos permitió construir un relato que nadie había contado y publicar una nota que daba información al lector que era única, exclusiva y que tenía uso práctico para los inversionistas.

Otro gran ejemplo de porque es clave "seguir la ruta del dinero" es el [reportaje](#) que hicimos de cómo el mismo Grupo Saieh inyectaba fondos desde su banco CorpBanca a su holding de retail que estaba en crisis, SMU.1

Lo hacía triangulando platas a través de Fondos de Inversión Privados (FIP), lo que violaba el espíritu de la Ley General de Bancos que estipula que no se puede prestar más del 5% del patrimonio efectivo del banco a una empresa relacionada. La ley fue el resultado de la crisis de 1982 cuando el país estaba prácticamente quebrado y el Gobierno tuvo que intervenir la industria bancaria, descubriendo que los grupos controladores le prestaban plata a sus diversos holdings a través de sus bancos.

Para probar nuestra nota seguimos la ruta del dinero, estudiando los Estados Financieros de SMU. Nos llamaba la atención que no tenían mucho que ver con los comunicados de prensa que el holding de Saieh mandaba. Eso despertó nuestra curiosidad para ver de dónde sacaba la plata SMU para seguir operando, ya que estaba en plena crisis. Siguiendo pistas y hablando con fuentes llegamos a los documentos de los FIP, el más relevante el de Sinergia (o Synergia) y los FIP Alpha 1 y Alpha 2. Los

documentos mostraban claramente que Saieh ponía plata en SMU a través de estos fondos de inversión, y que el origen era su banco y sus compañías de seguros.

Conseguirnos los documentos hizo posible que se publicara la nota. Ese reportaje y la serie que le siguió, gatilló una serie de eventos, incluyendo una corrida bancaria a CorpBanca, cambios en SMU, una nueva normativa de la Superintendencia de Bancos que ahora hace que las operaciones que hacía Saieh sean consideradas entre partes relacionadas e ilegales. También le provocó al grupo problemas de liquidez que lo obligaron a cancelar la Tercera TV, despedir personal en Pulso y La Tercera; y ahora lo tiene negociando para fusionar su banco y ceder el control a Itaú. El reportaje fue finalista en el [Premio Periodismo de Excelencia 2013 de la Universidad Alberto Hurtado](#).

Iván Weissman (@ivanwese)

Periodista y Editor General de El Mostrador Mercados, Chile. Ex Gerente Editorial Europa, África y Medio Oriente Bloomberg.

Experiencias con periodismo de datos en Iberoamérica

Casos de periodismo de datos en Uruguay

Maite Fernández

En Uruguay el periodismo de datos es una disciplina muy incipiente. Algunos de los grandes medios tradicionales han empezado a incursionar en esta área en los últimos años y simultáneamente han surgido nuevos medios y organizaciones de la sociedad civil que están trabajando fuertemente en este tema.

Medios y organizaciones trabajando en la apertura y el periodismo de datos

Diario El Observador

Uno de los medios que ha comenzado a experimentar con el periodismo de datos es El Observador, considerado uno de los diarios principales de Uruguay. El diario comenzó a trabajar en esta área desde hace al menos tres años, en un esfuerzo que se inició “no como una línea institucionalizada sino más bien por el esfuerzo individual de algunos periodistas”, señala Federico Comesaña, editor de Economía y Finanzas del periódico.

Lo que empezó como un equipo de análisis económico se transformó en un equipo de datos que trabaja como tal desde principios de 2014. “Si bien no funciona como un grupo de trabajo dedicado exclusivamente al periodismo de datos, se reúne periódicamente y busca compatibilizar la creación de productos de datos y el fortalecimiento de las capacidades de los periodistas para su desarrollo con sus otras tareas dentro del medio”, dijo Comesaña.

Gracias al trabajo de este equipo, El Observador publicó importantes proyectos de transparencia y rendición de cuentas públicas. Uno de estos proyectos fue [“Cuánto gastan los diputados en llamadas telefónicas”](#).

Sobre este proyecto, Comesaña explicó que el medio accedió a ese set de datos a través de un pedido de acceso a la información. “Lo que obtuvimos como respuesta fue un libro impreso de 200 páginas con el mayor nivel de desagregación de los datos. La sección que obtuvo la información consideraba que era una pérdida de tiempo destinar los recursos que se requerían para hacer inteligible tal volumen de datos”.

Fue entonces que intervino un grupo de periodistas y analistas que ya venía trabajando con datos. “Mandamos todo el material a *escaner* para que lo digitalizaran y luego, a través de software OCR, extrajimos la información. Esos datos fueron sistematizados por la Unidad de Análisis Económico y luego trabajados en conjunto por el periodista a cargo del tema, el infografista y la programadora. Esto nos permitió hacerle ver al medio la importancia que tenía dedicarle tiempo y recursos al trabajo con datos, desde un equipo interdisciplinario que abordara los proyectos junto con los periodistas de cada área, saliendo de las dinámicas del día a día y apostando por productos periodísticos de calidad”.

Otro importante proyecto realizado por El Observador fue, según Comesaña, “el desarrollo de una aplicación que permitiera visualizar el desempeño de cada legislador, considerando su porcentaje de asistencia y cada uno de los proyectos de ley que el legislador ingresó y acompañó”. Para la aplicación [“Qué hizo cada legislador en cuatro años”](#), los datos fueron complementados con los viajes realizados por cada legislador, cuánto tiempo estuvo en el exterior y cuánto dinero gastó. “Para eso se utilizaron técnicas de scraping sobre el sitio web del Parlamento y se complementaron con bases que los periodistas fueron construyendo a lo largo del tiempo”, señala el editor de El Observador.

Sudestada.com.uy

También en 2014 nació un sitio web dedicado enteramente al periodismo de datos y de investigación: [Sudestada.uy](#), que vio la luz el 5 de mayo. Conformado por periodistas con una larga trayectoria en

periodismo de investigación en Uruguay, (Pablo Alfano, Fabián Werner y Walter Pernas), Sudestada nació para llenar un vacío en la cobertura de los medios del país. “Creemos que tanto el periodismo de investigación como el de datos tienen un desarrollo insuficiente en los medios tradicionales en el Uruguay y no hay recursos en estas empresas para dedicarlos a eso por lo que resolvimos iniciar nuestro propio camino”, indica Fabián Werner.

Uno de los proyectos en los que Sudestada está trabajando es “Quién Paga. La ruta del dinero de la campaña electoral”, que participó del primer [DataBootCamp](#) de Uruguay (campamento de alfabetización en el uso de datos) y obtuvo el primer premio, entre más de [diez proyectos presentados](#).

[La aplicación](#) presenta los fondos usados por los partidos políticos para financiar sus campañas electorales en 2009. El proyecto “tiene el objetivo final de hacer lo mismo con la campaña que se realiza este año 2014 para las elecciones nacionales del próximo mes de octubre”, explicó Werner. Para ese proyecto utilizaron un set de datos de la Corte Electoral (organismo que regula las elecciones en Uruguay) con las rendiciones de cuentas de los partidos políticos. Pero, como suele suceder con los sets de datos públicos, la información estaba en un formato PDF y hubo que someterlos a un “engorroso proceso para extraer los datos y construir la base”.

DATA.uy

Como ocurre en varios países, parte del movimiento de open data y periodismo de datos está siendo impulsado no sólo por medios y periodistas, sino también por organizaciones no gubernamentales. En Uruguay, [DATA.uy](#) es una organización civil que comenzó a trabajar en forma fuerte en la apertura de datos en 2012.

Entre los proyectos más importantes de DATA.uy se encuentran el portal “¿Qué Sabés?” de acceso a la información; la co-organización de ABRE LATAM, una desconferencia regional sobre datos abiertos; la Primera Conferencia Regional de Datos Abiertos para América Latina y el Caribe, celebrada en junio de 2013 y organizada en conjunto con Presidencia del Uruguay, el Banco Mundial, Cepal, entre otros; y el próximo lanzamiento de la plataforma PorMiBarrio, basada en FixMyStreet, un sitio desarrollado por MySociety a través del cual los usuarios pueden denunciar baches en las calles, semáforos o alumbrado público en mal estado, etc.

Además de estos proyectos, DATA.uy también ha trabajado con medios de comunicación para desarrollar aplicaciones basadas en datos. Por ejemplo, la organización trabajó en 180DATA, un proyecto que realizaron junto a la sección Ciencia del portal 180.com.uy para crear aplicaciones en base a datos abiertos.

Otra de las aplicaciones fue [¿Qué tan popular es tu nombre?](#), basada en una idea original presentada en ABRE LATAM por Guillermo Moncecchi, desarrollador de software y jefe de desarrollo en la división tecnología de la información de la Intendencia Municipal de Montevideo.

“Resultó un enorme éxito de público, recibiendo decenas de miles de visitas y menciones en redes sociales,” señala Daniel Carranza, consultor en comunicación digital y co-fundador de DATA.uy. El segundo proyecto fue [Temporada de Pases](#), una aplicación que utilizó un juego de datos sobre salud y que se desarrolló en menos de una semana luego de publicados los datos. “La aplicación aprovechó un período especial que se abre para permitir el cambio de prestador de salud para mostrar datos publicados por el Ministerio de Salud Pública en forma de un ranking armado según las prioridades del usuario”, explicó Carranza.

Oportunidades y desafíos para el desarrollo del periodismo de datos en Uruguay

Los periodistas coinciden en que el periodismo de datos en Uruguay se encuentra en una etapa de desarrollo muy temprana y que aún son pocos los medios y organizaciones trabajando en esta área. “Se ven proyectos y aplicaciones puntuales interesantes pero no existe un medio u organización más

allá de DATA que tenga un proyecto sostenido al respecto”, dijo Carranza. Luego del Data Bootcamp organizado en marzo por AGESIC (organismo gubernamental que procura la mejora de los servicios ciudadanos a través de la tecnología) hay un mayor interés y se consolidó la creación de Hacks/Hackers Uruguay, dijo Carranza, “con lo que esperamos un panorama más interesante para la segunda mitad de 2014”.

Fabián Werner coincidió en que falta conocimiento entre los medios y periodistas sobre este tema. “Es bastante usual encontrar periodistas de larga trayectoria en los medios que no saben lo que es el periodismo de datos”, dijo. Federico Comesaña coincidió con esta postura y explicó que son pocas aún las redacciones que entienden la importancia de los datos como insumo fundamental para la investigación periodística.

Según Comesaña el problema comienza desde la formación del periodista en Uruguay: “existe un divorcio entre el periodista y los números que aleja a los medios del trabajo en proyectos intensivos en datos. Al mismo tiempo, no existe un vínculo entre el periodismo y las áreas de la tecnología enfocadas a la información. Los desarrolladores en Uruguay no son conscientes de su potencial de colaboración en áreas vinculadas a la apertura de datos y la relación entre la información y sus usuarios”.

Para Carranza la falta de fondos es también un obstáculo que no permite a los medios desarrollar proyectos en esta área. Un obstáculo es “la capacidad de dedicar fondos y personal en un panorama de medios muy ajustados a nivel económico. Además, algo a lo que se enfrenta todo proyecto relacionado con programación o desarrollo es el sobreempleo en el sector y los costos de los desarrolladores”, dijo.

Además de la falta de capacitación, Werner destacó la falta de interés de los medios de comunicación en desarrollar el periodismo de datos, “porque no encuentran un retorno económico y de prestigio para su marca en ello”, señaló. Otro problema es la falta de datos abiertos disponibles, tanto en la esfera pública como en organismos privados: “lo cual hace difícil que los periodistas identifiquen oportunidades de desarrollo de visualizaciones basadas en datos que resulten interesantes para las audiencias”.

Sin embargo, hay signos alentadores. En Uruguay recientemente se abrió un capítulo local de [Hacks/Hackers](#), la organización creada por el experiodista de la AP y creador de Storify But Herman y que busca unir a programadores, diseñadores y periodistas para que trabajen juntos para mejorar el periodismo. Estos capítulos, con presencia en más de 65 ciudades de todo el mundo, han probado ser una buena oportunidad para capacitar a periodistas en el periodismo de datos y atraer programadores y analistas para que trabajen en problemas inherentes al periodismo. En Montevideo, el capítulo local comenzó a funcionar en abril de 2014. “Hemos lanzado y estamos preparando nuevas instancias de formación en periodismo de datos tanto para periodistas como para diseñadores y programadores que quieran acercarse a esta disciplina”, señala Federico Comesaña, uno de sus organizadores. “La idea es generar espacios de intercambio, que presenten la labor del programador en equipos de periodismo de datos como una instancia de devolución a la sociedad, más que como una actividad de alto rédito (los medios no pueden competir con las remuneraciones del mundo del software); y también como un espacio para que los programadores con vocación hacia la información puedan conjugar sus áreas de interés”.

Medios de Comunicación desarrollando periodismo de datos en Uruguay:

[180.com.uy en asociación con D.A.T.A. y apoyo de ANII](#)

- [Temporada de pases](#)

- [¿Qué tan popular es tu nombre?](#)

Diario El Observador

Tienen una sección “Interactivos” que a veces se acerca más a periodismo de datos que otras veces, pero van en ese camino.

- [Qué hizo cada legislador en cuatro años](#)
- [¿Cuánto gastan los diputados en llamadas telefónicas](#)
- [Armá tu selección para el Mundial](#)
- [Circuitos limpios](#)
- [Aprobación de los presidentes en su último año de gestión](#)
- [Sepa dónde votar](#)

Sudestada

- [¿Quién Paga?](#)

Otras iniciativas con datos y recursos de datos abiertos

- [¿Dónde Reciclo?](#)
- [Observatorio de Políticas Públicas](#)
- [A dónde van nuestros impuestos](#)
- [Nomenclator de Calles](#)
- [Contenedores para depositar material reciclable](#)

Otras fuentes consultadas:

- [Guía para periodistas](#) elaborado por el Centro de Archivos y Acceso a la Información Pública (CAInfo) y el Banco Mundial
- [Portal de datos de la Intendencia Municipal de Montevideo](#)
- Portal de datos: <http://datos.gub.uy/> y <https://catalogodatos.gub.uy/>
- Página de Hacks/Hackers: <http://hackshackers.com/>

Para este artículo se entrevistó a Fabián Werner, director de Sudestada.com.uy; Federico Comesaña, editor de Economía y Finanzas de El Observador.

Co-autores: [equipo de D.A.T.A. Uruguay](#)

Maite Fernández (@maits)

Maite Fernández es la Directora de Comunicaciones para el Centro Internacional para Periodistas (ICFJ por su nombre en inglés), una organización sin fines de lucro dedicada a mejorar el periodismo a nivel mundial. Co-organizadora de Hacks/Hackers DC.

El proyecto “Una Vida es una Vida”

Claudia Méndez Arriaza

“Una vida es una vida” es un sitio de Internet que presenta a sus lectores tres productos esenciales: un mapa interactivo de datos para registrar las víctimas de homicidio en la ciudad de Guatemala, un blog que desarrolla notas exclusivamente sobre la violencia homicida y una sección para visualizar el análisis de datos sobre el fenómeno criminal (incidencia geográfica, perfil de víctimas, frecuencia en espacio y tiempo).

Sería redundante repetirlo, pero debo anotar que vivo en una de las ciudades más violentas de la región (67 homicidios por 100 mil habitantes). Y no puedes evadir esta condición cuando eres reportero o editor en medio de esta realidad. En ocasiones debí cubrir la hora de cierre en el diario el Periódico, donde trabajé durante 15 años, y entonces noté que la siguiente escena era repetitiva: el reportero de turno llegaba para decir que en determinada zona había un ataque y que tres personas murieron, como editora le pedía que averiguara los detalles del hecho para saber si apartábamos un espacio para nuestra edición; pero ambos sabíamos que sólo cuando se tratara de una circunstancia excepcional, íbamos a detener la producción del diario para incluir una nota más de homicidios. Aplicaba acá el principio fundamental que todos estudiamos cuando aprendemos: ¿qué es noticia?, ¿que un perro muerda a un hombre o que un hombre muerda a un perro?

Ese principio, no obstante, era cruel y desequilibrado. ¿Qué condición definía que ciertas víctimas sí ocuparan un espacio en nuestras páginas y ciertas otras no? Esta respuesta es más compleja que sencilla. Y para ser honestos, al debate debíamos añadir otra pregunta esencial: ¿qué periódico puede sostener 17 notas de homicidios en sus páginas? ¿hay uno solo cuyas páginas alcancen para registrar a cada víctima?

La respuesta a esta última pregunta es que ninguno, aunque si repetimos y repetimos y volvemos a repetir la pregunta, la solución es distinta: el periodismo en línea ofrece ahora un espacio ilimitado. Y, a medida que el periodismo de datos se desarrolla, las opciones para responder a esa interrogante se multiplican. [“Una vida es una vida”](#) nació entre 2011 y 2012, durante mi año como becaria de la Fundación Nieman en la Universidad de Harvard. Una de las condiciones de mi beca era concluir el año con una propuesta periodística. Y a lo largo de esos meses, especialmente cuando hice contacto con el Instituto de Sistemas de Análisis Geográfico, definí y desarrollé el proyecto.

No es el primero de este tipo, aunque gracias a los antecedentes, sirvió de mucho conocer los sitios que han desarrollado grandes medios y también periodistas independientes: el [mapa de homicidios](#) de The New York Times, por ejemplo, o la propuesta de [cobertura de muertes violentas](#) de la sección de datos de Los Angeles Times y, en especial, fue inspirador el trabajo [Homicide Watch](#) de la periodista Laura Amico.

La balanza, la independencia y la agenda propia

Emprender la realización del sitio implicó varios esfuerzos. Supone cambiar un marco mental: Comprender el lenguaje y la dinámica del nuevo ecosistema. No puedo codificar, pero he aprendido que un programador identificado con las motivaciones logra hacer “mover” la pantalla.

El sitio nació, se modificó y transformó a lo largo de 2011 y 2012. Su propuesta es balancear la cobertura de muertes violentas en la Ciudad de Guatemala. Al crear una base de datos que registra a cada víctima y la respuesta judicial a cada crimen, pretende generar debate y discusión abierta sobre la violencia que afecta a esta capital.

Quien navegue en el sitio podrá enterarse además de la identidad de quienes mueren de manera violenta, de las condiciones y factores que describen el fenómeno de homicidios en la ciudad

guatemalteca: horas y días, por ejemplo, o meses y evolución judicial de los casos.

En su fase inicial (enero-junio 2012) se basó en los reportes consolidados que proporcionó el Centro de Coordinación de Información Institucional (CECOIN) del Ministerio de Gobernación de Guatemala. Allí un equipo, con mística y pasión, trabaja en el registro y análisis cuidadoso para comprender las condiciones de la violencia a nivel nacional. En su fase más reciente (a partir de junio 2012), la base de datos se construyó semanalmente alimentada por los reportes diarios de novedades de Policía Nacional Civil (PNC).

La Fiscalía de Delitos contra la Vida en el Ministerio Público (MP) es clave para ampliar el alcance del proyecto. Aunque no existe un reporte público que permita conocer el avance de cada caso, a partir de entrevistas con Ricardo Guzmán, el jefe de la unidad, se identificaron variables que ampliarán la base de datos original para observar la respuesta judicial a cada crimen.

Hubo momentos clave en la concepción de este trabajo: observar y seguir sitios como el Proyecto Mapa de Homicidios de Los Angeles Times y conversar con los creadores de Homicide Watch fue determinante para completar las ideas. El apoyo del [Center for Geographic Analysis](#) de la Universidad de Harvard fue fundamental para estudiar el tema desde la perspectiva espacial. Una poderosa discusión de arte, números y periodismo, en el [International Symposium for Journalism 2012](#) fue inspiradora para propulsar la sección de análisis y visualización de datos.

¿Qué logra Una vida es una vida?

- Equilibra la balanza en cobertura de homicidios
- Desarrolla una agenda propia sobre el tema para el medio y los periodistas
- Una conexión interactiva con sus lectores

Claudia Méndez Arriaza (@cmendeza)

Reportera de investigación enfocada en casos de derechos humanos, corrupción y narcotráfico. Actualmente trabaja en Revista Contrapoder y A las 8:45 de Canal Antigua. Trabajó en el Periódico durante 15 años. Creadora del sitio www.unavidaesunavida.org, el mapa interactivo de datos sobre los homicidios en ciudad de Guatemala.

Juntando periodistas y programadores, las experiencias de Data'n'Press

Eduard Martín-Borregón

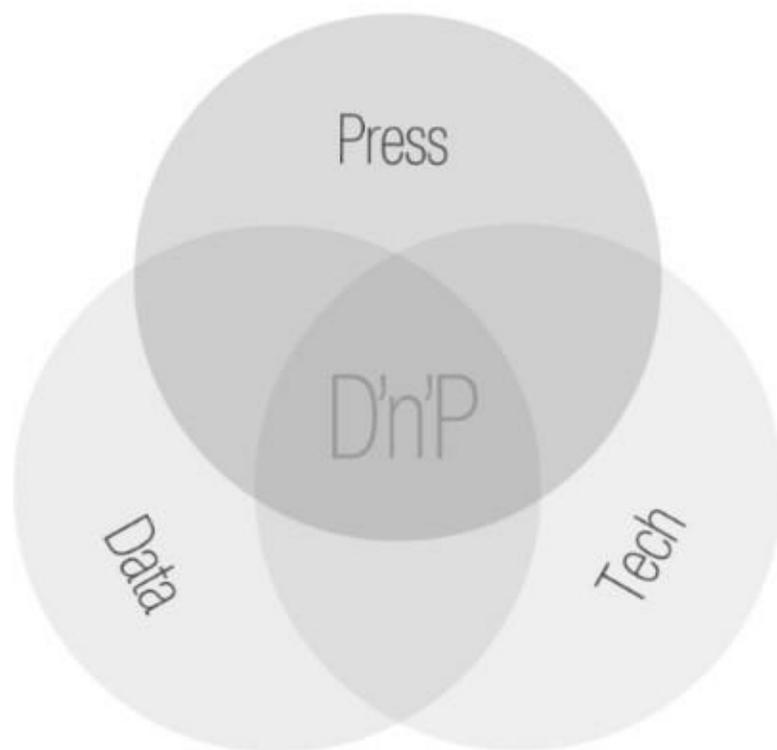
[Data'n'Press](#) es un estudio de periodismo de datos y análisis fundado en noviembre de 2011. Lo integran el Ingeniero en telecomunicaciones David Martín-Borregón, el informático Arnau Udina y el periodista Eduard Martín-Borregón. Han trabajado para El Periódico de Catalunya, el Confidencial, Vilaweb y eldiario.es, y también han realizado proyectos propios como [twitterencatala.org](#) y [@resultados20n](#).

El estudio trabaja tanto proponiendo temas a los medios, como desarrollando proyectos para clientes. Entre sus especialidades están la cobertura de elecciones, análisis y visualización de grandes cantidades de datos, trabajar con la API de redes sociales, integrar proyectos de datos con social media y el análisis de grandes textos.

Visión del Periodismo de Datos

Desde Data'n'Press vemos el periodismo de datos (cabría decir que nos sentimos más cómodos con las expresiones “periodismo de base de datos” o “periodismo inducido por datos”) como la aproximación desde el periodismo al nuevo paradigma de la sobreinformación, o mejor dicho del exceso de datos.

Internet y la digitalización de la sociedad permiten 'datizar' y guardar millones de acciones y registros que en su mayoría han sucedido siempre, pero que hasta ahora no se podían almacenar para ser consultados de forma óptima y ágil. Para convertir estos datos en conocimiento necesitamos las herramientas tecnológicas necesarias, y la interpretación o intención que le damos es periodística. Existen otras disciplinas que hacen aproximaciones distintas a este mismo paradigma, existen economistas de datos, científicos de datos o visualizadores de datos.



Nosotros describimos el proceso del periodismo de datos en tres etapas: obtención, procesado y visualización. Cada uno de estos puntos los afrontamos desde tres visiones metodológicas distintas: tecnologías a desarrollar, diseño de interface e historia periodística. El objetivo final es siempre

contar la historia de la mejor forma posible.

Los proyectos que realizamos pueden empezar en cualquier parte del proceso: tenemos unos datos nuevos que queremos trabajar, a partir de un nuevo algoritmo o proceso podemos darle una visión nueva a la historia, o querer visualizar unos datos de una forma más clara y concisa de lo que se ha dado hasta ahora. Aunque en la concepción del proyecto sea distinta, cada proyecto pasa por los tres pasos.

Obtención de datos

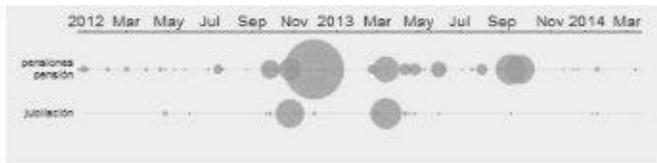
Buscamos datos tanto en redes sociales, en portales de open data o haciendo *webscrapping*. La importancia no es tanto como obtenerlos, sino conocer a la perfección el dataset. Un error en la base de datos será un error que se arrastrará en todo el proyecto y nos puede llevar a conclusiones erróneas que invalidarán el trabajo. Para evitarlo debemos repasar todos los campos y limpiarlos adecuadamente.

Conocer bien un dataset no permitirá sacar historias que no tiene nadie más. Un ejemplo de ello fue nuestro proyecto twitterencatala.org. En julio de 2012 Twitter sacó la versión en catalán de la red social, un hito dentro de la comunidad catalana que se había implicado primero en presionar a la red de microblogging y después en su traducción. La historia a contar era como su aceptación, pero Twitter no da estos datos de forma global. Lo que sí da a través de su API, es el idioma en que cada usuario tiene configurado su cuenta.

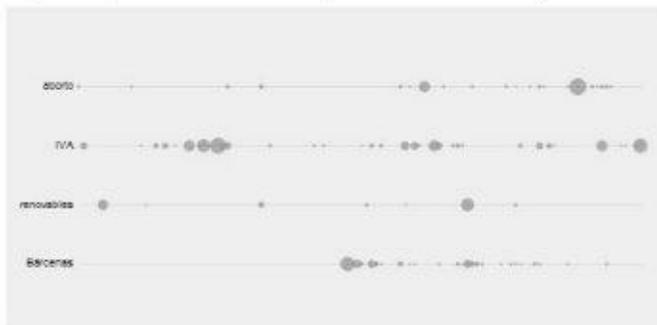
A partir de esta consulta de la API armamos un algoritmo que buscara usuarios con Twitter configurado en catalán y los fuera contando. Los resultados se daban tanto en una web como en una cuenta específica de Twitter. La aplicación tuvo un gran resultado, con más de 1000 seguidores en la cuenta de Twitter, casi 5.000 páginas vistas en la web y nuestros datos aparecen en todos los grandes medios de comunicación catalanes. Cifras muy elevadas para una comunidad tan pequeña.

Analizar y procesar los datos

Los datos nos dan aproximaciones distintas a una historia, e incluso si procesamos de distintas formas una base de datos podremos llegar a conclusiones distintas cada vez. Este es el ejercicio que hacemos con eldiario.es con la serie [“El lenguaje del Gobierno”](#) que analiza las ruedas de prensas posteriores al Consejo de Ministros del Gobierno de España. Prácticamente cada viernes el Gobierno da una rueda de prensa al finalizar el consejo de ministros. Asisten multitud de periodistas nacionales e internacionales que al finalizar la exposición realizan preguntas. La cobertura es total, lo que digan los ministros y portavoces aparecerá en diarios e informativos, pero través de los datos podemos conseguir nuevas historias.



Pensiones, pensiones, pensiones. Es una de las palabras más repetidas por la vicepresidenta y los ministros. Parte de la culpa la tiene Luís de Guindos y su reforma.



Llama la atención los equilibristas que ha tenido que hacer la vicepresidenta para no mencionar la palabra Bárceas, a pesar de las preguntas de los periodistas. Del Gobierno, solo Sáenz de Santamaría ha pronunciado la palabra "Bárceas" en ruedas de prensa. Fue en tres ocasiones. Los periodistas, sin embargo, han utilizado el nombre del exesorero del PP hasta en 83 ocasiones para preguntar al Gobierno.



Desde 1996 el Gobierno transcribe y sube al web cada una de sus ruedas de prensa. A partir de este dataset podemos ver cómo cada gobierno usa unas palabras u otras, como plantea su estrategia comunicativa o incluso que índice de legibilidad tiene su exposición. Historias que nos permiten conocer desde una aproximación distinta el Gobierno y sus políticas.

Visualizando datos

Si hay algunos datos que se han visualizado siempre en los medios de comunicación son los resultados electorales. Desde el periodismo de datos podemos intentar aportar a partir de la visualización interactiva, las redes sociales, el procesado de datos en tiempo real y agregar otros datos que permitan generar nuevo conocimiento.

En las elecciones Europeas del 2014 Data'n'Press trabajó con [El Confidencial](#) en la aplicación de resultados del diario electrónico. Las claves de la visualización, una de las más visitadas del estado, fueron: actualización en directo, desagregación hasta el detalle municipal (en directo también), datos históricos desde 1990 y gráficos que explicaban la historia de las elecciones, en aquel caso la caída del bipartidismo y el incremento de la abstención.

Uno de nuestros mantras cuando visualizamos es “la mejor visualización es la más simple” y seguramente @resultados20N ha sido la más simple que hemos realizado hasta el momento. A partir de todo el juego de caracteres ASCII creamos tuits que visualmente mostraban los resultados electorales, al más puro estilo ASCII ART. Fue volver a los orígenes, a la línea de comandos, al MS-DOS. Y es que para hacer buenas visualizaciones (y buen periodismo de datos) no es necesario usar las últimas tecnologías, sino conocer tus herramientas a fondo y usarlas de forma imaginativa.

Eduard Martín-Borregón (@emartinborregon)

Periodista de datos, cofundador de Data'n'Press. Licenciado en Periodismo con el 1r Ciclo de Economía, fotógrafo con un Máster en Dirección de Comunicación y ganador de beca Google para

estudiar el Máster en Periodismo de Investigación, Datos y Visualización.

Trabajar con datos en Perú: Comparamuni y Emergencias.pe.

César Soplín

El 26 de mayo del 2014 el Poder Judicial peruano publicó [29.486 sentencias](#) para su libre acceso a los ciudadanos y ciudadanas. Estos documentos son sobre materia constitucional, penal, civil, comercial, familiar, laboral, previsional y contencioso administrativo. Pero el sistema sólo permite la búsqueda manual. Obtener los documentos automáticamente es imposible porque la plataforma tiene instalada un [captcha](#) que impide la descarga masiva. Esto hace difícil poder explorar los documentos para la investigación, ya que sin la automatización un periodista tendría que descargar todos y cada uno de los documentos manualmente, uno por uno. Esto un programador lo podría hacer en minutos y con relativo poco esfuerzo, siempre y cuando se pueda acceder con una interfaz amigable (API).

La cuestión técnica es uno de los problemas principales, la fuente principal de datos no es accesible: el Estado aún no comprende completamente las ventajas del liberar los datos y pone trabas técnicas que impiden a los sistemas conectarse y descargar información rápidamente. Los *captchas* están sembrados en la mayoría de las fuentes de datos del Estado peruano.

Otro ejemplo: el 7 de julio del 2014 se supo que en el [Registro de Organizaciones Políticas](#) la Alcaldesa de Lima figuraba como [militante del partido de Gobierno](#), del cual sería adversaria en las próximas elecciones municipales. Esta podría ser la oportunidad para un periodista de averiguar con qué otros personajes de la política sucede lo mismo. Usando un listado de personajes se podría realizar una comparativa y encontrar diferencias rápidamente. ¡Podríamos imaginar 10 mil comparativas en un segundo! Pero sólo se puede hacer manualmente, no es posible hacer consultas automatizadas, este sistema también tiene instalado un captcha que lo impide.

Con este horizonte y sin una posición clara del Gobierno hacia los datos abiertos, pensar en un API de datos estatales es pedir mucho.

Comparamuni

En el primer hackatón de Hack/Hackers Lima participamos junto a un equipo con el proyecto [ComparaMuni](#). Buscábamos saber cuánto se gastaba en los rubros de viajes, asesorías y publicidad en el Municipio de Lima haciendo una comparación entre las cifras de la gestión actual y la anterior. Una comparación simple. La Municipalidad cuenta con una plataforma para [obtener los datos](#) lo cual ahorró bastante tiempo en la tarea, sin embargo cuando comenzamos a revisar la información encontramos que faltaban años de datos o que los rubros no eran claros. Sencillamente los datos no seguían un formato concreto, por lo que los periodistas tenían que procesar a mano esta información, para luego recién pensar en pasarla al sistema para su visualización.

Esto generó un hecho curioso. Algunas de las sumas se hicieron en Google Spreadsheets, pero automáticamente la herramienta redondeaba las cifras o interpretaba los separadores por comas como puntos (no se había configurado correctamente), situación que se advirtió al pasar las cifras al prototipo. Al final para asegurarnos que las sumas eran correctas usamos en una calculadora de toda la vida.

Finalmente obtuvimos un [prototipo base funcional](#) donde probar las visualizaciones (aún se debe mejorar), y nos entregó un aprendizaje de cómo trabajar periodistas y programadores juntos frente a los datos.

Emergencias.pe

Este proyecto surgió hace unos años como un experimento personal. La idea principal es conectar la central del Cuerpo de Bomberos del Perú (CGBVP) con Twitter. @bomberos publica las emergencias que se reportan a la central en tiempo real, tal cual se reciben. Permite a diversos diarios y

comunidades mantenerse al tanto de las emergencias de lo que sucede en la ciudad. Al momento de escribir este texto cuenta con 14,530 seguidores.

No existe un API. Estos datos se obtienen de la central de bomberos vía *scraping*. El sistema procesa estos datos y los muestra en Emergencias.pe. La idea es que los periodistas y ciudadanos accedan fácilmente a esta información sobre las emergencias que suceden en la ciudad.

La elección de la tecnología es una configuración clásica: PHP. Usando Drupal se obtiene, procesa y envía los datos a Twitter. Para la visualización de los datos se utiliza Highcharts.js y Google Maps. Todavía es un trabajo en progreso y en los planes está usar MongoDB para mejorar la velocidad al obtener los reportes.

El programador en la redacción local

En los medios digitales peruanos el trabajo con datos todavía está comenzando, se recurre a soluciones como Tableau pero aún no hay desarrollos a medida. Trabajar con datos requiere que periodistas y programadores (y diseñadores) formen juntos un solo equipo, pero en la mayoría de redacciones esto es un [choque cultural](#) que aún no se puede superar. Incluso en redacciones integradas aún no se ven desarrolladores trabajando con datos a tiempo completo.

Otras iniciativas

El estado actual de las cosas en Perú es de iniciativas aún en progreso. En el ámbito local [@aniversarioperu](#) (es un seudónimo) tiene varios proyectos de datos libre en su sitio web. También hay una iniciativa de liberar datos en la Municipalidad de Lima. Un ejemplo de app que consume estos datos abiertos es [Museos Lima](#).

César Soplín (@cesars)

Cesar Soplín Sánchez es arquitecto de información y desarrollador web. Antes trabajó en medios digitales en el Grupo El Comercio de Perú (elcomercio.pe, peru21.pe, gestion.pe, trome.pe, depor.pe) y en el sitio de Todo Noticias de Grupo Clarín en Argentina. Actualmente es profesor de Arquitectura de Información en Área 51 y consultor en Experiencia del usuario. Le apasionan los medios digitales, como se organizan y la manera en que las personas los usan.

Los esfuerzos individuales impulsan el periodismo de datos en Colombia

Por Miriam Forero y Óscar Parra

Varias prácticas relacionadas con el periodismo de datos han sido exploradas desde hace años en Colombia, especialmente por reporteros que trabajan investigaciones a profundidad. Estas prácticas han evolucionado de manera empírica y gracias a los mismos periodistas, ya que hasta hace pocos años, ningún medio de comunicación había planteado la capacitación de sus periodistas en esta área, ni mucho menos, la conformación de equipos interdisciplinarios para trabajar con bases de datos.

Existen dos estudios académicos que evaluaron el estado de desarrollo del periodismo de datos en el país. Miriam Forero ([Los primeros pasos del periodismo de bases de datos en Colombia](#)), Óscar Parra y Danghelly Zúñiga (Desafíos en la vinculación de nuevos actores en el periodismo investigativo en Colombia), analizaron casos puntuales con encuestas y entrevistas, y encontraron coincidencias relacionadas con el escaso interés de los medios a la hora de crear unidades interdisciplinarias, en contraste con el impulso que varios periodistas le están dando a esta práctica al interior de las redacciones, por su propia cuenta.

Los periodistas entrevistados en los dos estudios afirman que tienen algún tipo de conocimiento en herramientas para el periodismo de datos, especialmente para el uso de tablas, como Excel. En la mayoría de los casos, el desarrollo de esas capacidades se ha dado de manera empírica, con manuales de herramientas en Internet, consulta con expertos informáticos o tomando cursos de conceptos y métodos de trabajo con bases de datos.

En Colombia, la organización Consejo de Redacción² fue pionera en este trabajo de formación. Desde 2006 comenzó a organizar talleres de periodismo asistido por computador, que rápidamente evolucionaron en pequeños seminarios en los que reporteros regionales aprendieron las técnicas básicas y comenzaron a aplicarlas en investigaciones que realizaban en su cotidianidad en sus medios de comunicación.

Por un efecto viral y gracias a los resultados de los primeros talleres, empezó a crecer el interés en las redacciones por capacitarse en periodismo de datos. Así, hoy en día no es uno solo sino dos o tres los periodistas que conocen e intentan usar ese tipo de herramientas en medios regionales como El País, de Cali; Vanguardia Liberal, de Bucaramanga; La Patria, de Manizales; y El Espectador, Semana, Verdad Abierta, La Silla Vacía y, más recientemente, El Tiempo, a nivel nacional.

Los medios despiertan

A pesar de estas iniciativas particulares, nacidas de los mismos periodistas, el apoyo de los medios aún carece de estructura y definición: son pocos los esfuerzos para construir unidades especializadas en este campo o integrar profesionales en áreas como el diseño o la ingeniería. Según los estudios de Forero, Parra y Zúñiga, hasta 2013 sólo existían ejemplos puntuales de trabajo interdisciplinario en contados casos y siempre por iniciativa de periodistas que buscan asesoría para determinadas investigaciones; pocos medios pagan capacitaciones en periodismo de datos para sus periodistas y faltan ayudas como la adquisición de servicios informativos en línea.

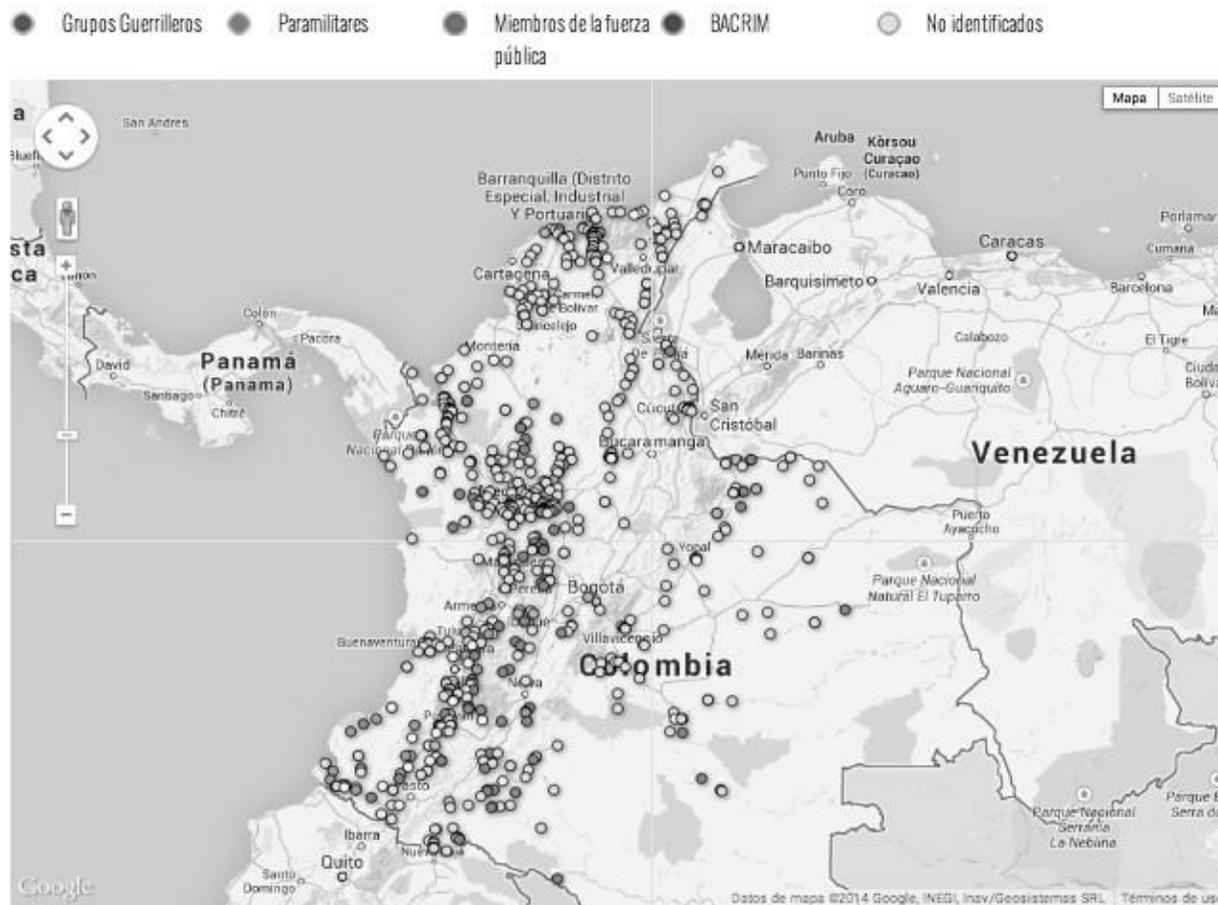
Pero las primeras luces empiezan a aparecer. Ya existen medios –principalmente los digitales– que han tomado el trabajo de datos como una línea a priorizar. La Silla Vacía creó una base de datos llamada [‘Quién es Quién’](#), para encontrar relaciones entre políticos; además, desde hace tiempo publica múltiples artículos que resultan de analizar bases de datos y que, desde 2013, aglomera en su sección [‘Hágame el cruce’](#), donde, en algunas ocasiones, los usuarios tienen acceso a los datasets en bruto, una práctica de transparencia mediática en la que este portal es pionero.

VerdadAbierta.com construyó [Rutas del Conflicto](#), una base de datos propia para organizar la

información relacionada con masacres, que el mismo medio había documentado durante más de seis años. En este caso existió un equipo interdisciplinario que incluyó ingenieros, diseñadores y periodistas, pero que sólo existió para el desarrollo puntual del proyecto.

GEOGRAFÍA DEL TERROR

Este mapa muestra las masacres que se han cometido en Colombia, discriminadas por grupos armados. Haga zoom sobre el mapa para obtener información más detallada sobre las masacres.



En enero de 2014, el diario El Tiempo, se convirtió en el primer gran medio nacional en conformar una unidad de datos encargada de asistir a las demás áreas en temas relacionados con la documentación, análisis y visualización de datos. La unidad, dirigida por Gina Morelo, una de las periodistas que comenzó a capacitarse por cuenta propia en los talleres de Consejo de Redacción, también tiene una agenda de producción propia. El periódico construyó un equipo que incluye periodistas, ingenieros y diseñadores, y ha invertido en capacitación para sus integrantes, así como para los editores de diferentes áreas.

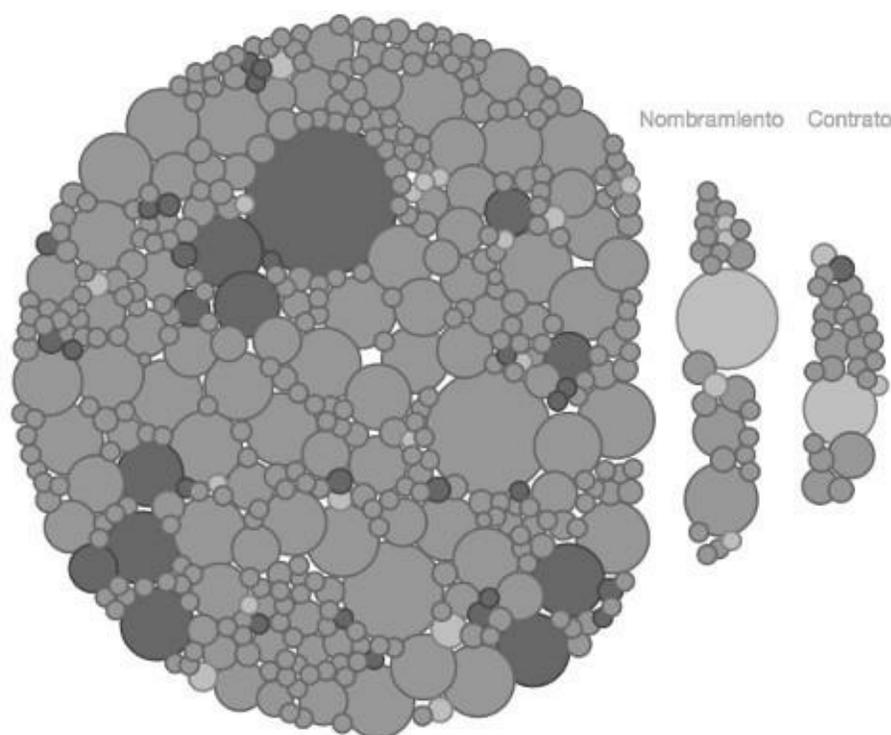
Principales prácticas investigativas

El estudio de Forero analizó varios casos de El Espectador y La Silla Vacía y reconoció cómo el uso del periodismo de datos ha enriquecido su labor de fiscalización de los poderes políticos y económicos, principalmente en la vigilancia del gasto público. Así, estos dos medios y los mencionados arriba acuden al análisis de datos para indagar temas como contratación pública, conglomerados económicos, hojas de vida y conexiones de funcionarios de Gobierno, antecedentes y sanciones de candidatos electorales, financiamiento de campañas políticas, asignación de subsidios, víctimas del conflicto, etc.

La principal herramienta, en la mayoría de los casos, es Excel, con la cual se hacen análisis y depuración de datos para luego convertirlos en visualizaciones, que generalmente desarrollan los diseñadores de cada medio. Algunos periodistas ya incursionan en aplicaciones gratuitas en Internet

como Open Refine, Timeline JS, Tableau, Fusion Tables y Google Docs. También son muy populares las consultas múltiples a bases de datos en línea para rastrear personas y empresas.

Como parte del esfuerzo individual de los reporteros en sus redacciones, algunos como [Alexánder Marín](#) (El Espectador) y [Juan Esteban Lewin](#) (La Silla Vacía), se han habituado a la práctica de construir sus propias bases de datos con información que reúnen de diversas fuentes y luego procesan. En otras ocasiones se dan alianzas puntuales para fortalecer trabajos específicos, como la de La Silla Vacía con [Aentrópico](#), una empresa de análisis y procesamiento de datos, que elabora [las visualizaciones más grandes de ese medio](#).



Otro ejemplo es el de periodistas de medios como [Semana](#) y [El País](#), que están asociados a Consejo de Redacción y que aprovechan la base de datos [Zoom Online](#), desarrollada por esa organización, en la que ya hay más de dos millones de registros relacionados con candidatos y financiamiento electoral, congresistas, personas extraditadas, estados financieros de compañías, sanciones fiscales, fondos de recursos públicos, etc. El mecanismo de estas alianzas es cruzar con Zoom grandes bloques de datos que obtienen los periodistas en su reporte, para identificar conflictos de interés, antecedentes, etc.

Otros ejemplos a destacar

[Las fichas para la reelección del Procurador](#). La Silla Vacía, medio digital. Uno de los primeros grandes trabajos con bases de datos en este medio reveló posibles conflictos de interés en la reelección del procurador. Mecanismo: construcción propia de bases de datos para reportajes puntuales.

[Fraude agrario histórico: ¡Qué tierrero!](#). [Semana](#), revista. Con el análisis de grandes cantidades de datos, la revista encontró que muchos beneficiados con la entrega de terrenos del Estado no cumplían

el perfil que la ley exige. Mecanismo: cruce con la base de datos Zoom Online.

[Los contraticos de la Cámara](#). El Espectador, periódico. Mediante contratos pequeños de personal de apoyo, la Cámara de Representantes vincula múltiples personas con sueldos altos y labores inexplicables. Mecanismo: solicitud de información pública.

[En Cali, los contratos de obras civiles tienen sus 'dueños'](#). El País, periódico. Los periodistas encontraron una concentración en las personas y empresas contratistas que realizan obras civiles en Valle del Cauca y su capital, Cali. Mecanismo: solicitud de información pública.

Miriam Forero (@Miriescribe)

Periodista independiente que se ha especializado en el análisis de datos para investigaciones periodísticas, periodismo asistido por computador y elaboración de bases de datos. Actualmente es editora de Poderopedia Colombia, proyecto en línea que mapea las relaciones de poder político y económico. Durante seis años estuvo vinculada a Consejo de Redacción -organización que promueve el periodismo de investigación en Colombia-, donde coordinó el área de investigación y dirigió la plataforma Zoom Online, una base de datos que cruza más de dos millones de registros públicos para uso periodístico. Apoya y asesora investigaciones periodísticas mediante técnicas de periodismo de datos. Es catedrática de la Maestría en Periodismo de la Universidad del Rosario y Semana, y ha sido tallerista sobre periodismo y minería de datos en diversos espacios para periodistas.

Óscar Parra

Ingeniero de sistemas de la Universidad Nacional de Colombia, con una especialización en periodismo en la Universidad de los Andes y un máster en periodismo con el diario El Mundo de España. Cuenta con ocho años de experiencia en medios digitales en portales relacionados a temas políticos y de derechos humanos como Semana.com, Votebien.com y VerdadAbierta.com. Reportero y diseñador de reportajes multimedia y programador de herramientas de bases de datos con contenido periodístico: Voto en Alerta, Votebien 2011 y Rutas del Conflicto, Verdad Abierta, 2013. Profesor de pregrado y maestría de la Universidad del Rosario en asignaturas relacionadas con la redacción periodística, herramientas digitales para periodistas y minería de datos.

Periodismo de datos transnacional para solucionar los retos de periodismo de datos en países en desarrollo

Eva Constantaras

Internews, una ONG dedicada a mejorar el acceso y la calidad de información en países en desarrollo, reconoce al periodismo de datos como una oportunidad para evitar la crisis de audiencia sufrida por muchos medios en países donde la revolución digital ya es la realidad y aún más importante, para enfrentar problemas de corrupción que impiden el progreso de estos países. Existen varios retos para realizar proyectos basados en datos por razones sociales, políticas y tecnológicas.

En muchas sociedades no existe aún una comunidad de datos abierta compuesta por periodistas, civic hackers, académicos, diseñadores y activistas que tengan la capacidad de realizar proyectos. En términos políticos, muchos medios están controlados por intereses comerciales o políticos, por lo que no les sirve promover la transparencia y el acceso a datos. En muchos casos, no existen leyes para facilitar acceso a datos ni para proteger a periodistas que publican información crítica.

El consumo de información basada en datos hasta ahora se ha concentrado en la porción de la población que tiene acceso a la tecnología adecuada. Una gran cantidad de personas en el mundo tiene acceso sólo a la radio. Esto implica que la gente más marginalizada no tiene la oportunidad de aprovechar los datos y tomar mejores decisiones para su familia y su comunidad.

Colaboración Sin fronteras en América Latina

En abril de 2011 Internews empezó un experimento en datos abiertos que esperaba presionar a los Gobiernos de América Latina a abrir más datos y fomentar una demanda por los mismos en países afectados por el crimen organizado. [A la Sombra del Crimen](#) fue una investigación transnacional entre Verdad Abierta y la Fundación InSight Crime en Colombia; El Faro en El Salvador; Plaza Pública en Guatemala y Animal Político en México para buscar lazos entre los carteles que operan en cada país usando base de datos e investigación. El resultado del proyecto, se enfocó en desplazamiento y esclavitud moderna y fue finalista para el premio de periodismo de datos de Daniel Pearl y el premio Gabriel García Márquez por innovación en periodismo. Las historias realizadas por Verdad Abierta ganaron el premio nacional para periodismo de datos en Colombia.

Sin embargo, [A la Sombra del Crimen](#) no era la investigación regional de periodismo de datos que esperábamos. En nuestra [evaluación](#) del proyecto, publicado por Universidad de Porto, descubrimos que todos los medios creían que su medio había mejorado como resultado de la colaboración, les inspiraba explorar nuevos formatos de periodismo multimedia, mejoraron sus mecanismos para planear, verificar y organizar una investigación. Pero no existía ni la seguridad, ni la confianza ni las herramientas para compartir datos e información para una verdadera colaboración en que cada medio contribuyera los datos en la búsqueda de identificar redes transnacionales de fondos, personas y conexiones políticas y comerciales. Habían logros nacionales: Animal Político en México montó un mapa de secuestrados con datos del Gobierno y [Verdad Abierta investigó el desplazamiento](#) histórico en la frontera colombiana-venezolana en una narrativa interactiva.

En búsqueda de modelos en España

Con esta experiencia, realizamos un proyecto en España con el fin de diseñar un currículo de periodismo de datos adecuada para países hispanohablantes con el apoyo de Google y el diario El Mundo. La realidad de realizar periodismo de datos resultó muy parecida que la de América Latina. No existía una tradición de colaboración entre personas de distintos medios, no existía una ley de acceso de información y los medios tradicionales no estaban dispuestos a invertir en su presencia digital.

En un reportaje publicado en [El Confidencial](#) sobre la situación del trastorno mental dentro del sistema penitenciario, se concluyó que el 41% de los presos españoles sufre algún tipo de trastorno mental, un nivel cinco veces superior a la media de la población general. El primer paso en la búsqueda de los datos fue a través de una petición de información realizada por tuderechosaber.es, una ONG que facilita el acceso a información, a la Secretaría General de Instituciones Penitenciarias. La petición solicitó el número de personas declaradas como "no culpables" por razones de trastorno mental clasificadas por sexo, edad, cárcel, enfermedad y tratamiento que reciben en la cárcel, entre otros datos.

La Secretaria General no contestó esta petición ni las siguientes, ni las llamadas. La página de estadísticas de la Secretaria de Instituciones Penitenciarias también incluye la población reclusa según situación procesal-penal por sexo, que incluye personas condenadas a "medidas de seguridad", pero sin especificar dónde las están cumpliendo. El Secretario General de Instituciones Penitenciarias, Arroyo Cabo, confirmó que solo tenía estimaciones del número de enfermos mentales dentro del sistema penitenciario. Sin datos, era imposible evaluar la gravedad del problema, las cárceles son negligentes y no tienen los recursos necesarios. La investigación que salió se basó en casos particulares de personas afectadas y las experiencias de expertos que trabajan dentro de las cárceles.

Land Quest: norte-sur colaboración en periodismo y desarrollo

[Land Quest](#) es una colaboración experimental entre periodistas particulares basados en España, Kenia y Brasil para mostrar cómo Kenia se ha convertido en un campo de batalla en el que se enfrentan dos intereses: la ayuda europea por un lado; la búsqueda del beneficios de las grandes multinacionales del petróleo y flores por el otro. Datos proporcionados por los Ministerios de Agua, Medioambiente, y Petróleo en Kenia, el Banco Mundial, la Comisión Europea, la Agencia Española para la Cooperación y varias ONG nos permitieron mapear los flujos de dinero y el impacto en acelerar la desigualdad en Kenia. Por primera vez, salió a la luz la base de datos de las [personas contratadas por Tullow Oil](#), la empresa petrolera más grande, un tema que había generado muchas manifestaciones en la zona y una militarización del área de operación, además de un contrato secreto entre el Gobierno y Tullow Oil para parar las manifestaciones.

El Mundo, el medio comprometido a publicar la investigación, solo publicó dos de los reportajes y ningunos de los gráficos y mapas analizando la situación a partir de los datos. Sin embargo, los datos han sido utilizados por periodistas en Kenia para informar el debate sobre la desigualdad y para fomentar el desarrollo de leyes que regulen la industria del petróleo.

Las ventajas de los proyectos de periodismo de datos globales

La potencia del periodismo de datos transnacional es alta en particular en los países que más sufren de corrupción y falta de transparencia. El riesgo de los periodistas locales se disminuye cuando son parte de un proyecto global. Además, después de que sale la noticia en un medio internacional, los medios nacionales están más dispuestos a publicar temas polémicos. La experiencia también ayuda a periodistas en países sin una comunidad de periodistas de datos a integrarse en la comunidad internacional y a conocer recursos para fortalecer la producción de periodismo de datos a nivel local. En nuestra experiencia, trabajar con periodistas particulares en vez de montar colaboraciones entre medios facilita mucho el proceso. Se puede identificar periodistas realmente apasionados sobre el tema y el aprendizaje de periodismo de datos y se puede formar un equipo con todas las capacidades periodísticas y técnicas necesarias para realizar un proyecto.

En el futuro, buscaremos más oportunidades de cooperación Sur-Sur para proyectos de periodismo de datos en el mundo en desarrollo. Actualmente, trabajamos en Kenia, Afganistán, Palestina, China y Sri Lanka. Temas comunes que aparecen como posibles de explorar al nivel global son: las industrias

extractivas, el medioambiente, los gastos del Gobierno y el uso de fondos de desarrollo. Con más cooperación entre la comunidad y con una audiencia internacional, el periodismo de datos tiene una alta probabilidad de cambiar el futuro del periodismo y el futuro digital de medios en países en desarrollo.

Proyectos de periodismo de datos destacables en España:

- [De Castelldefels a Santiago, cuando la seguridad ferroviaria queda en entredicho](#)
- [La demografía de la crisis: consulte la variación de los habitantes de su municipio](#)
- [El Confidencial.Lab](#)
- [Trece años de inmigración: más de 23.000 muertos por intentar alcanzar Europa](#)
- [Quién Manda](#) y [España en Llamas](#), de la Fundación Civio
- [Fuga2](#), de Eli Vivas

Eva Constantaras (@EvaConstantaras)

Periodista de datos dedicada a la producción de investigaciones transnacionales para combatir la corrupción. Actualmente, maneja programas de periodismo de datos en Kenia, Afganistán, China, Palestina y Sri Lanka sobre temas de petróleo, salud pública, economía y política. Sus proyectos se han publicado en El Mundo y El Confidencial de España y en blogs de PBS Mediashift, Open Knowledge Foundation y Knight-Mozilla OpenNews Source. Vive en Nairobi, Kenia.

Academia y periodismo de datos en Portugal: una relación feliz

Dora Santos Silva

La práctica del periodismo de datos en Portugal -o periodismo computacional, como también se suele nombrar - fue particularmente impulsada por el medio académico. Dicha colaboración entre media e investigadores ha culminado en la publicación de varios proyectos en el diario [Público](#).

El entusiasmo relacionado con las potencialidades de esta forma de periodismo nació con los ejemplos de The Guardian y del New York Times, referencias en las redacciones y clases de los futuros periodistas portugueses, y crece a través de las redes sociales y de los blogs de profesores, investigadores, alumnos y profesionales de periodismo y medios digitales. [Ponto Media](#) y [Digital Distribution](#) son algunos ejemplos.

Al mismo tiempo, iniciativas de conferencias y workshops sobre periodismo y visualización de datos en varias instituciones de enseñanza superior ([UNL](#), [ISEG](#), [ESEC](#), entre otras) y la adhesión de portugueses al MOOC sobre periodismo de datos del Centro Knight para el Periodismo en las Américas refuerzan este entusiasmo.

Sin embargo, no seamos ingenuos: el entusiasmo no se refleja en la misma amplitud en la práctica efectiva. Aunque sólo en el último año hayan nacido dos proyectos de media exclusivamente online - [Observador.pt](#) y [Expresso Diário](#) - el periodismo de datos, por el tiempo que requiere, por la influencia en las rutinas periodísticas, por el equipo multidisciplinario que se necesita y por la propia dificultad de acceso a determinados datos e incluso la dificultad de su interpretación, no es todavía una práctica corriente.

Se puede considerar una excepción el diario Público, resultante en gran parte de la colaboración con el proyecto académico REACTION (Retrieval, Extraction and Aggregation Computing Technology for Integrating and Organizing News).

REACTION (octubre de 2010 – abril de 2014) fue desarrollado en el ámbito del programa internacional UT Austin | Portugal y apoyado por la Fundação Para a Ciência e Tecnologia. Involucró investigadores de varias universidades portuguesas y tuvo como socios Sapo Labs, del portal Sapo, y el periódico Público. La periodista y becaria del proyecto, Raquel Albuquerque, trabajó en la misma redacción del periódico y estableció la conexión entre este y el equipo de investigadores.

Esta colaboración generó varias piezas periodísticas que resultaron de la exploración de datos específicos y algunos proyectos integrados en un enfoque computacional a los datos.

Algunos ejemplos de periodismo de datos en Portugal

[Floresta em Perigo](#) (reportaje multimedia presentado en Público en julio de 2013 y actualizado hasta octubre de 2013). En la multimedia “Doce años de incendios”, integrada en el reportaje Floresta em Perigo, los periodistas utilizaron datos compartidos por el ICNF (Instituto da Conservação da Natureza e das Florestas) relacionados con los incendios ocurridos entre 2001 y 2012 y los interpretaron de manera que mostraran qué años hubo más incendios, los respectivos municipios y las horas en las que se dieron los alertas para esos fuegos. Los datos fueron presentados a través de infografías interactivas. Fue particularmente importante la confirmación a través de los datos de que, por ejemplo, la alerta para cerca de diez por ciento de los fuegos anuales se dieron, en promedio, entre las dos y las tres de la madrugada, lo que fue revelador de acción humana. Este trabajo, resultante de la colaboración con el proyecto REACTON, fue distinguido con una medalla de oro en la categoría “innovación” por la Society for News Design, en el congreso ÑH.

[O financiamento das duas últimas campanhas autárquicas](#) (Público, septiembre de 2013). En época de elecciones municipales, los periodistas compilaron datos compartidos por la Entidade das Contas e

Financiamentos Político (ECFP), órgão del Tribunal Constitucional, y por el Parlamento. Se basaron en las cuentas de los partidos relativas a las últimas campañas municipales y crearon una visualización que permitió presentar los gastos y los ingresos de las respectivas candidaturas por municipio, en 2005 y 2009, y los valores de subvención pagados por el Estado a cada candidatura municipal. Este fue un trabajo pionero que contribuyó para el debate público en torno de las fuentes legales de ingresos de los partidos y del aumento del financiamiento público a las campañas. Dicho trabajo, también resultante de la colaboración con el proyecto REACTION, fue distinguido con un premio del Observatório de Jornalismo, en la categoría “Infografía digital”.

[Twittero](#) (presentado por el portal Sapo en junio de 2012 y actualizado a lo largo del campeonato de fútbol EURO 2012). Por ocasión del EURO 2012, se desarrolló en el ámbito del proyecto REACTION, en colaboración con Sapo Labs, un barómetro de popularidad de los equipos que disputaban el campeonato europeo de fútbol y de los respectivos jugadores, basado en tweets de los usuarios. Esta aplicación en tiempo real implicó la recolección de datos en larga escala. Los tweets fueron captados de todo el mundo, en un promedio de 600 mil al día. Fue también concebida una aplicación de la herramienta para *smartphones* y tabletas, de manera que fuera más fácil compartir información.

[25 de Abril, 40 años de democracia](#) (presentado en Público y actualizado a lo largo de abril de 2014). Con ocasión de los 40 años de la Revolución de los Cláveles en Portugal, Público presentó en abril un micro-site dedicado al tema, en el que reveló un conjunto de documentos originales de ese período, actas de reuniones militares y comunicados post revolución, cedidos por el Centro de Documentação 25 de Abril da Universidade de Coimbra.

Dora Santos Silva (@culturascope)

Actualmente becaria de Doctorado en Media Digitales del programa UT Austin | Portugal, con la tesis “Cultural Journalism in a Digital Environment - New Models, Practices and Possibilities”. Graduada en Ciencias de la Comunicación y Master en Cultura Contemporánea y Nuevas Tecnologías (FCSH-UNL). Es profesora invitada de la FCSH-UNL en el Master de Periodismo. Lleva diez años de experiencia como periodista y guionista en proyectos nacionales e internacionales.

Los proyectos “Alertas Universitarias” y “Cuánto ganan los rectores”

Aramis Castro

El proyecto [Alertas Universitarias](#) nació con el objetivo de informar sobre situaciones que afectan a la comunidad universitaria peruana. Se presentó bajo la campaña #AlertasUniversitarias, gracias a la plataforma Datea.pe basada en Google Maps, en la que los estudiantes reportaron y denunciaron los problemas en sus universidades. En tres meses de campaña se registraron 54 casos. La mayor cantidad de alertas vino de Lima (60%), seguida de Ica y Ayacucho.

Las denuncias estuvieron clasificadas por tipos como: problemas académicos o en la infraestructura, inseguridad y robos, corrupción o deficiente gestión de recursos, cobros excesivos, malos servicios, acoso sexual y tráfico de notas, entre otros.

La finalidad de la campaña fue solucionar problemas universitarios. Con las denuncias Corresponsales.pe preparó un informe trimestral para que las autoridades tuvieran conocimiento de la problemática. Entre algunos casos emblemáticos se encuentra el de la Universidad Nacional Federico Villarreal que se vio obligada a reducir los costos en trámites administrativos tras las protestas estudiantiles que se reportaron en el mapa. Otro caso ocurrió en Cajamarca, al norte del Perú, cuando las autoridades transitorias de una universidad renunciaron en bloque luego de que los estudiantes denunciaran que se dictaban las clases dentro de un mercado.

Con el objetivo de compartir a la comunidad universitaria los [ingresos de sus máximas autoridades](#), Corresponsales.pe diseñó una herramienta digital que permitía conocer los ingresos de los rectores y el valor de sus bienes. Para acceder a la información oficial, se usó la Ley de Transparencia. Este fue, a grandes rasgos, el proceso:

Búsqueda y obtención de la data

Ante la falta de información sobre los ingresos y bienes de las autoridades universitarias que, por ley, deberían ser publicados en los portales web de las universidades públicas, Corresponsales.pe solicitó los datos a la Contraloría General de la República, órgano de control de los recursos públicos del Perú.

El pedido se realizó en la web de la Contraloría a través de un formulario en línea.

El plazo legal para entregar la información venció (7 días hábiles y 5 de prórroga de ser notificado a quien solicita la información) y no recibimos respuesta. Tras casi dos meses de insistencias, la Contraloría nos entregó los documentos.

Procesamiento y análisis de la data

Dos reporteros ingresaron los datos -cerca de 200 páginas en formato PDF- en una hoja de cálculo. Al mismo tiempo un programador diseñó la base de datos.

Visualización y programación

Para la etapa de visualización se optó por un mapa del Perú que mostrara la región donde se ubicaba la universidad. La información incluyó la declaración jurada en PDF, el CV del rector y un gráfico que mostraba la evolución en sus ingresos.

Para el diseño de la base se decidió no usar Flash u otro formato que retrasara la descarga de la información ya que varias regiones del Perú no cuentan con buena conexión a Internet.

Difusión

Se contactó a medios regionales y representantes estudiantiles de diferentes universidades para que conocieran la herramienta y pudieran compartirla en sus redes sociales.

Principales obstáculos

- Retraso en la entrega de la información.
- Entrega de la información oficial en formato PDF (como imagen), esto imposibilitó el uso de programas de extracción de datos para un procesamiento más sencillo.
- No todos los rectores presentan sus declaraciones juradas y las universidades no brindan información al respecto.

Open Data

La hoja de cálculo se puso a disposición de los lectores para que analicen y conozcan la información completa en un solo archivo.

Cuánto dinero tiene tu universidad

Ante el desconocimiento y la poca rendición de cuentas en la comunidad universitaria; lo que muchas veces causaba protestas estudiantiles que exigían mayores recursos para sus universidades, Corresponsales.pe presentó la evolución del presupuesto público de las universidades peruanas. Con ello se evidenció que no todo el problema recae en el Estado sino que las propias autoridades universitarias no gastan los millones que tienen al inicio del año.

Las cifras en el mapa del presupuesto universitario van desde el año 2001 hasta el 2014 (dependiendo de la fecha de creación de la universidad), e incluye lo que recibieron cada año y lo que dejaron de gastar. La fuente usada fue el [Sistema Integral de Administración Financiera \(SIAF\)](#) del Ministerio de Economía y Finanzas, una herramienta del Estado que permite saber la ejecución del presupuesto público, se actualiza cada dos días e incluso permite descargar la data en formato Excel.

Como en el caso del mapa de los rectores, se apostó por un [gráfico interactivo sencillo](#) para que no hubiera problemas en la carga de la web en las diferentes regiones del país.

Aramis Castro (@aramis19)

Co-Director e investigador en Corresponsales.pe (plataforma de periodismo de datos que desarrolla temas universitarios). Especializado en las nuevas herramientas de extracción, procesamiento y visualización de la información. Expositor nacional e internacional en temas de Open Data, transparencia y acceso a la información.

Universidad Nacional de Rosario y Calles Perdidas: el avance del narcotráfico en la ciudad de Rosario

Anahí Lovato

“Somos, fundamentalmente, narradores de historias”

Al fin y al cabo, los periodistas somos narradores del mundo. A veces son historias pequeñas y simples. Otras, más complejas. Pero siempre se trata de relatos. Así lo entendemos en nuestra práctica profesional en la [Dirección de Comunicación Multimedial de la Universidad Nacional de Rosario](#), bajo la dirección del Mg. Fernando Irigaray.

El periodismo de datos es, antes que nada, periodismo. Eso quiere decir que requiere de las habilidades investigativas y narrativas más esenciales del periodismo, aunque nos permite sumar, también, la posibilidad de valernos de grandes volúmenes de datos para narrar. En este punto, compartimos la visión de [Paul Bradshaw](#), quien sostiene que los datos pueden ser tanto la fuente como la herramienta para contar la historia.

La digitalización de los datos -y, con ellos, la digitalización de la cultura- es el principio básico de nuestro actual ecosistema de medios. Un ecosistema en permanente transformación. Así lo expone Lev Manovich, para quien la representación numérica, el código binario -por lo general, invisible para los prosumidores- es el lenguaje que sustenta nuestra comunicación digital. En los medios digitales, los elementos son modulares, pueden extraerse, trasladarse y mostrarse en otros espacios, y transcodificarse: traducirse a otros formatos, sin dejar de ser nunca, en la base, largas cadenas de ceros y unos. La automatización también es un principio del lenguaje de los nuevos medios. Y es, además, un proceso que puede aparecer en diferentes estadios del periodismo de datos.

Con esas ideas en mente, comenzamos a experimentar con herramientas del periodismo de datos en el marco de las producciones digitales de la Universidad Nacional de Rosario. Empezamos por preguntarnos, cuestionarnos, hacer, errar, y aprender.

El data journalism emergió con mucha fuerza en ocasión de las ponencias del [4° Foro Internacional de Periodismo Digital](#), organizado por nuestra universidad y celebrado en Rosario en 2012. Allí, nos referimos al periodismo de datos como “un nuevo modo de pensar las noticias, sacarlas por un momento de la tradicional lógica del artículo, que aún cuando es multimedia, muchas veces es una simple yuxtaposición de lenguajes precedentes. El periodismo de datos promete no sólo abrir un nuevo modo de contar y acceder a las noticias, sino también quizá tenga una de las claves de la desesperada búsqueda del modelo de negocios que tanto persiguen las empresas”.

Fue a partir de esas primeras conversaciones que, en la mesa de producción de contenidos de la Universidad, decidimos ponernos a estudiar qué se estaba haciendo, en otras latitudes, en materia de periodismo de datos. “Scraping for journalists”, el e-book en beta permanente publicado por Paul Bradshaw, fue nuestro texto de cabecera para iniciar la exploración. Allí, el periodista sugiere que existen cuatro formas de obtener datos: recibirlos directamente de una fuente, pedirlos mediante leyes de acceso a la información, encontrarlos a través de métodos de búsqueda avanzada, y “scrapearlos”, es decir, extraerlos “rascando” recursos digitales con la ayuda de una PC. Comenzamos, entonces, por hacer pruebas básicas de *scraping* utilizando fórmulas para importar HTML en hojas de cálculo de Google Drive. Descubrimos allí un gran aliado para capturar datos, pero también, fundamentalmente, para compartir sets de datos y trabajar colaborativamente en las producciones del equipo.

Siguiendo a Bradshaw, comprendimos también que el periodismo de datos requiere de una serie de procesos desarrollados con el objetivo de comunicar historias de forma clara. Compilar, limpiar, contextualizar y combinar datos son etapas necesarias en el proceso. Probando herramientas y

aplicaciones para cada instancia, pasamos por:

- [Google Fusion Tables](#) y [Open Refine](#)
- [Junar](#)
- [Datawrapper](#)
- [Visual.ly](#), [Infogr.am](#), [Piktochart](#), [Sprites](#)
- [Many Eyes](#)
- [Tableau Public](#)
- [Google Code](#)

El tiempo dedicado a estudiar procedimientos y probar herramientas se tradujo en propuestas de periodismo de datos para el tratamiento de los temas abordados como contenido del [portal de la Universidad](#). Produjimos una serie de notas y especiales multimedia donde nos animamos a incorporar propuestas de visualización de datos como una estrategia de contextualización de la información, pensando en facilitar la exploración de las temáticas por parte de los usuarios.

Así, en el especial multimedia titulado ¿Quiénes estudian en la UNR?, nos adentramos en el análisis de cómo se compone el alumnado de la Universidad, qué variaciones ha tenido a lo largo de los últimos diez años, de dónde llegan los estudiantes que pueblan las aulas de las facultades. El especial se organizó con tres nodos informativos, donde se incluyeron entrevistas en formato audiovisual, artículos hipertextuales y dashboards de datos, organizados con Tableau Public. En el [primer nodo](#), se propone a los usuarios explorar las variaciones en el alumnado de la Universidad, por Facultad, entre los años 2002 a 2012. El [segundo nodo](#) muestra la procedencia de los ingresantes en 2012. El [último nodo](#) muestra la progresión del alumnado, con datos desagregados por género en cada Facultad.

Para la sección I+D, un espacio dedicado a la divulgación de conocimiento científico producido en la UNR, decidimos volver a aprovechar la versatilidad de los dashboards de Tableau Public para mostrar las características del mercado del suelo en la [periferia de Rosario](#). Recopilamos allí información sobre precio del suelo, diferencias entre zonas, variación anual, precio de alquileres, oferta, permisos para construcción, etc. Además, aprovechamos [Google Maps Engine](#) para georreferenciar áreas urbanas con propiedades particulares.

Luego, en casos como la nota titulada [Taller vocacional inclusivo](#), hicimos pruebas con [iCharts](#). Para la nota [Migrantes paraguayos en el Gran Rosario](#), intentamos con una infografía de desarrollo propio, creada con jQuery. Finalmente, también la herramienta [Thinglink](#) nos resultó potente para mostrar datos sobre imágenes interactivas.

Con esas primeras experiencias desarrolladas y afianzadas, organizamos en 2013 un [Curso de Formación y Capacitación profesional para periodistas](#), junto al Sindicato de Prensa Rosario. Allí, bajo una metodología teórico-práctica, con una buena cuota de taller, abordamos cuestiones como la búsqueda y acceso a base de datos en línea, la recuperación documental de textos, búsquedas inteligentes en la web (motores de búsqueda, metabuscadores, filtrado mediante operadores, búsqueda combinatoria especial, buscadores específicos por formato, buscadores académicos), aplicaciones multiplataforma y multisoporte para la organización de información a través del guardado de notas, antecedentes del periodismo de datos (periodismo de precisión y periodismo de investigación asistido por ordenador), compilación, filtrado y limpieza de datos, contextualización de la información y comunicación de los datos (visualización de datos, georreferenciación).

También desarrollamos, a partir de este año, una experiencia con estudiantes de la Licenciatura en Comunicación Social (Facultad de Ciencia Política y Relaciones Internacionales, Universidad

Nacional de Rosario), incorporando al programa académico del Seminario de Integración y Producción (modalidad: Periodismo Digital) una unidad específica sobre periodismo de datos. De este modo, el programa tradicional se completará con aprendizajes específicos del campo del *data journalism*.

Sin embargo, el desarrollo más interesante se produjo en el marco del trabajo de investigación que culminó con la publicación del DocuMedia: [“Calles Perdidas, el avance del narcotráfico en la ciudad de Rosario”](#). Se trata de un documental multimedia interactivo –el cuarto DocuMedia producido por nuestra Universidad- donde se conjugan técnicas del documentalismo, el periodismo de investigación y el periodismo de datos, produciendo un conjunto de recursos integrados en una plataforma interactiva. El trabajo fue recientemente reconocido con el Premio Internacional Rey de España 2013 en la categoría Periodismo Digital.

Con “Calles Perdidas” propusimos un recorrido por la trama del narcotráfico en la ciudad de Rosario, analizando la realidad de los distintos barrios y repasando la complejidad de actores sociales implicados. Uno de los principales aciertos de ese trabajo fue, probablemente, la idea de incluir un mapa interactivo donde nos ocupamos de georreferenciar los homicidios y hechos violentos ocurridos en la ciudad desde enero de 2012 a la actualidad.

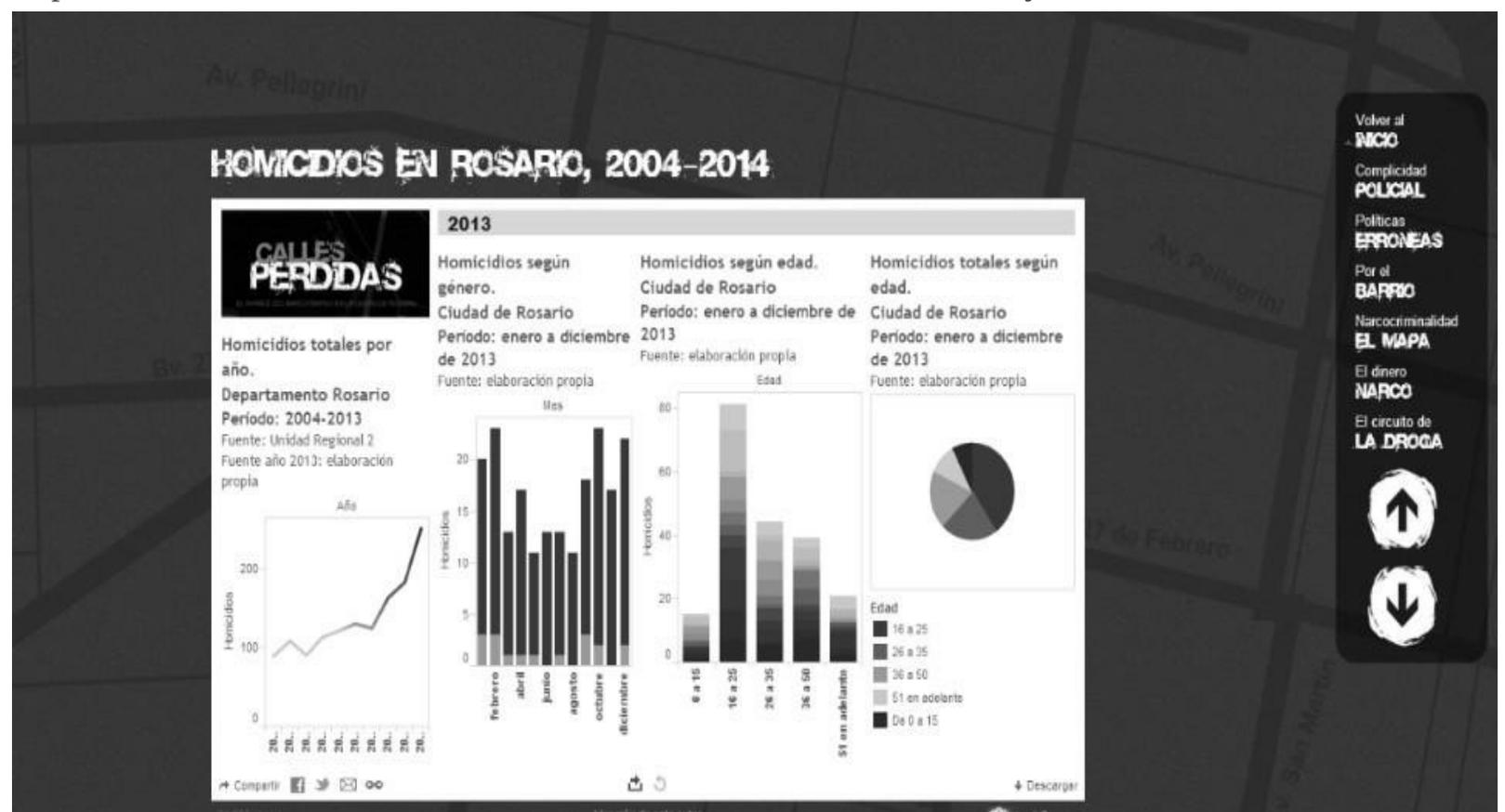
Para el desarrollo del DocuMedia se conformó un equipo multimedia con profesionales de diferentes perfiles y experiencias, incluyendo periodistas, realizadores audiovisuales, diseñadores, desarrolladores web, infografistas, fotógrafos e ilustradores. Trabajaron en “Calles Perdidas”: Fernando Irigaray, Marcelo Colman, Patricio Irisarri, Matías Loja, Anabela Alemanno, Gisela Moreno, Anahí Lovato, Andrés Aseguiolaza, Agustín Pagliuca, Alejandro Coscarelli, Martín Pérez, Joaquín Paronzini, Aldo Iñiguez, Marcos Riganti y Sebastián Suárez Meccia.



Hasta el año 2012, la Unidad Regional 2 de Policía de Santa Fe había hecho públicos los datos de homicidios totales ocurridos en el departamento Rosario. Sin embargo, las crónicas policiales diarias

daban cuenta de una escalada de homicidios violentos en los barrios de la ciudad. Esos crímenes se comunicaban como hechos aislados y estaban desperdigados por los portales de noticias de la región, caratulados como “ajustes de cuentas”.

Al interior del equipo periodístico de la Universidad, nos propusimos, luego, sistematizar la información dispersa y organizarla en una tabla de datos que diera cuenta del número de víctimas que se producían mensualmente. Resolvimos identificar, también, el sexo y la edad de las víctimas.



Además, decidimos ubicar esos hechos sobre un mapa de Rosario. Esa estrategia nos permitió ver que la totalidad de la ciudad está implicada en la problemática. Sobre cada hecho marcado, agregamos una descripción breve y un enlace a la noticia donde se refería el homicidio. Comenzamos a observar, de este modo, que muchos homicidios se producían utilizando la misma metodología, en un radio de pocas cuadras de distancia.

Decidimos agregar otras capas de información al mapa, sumando a los homicidios el conjunto de acciones policiales y vecinales producidas en relación al narcotráfico: incautaciones, destrucción de bunkers, detenciones. También añadimos las jurisdicciones policiales, es decir, las áreas que competen a cada comisaría de la ciudad. Esa transposición de datos nos permitió observar, claramente, las relaciones entre homicidios, balaceras y localización de kioscos y bunkers; e identificar los lugares más “calientes” y su relación con la competencia policial sobre cada zona.

La tecnología de Google Maps fue el sustento utilizado para la georreferenciación. A ella se sumó un desarrollo con [jQuery](#) para permitir a los usuarios interactuar con la información de las capas, filtrando datos y seleccionando diferentes opciones de visualización. Definitivamente, esa tarea de sistematización y visualización de datos marcó la diferencia entre ver los fragmentos de una realidad problemática y armar el rompecabezas. Por otra parte, el filtrado de datos por edad y género nos permitió evidenciar que cerca del 50% de los muertos en Rosario son varones menores de 25 años.

Para mostrar esa información decidimos utilizar una doble estrategia, siguiendo la propuesta conceptual de [Alberto Cairo](#): las infografías nos sirven para explicar; la visualización de datos, para explorar”. Entonces, organizamos dashboards de gráficos interactivos que se actualizan

mensualmente.

Luego reforzamos esa visualización con piezas infográficas animadas, diseñadas para destacar datos particulares. El diseño de infografías interactivas nos permite explicar relaciones entre datos puntuales puestos en contexto. Los sets de datos, en cambio, ponen al alcance de los usuarios la posibilidad de explorar por sí mismos las propiedades del fenómeno.

En su conjunto, ciertamente, los procedimientos del periodismo de datos le dieron sustento denso al relato de una historia ardua y compleja como la que contamos en DocuMedia: “Calles Perdidas, el avance del narcotráfico en la ciudad de Rosario”.

Anahí Lovato (@anahilo)

Lic. Anahí Lovato. Coordinadora de contenidos multimedia (Dirección de Comunicación Multimedial, Universidad Nacional de Rosario). Docente en Seminario de Integración y Producción (Licenciatura en Comunicación Social, UNR). Guionista e investigadora en proyectos transmedia. Community Manager.

Periodismo de datos en Puerto Rico

Natalia Bonilla

La falta de tiempo y escasez de fondos son dos factores que constantemente afectan el ejercicio del periodismo de datos en Puerto Rico. A mitad del siglo 20, periódicos como El Mundo, el San Juan Star y El Nuevo Día destacaron con investigaciones sobre corrupción gubernamental e irregularidades en las operaciones de empresas grandes del país.

Según Israel Rodríguez, autor del libro [Escándalos políticos y el periodismo en Puerto Rico](#), desde 1962 se han publicado múltiples historias meritorias de distinción: [la contratación de 300 empleados “fantasmas” en la Legislatura](#); [los asesinatos de dos jóvenes independentistas en el Cerro Maravilla en el 1978](#); [la desaparición del Girod Trust Company](#); [las transacciones fraudulentas del empresario Andrew Cipollo](#); la excarcelación ilegal de los asesinos de los independentistas en Maravilla; y la condonación gubernamental de su deuda millonaria a una compañía privada, [la Caribbean Petroleum Corporation \(CAPECO\)](#).

Sin embargo, así como se han trabajado grandes historias, hay muchas otras que siguen un lento proceso o simplemente, permanecen desconocidas, por la invariable dificultad de acceso a la información pública.

“En Puerto Rico hay un problema de acceso pero también de confiabilidad en los datos que sí son accesibles. En el proceso de trabajar con ellos, hay que entender sus limitaciones y la necesidad de seguir ampliando la información disponible para poder corroborar lo que pensamos es la realidad”, expresó Alvin Quiñones, cofundador del [Centro de Investigación y Política Pública \(CIPP\)](#).

Por su parte, Edgar Ríos, oficial de proyecto de Abre Puerto Rico -una iniciativa cibernética del CIPP lanzada el 5 de marzo de 2014 para fomentar la transparencia gubernamental-, explicó que usualmente los empleados públicos desconocen qué información contiene los documentos que se solicitan por lo que dudan cómo proceder, temiendo que si acceden, la información publicada ponga en peligro sus empleos o sea utilizada para atacar al Gobierno.

“Es mucho más fácil que el Gobierno le dé información a un ciudadano que a un representante de una organización”, destacó. “Estamos a merced de los datos que nos provea el Gobierno” lo cual aseveró que, en ocasiones, imposibilita garantizar la fiabilidad de los documentos.

La Directora Ejecutiva del Centro para la Libertad de Prensa en Puerto Rico, Helga Serrano, recordó un evento cuando, mientras ejercía como reportera del desaparecido rotativo El Mundo, recibió documentos públicos incompletos.

“Recuerdo que cuando yo estaba en El Mundo fuimos a Corte para pedir unos documentos sobre los viajes del entonces gobernador Rafael Hernández Colón (1973-1977), y, finalmente, los dieron. Pero había mucho texto que no podía leerse (con tachaduras) y el Gobierno alegó que era relacionado con seguridad nacional”, relató Serrano.

Esta tendencia se une también a otras violaciones al derecho de libertad de prensa e información, cobijado bajo el [Artículo 2, Sección 4, de la Constitución del Estado Libre Asociado de Puerto Rico](#). De acuerdo con Serrano, resaltan los siguientes atentados recientes: [la cobertura de la salida de la Marina en Vieques, cuando fueron arrestados varios periodistas en el 2003](#); en el 2006, [las agresiones a periodistas por parte de agentes del Negociado Federal de Investigaciones \(FBI, por sus siglas en inglés\)](#), que tuvo un largo pleito judicial que terminó a favor de los agentes; y [el cierre del Hemiciclo del Senado a periodistas](#), que fue ordenado por el entonces presidente de ese cuerpo legislativo, Thomas Rivera Schatz, en el 2010, entre otras.

Mientras, Serrano enfatizó en la necesidad de fomentar la lectura crítica de medios y la

responsabilidad que tiene el ciudadano de fiscalizar al Gobierno; Ríos destacó la importancia del periodismo de datos para entender la crisis económica en el país que se ha extendido desde el 2008 hasta el presente. Recientemente, tanto Quiñones como Ríos colaboraron con el periódico El Nuevo Día para producir una serie investigativa sobre irregularidades en los presupuestos gubernamentales en los últimos 30 años.

“El periodismo de datos le provee cierta continuidad a los distintos temas que se discutan. Si nosotros revisamos los datos, podemos identificar patrones y discrepancias. Ahí hay historias que deben contarse”, señaló Ríos.

Otros datos importantes

Aparte del derecho constitucional de acceso a la libertad de prensa y expresión (de la cual se desprende una interpretación a la libertad de información), en 1955 se aprobó la [Ley 5](#) para la Administración de Documentos Públicos en Puerto Rico. Sin embargo, los periodistas y los ciudadanos también recurren al [Freedom of Information Act](#) -que aplica a Puerto Rico por su relación política con los Estados Unidos- y a casos judiciales como precedentes para validar sus solicitudes. Casos recientes incluyen [Romeu Matta v. PRIDCO](#), [Ortiz v. Administración de Tribunales](#) y [Angueira Navarro v. Junta de Libertad Bajo Palabra](#) (2000).

Natalia A. Bonilla Berríos (@nataliabonilla)

Periodista y productora. Posee una maestría en Relaciones Internacionales de la Universidad de York, Inglaterra, y un bachillerato en Información y Periodismo de la Universidad de Puerto Rico, Recinto de Río Piedras. Se especializa en derechos humanos, conflictos étnicos y libertad de prensa. Durante cuatro años, ejerció como directora de proyectos de Grupo Latitudes, una organización sin fines de lucro gestora de proyectos periodísticos, culturales y humanitarios de índole internacional. Ha producido y dirigido los documentales Haití espera (2011) y Ecos del Exilio (2013). También, colabora con el Centro para la Libertad de Prensa en Puerto Rico como una columna bimensual sobre los ataques y proyectos de libertad de prensa en el mundo.

Blog: www.consciente.wordpress.org

SocialTIC y la Escuela de Datos

Sergio Araiza

A inicios de 2013, en el equipo de [SocialTIC](#) decidimos enfrentar una tarea que consideramos fundamental para el habilitamiento tecnológico de los actores de cambio en América Latina: la formación en el uso de datos. Después de varios años impulsando el empoderamiento digital era cada vez más evidente que para poder lograr articular acciones de incidencia basada en información y argumentos, se debían trabajar capacidades básicas de análisis y manejo de datos. A medida que las corrientes de apertura de datos crecieron, también lo hicieron las exigencias para periodistas y activistas en poder hacer uso óptimo de la información accesible y generable utilizando las nuevas tecnologías, que se ven frenadas por una formación tradicional que no sólo está distanciada de las herramientas útiles para gestiones de datos sino que no fomenta que se forjen nuevas narrativas propias en ambientes digitales.

Al revisar muchos de los espacios existentes para aprender a obtener, manejar, analizar y visualizar datos, notamos tres grandes deficiencias en los ecosistemas de recursos sobre datos. Primero, la mayoría de los recursos disponibles en Internet no están en nuestro idioma, excluyendo a miles de actores de cambio en América Latina. Segundo, las referencias prácticas en el uso de datos eran extranjeras, principalmente anglosajonas y gestadas en contextos tecnológicos, institucionales y de acceso a datos, muy distintos a los que se viven en América Latina. Y por último, los recursos didácticos se encuentran en tantos distintos espacios digitales y comunidades de práctica que el acceder a contenidos relevantes para actores de cambio implica un ejercicio de inherente búsqueda y curación de recursos.

Ante lo anterior, nuestra primera reacción fue la de crear una currícula y contenidos propios con base en la experiencia didáctica inicial formando a activistas e impulsando el hacking cívico. No obstante, al conocer el proyecto School of Data de Open Knowledge Foundation, optamos por sumar esfuerzos e iniciar la traducción de contenidos para formar [Escuela de Datos](#). El proyecto replicó la currícula inicial, la lógica de aportación comunitaria y los enfoques didácticos de la versión en inglés.

La currícula inicial de Escuela de Datos abarca muchos de los aspectos más fundamentales (y necesarios) a los que se enfrenta una persona que trabaja con datos. Los temas generales que se abarcan son:

- Fundamentos de datos
- Cómo acceder a datos
- Extracción de datos de PDF, web y otros formatos cerrados
- Limpieza de datos
- Manejo de datos georeferenciados
- Visualización de datos
- Gastos y presupuestos
- Narrativas basadas en datos

Adicional a los recursos, el proyecto acuña la "expedición de datos", un concepto didáctico que ha sido rápidamente adoptado y posteriormente "hackeado" en América Latina. Una expedición de datos consta de una sesión de entre 4 y 6 horas que une a un grupo multidisciplinario (analistas, programadores, diseñadores, investigadores y narradores) para poder dar solución a una pregunta básica en el trabajo con datos: ¿Qué dicen los datos? Lo relevante de la expedición son todos aquellos

procesos de aprendizaje e intercambios de enfoques que se dan entre personas con formaciones distintas, bajo una dinámica de análisis constante de una serie de datos que normalmente implica enfrentar retos técnicos (ej. extracción, limpieza, manejo de datos, etc.), analíticos (ej. correlación de variables, validación estadística, etc.) y contextuales (ej. verificación de información, identificación de relevancia de la información, etc.). Además, después de horas de análisis se refuerza o descarta la hipótesis sobre aquello que se está capturado por los datos, se profundiza sobre el tema en cuestión y se tiene mayor agudeza sobre qué se puede hacer con esa información (ej. tácticas para la incidencia, investigaciones periodísticas, visualización de datos, etc.).

Escuela de Datos, además del contenido inicial, acuñó dos principios que consideramos fundamentales para el contexto latinoamericano: el tener una marca común y desvinculada a instituciones de la región y que su utilidad está basada en la participación de los individuos y organizaciones que integren la comunidad.

A medida que Escuela de Datos y sus subsecuentes versiones lingüísticas en distintas partes del mundo crecen, también aumentan y se diversifican contenidos y enfoques para la enseñanza de datos. Por ejemplo, desde México hemos definido una metodología para integrar la enseñanza de distintas herramientas para el manejo de datos como parte de una expedición de datos. En Francia, han buscado llevar los contenidos a públicos infantiles con tal de formar en materia de datos a las nuevas generaciones. Y a medida que aumentan las colaboraciones hispanoamericanas, se incrementa la exposición de los proyectos regionales, inspirando y formando a grupos que trabajan con datos o están por hacerlo.

Participar en la Escuela de Datos es sencillo. Si ya trabajas con datos, comparte tus experiencias y proyectos escribiendo en el blog. También puedes ser parte de la red de especialistas que desarrollan contenidos nuevos y resuelven dudas técnicas en la comunidad regional e internacional. Si ya entrenas en el uso de datos, participa en los llamados para fellows o forma un capítulo de la Escuela De Datos para apoyar de manera constante a los actores de cambio de tu país. Y si estás aprendiendo, simplemente participa en las redes sociales y en las comunidades dateras enviando dudas y sugiriendo mejoras a contenidos y recursos existentes.

Sergio Araiza (@mexflow)

Licenciado en Ingeniería en Sistemas Computacionales por el Instituto Politécnico Nacional - ESCOM Escuela Superior de Computo. En el ámbito social es líder de activismo digital y seguridad informática para organizaciones de la sociedad civil y grupos de activismo. Es Miembro de Internet Society México, SocialTIC y Escuela de Datos.

Mapas y dibujos que cuentan historias

Infografías y visualizaciones que simplifican en exceso o confunden datos

Aberto Cairo

Exposición en Chicas Superpoderosas, Miami

El origen de esta presentación viene de una plática que tuvimos mi amiga Mariana y yo. Un día me dijo: ¿podrías dar una plática en Chicas Superpoderosas? Y yo dije: por supuesto que sí... pero ¿una plática sobre qué? Obviamente pensé en visualizaciones e infografía. Ella me contestó: puedes escoger el tema que quieras. Entonces pensé y pensé en puras estupideces, sinceramente.

Esto fue lo que hice: me aislé de todo, de redes sociales, ruidos, de todo lo que me distrajera. Cuando quiero aislarme voy a las librerías. Bueno, entonces la semana pasada fui a Barnes & Noble, empecé a recolectar libros al azar y a ponerlos sobre la mesa y encontré uno muy interesante, que terminé comprando: "Reinventing American Health Care". El libro literalmente me atrapó ya que explica por qué el servicio de salud de Obama es tan innecesario en EE.UU., utilizando información gráfica y muchos ejemplos a través de la visualización.

Todos estos gráficos y visualizaciones dentro de este libro está rodeados de textos, contextos, y esa es la clave para hacer visualizaciones o infográficos, no olvidarnos del contexto que queremos presentar y eso es lo que les deseo advertir, porque ustedes trabajan con visualizaciones y les quiero dar algunos consejos.

¿Cómo pueden mejorar su trabajo? Les daré tres recomendaciones:

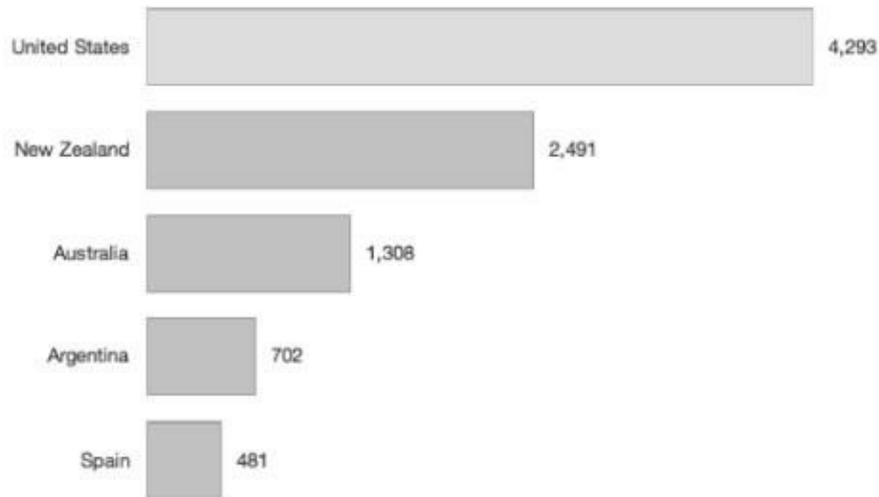
Primera recomendación. Las historias son siempre más complicadas de lo que parecen al principio. No traten de simplificar datos en exceso. Les daré algunos ejemplos de lo que digo:

Existe un sitio que se llama Vox.com, que intenta explicar temas complejos a lectores en general y es un excelente ejemplo de como a veces simplificamos en exceso. Hace unas semanas publicaron un montón de infográficos y un grupo de visualizaciones para explicar porqué el sistema de salud actual de los Estados Unidos es tan caro. Utilizaron diagramas, gráficos, cuadros, comparaciones, etc. Leí la publicación y me fui directamente a la fuente principal de todos estos datos: International Federation of Health Plans, y encontré el reporte en el cual Vox.com se basó para crear los infográficos. Me di cuenta que en el reporte no se mencionan muchas cosas interesantes sobre los datos, no se menciona que los precios para cada país se presentan por miembros participantes de International Federation of Health Plans, y la forma en que recopilaron los datos en Estados Unidos y en el resto de los países.

Para Estados Unidos, los precios fueron calculados sobre una base de datos de más de 100 millones de demandas que reflejan los precios negociados y pagados entre miles de proveedores y aproximadamente cien planes de salud comerciales. Cuando consultas los datos recolectados en otros países, los precios para Australia, Argentina y España son del sector privado, con datos proporcionados por un plan de salud privado en cada uno de esos países. No es un indicador. No puedes comparar manzanas con naranjas, y los periodistas que crearon esta historia no tomaron en cuenta estos hechos que no se pueden comparar, porque están comparando una gran base de datos con un sólo proveedor. Otra cosa a la que no pusieron atención fue que las circunstancias de cada uno de los países también afecta los precios, si se tiene un proveedor público de servicios de la salud eso baja los precios porque existe competencia, eso se tiene que tener en cuenta, de hecho la fuente original lo dice: la comparación entre los diferentes países es complicada debido a las diferencias en los sectores, tarifas y sistemas. ¿Por qué no se mostró esta leyenda en Vox.com? Estos datos son muy importantes para entender lo que la información significa, al no tomar en cuenta esto básicamente están creando una historia que no está completa.

A day in a hospital costs \$4,293 here – and \$702 in Argentina

Average cost of one day in a hospital (2013)



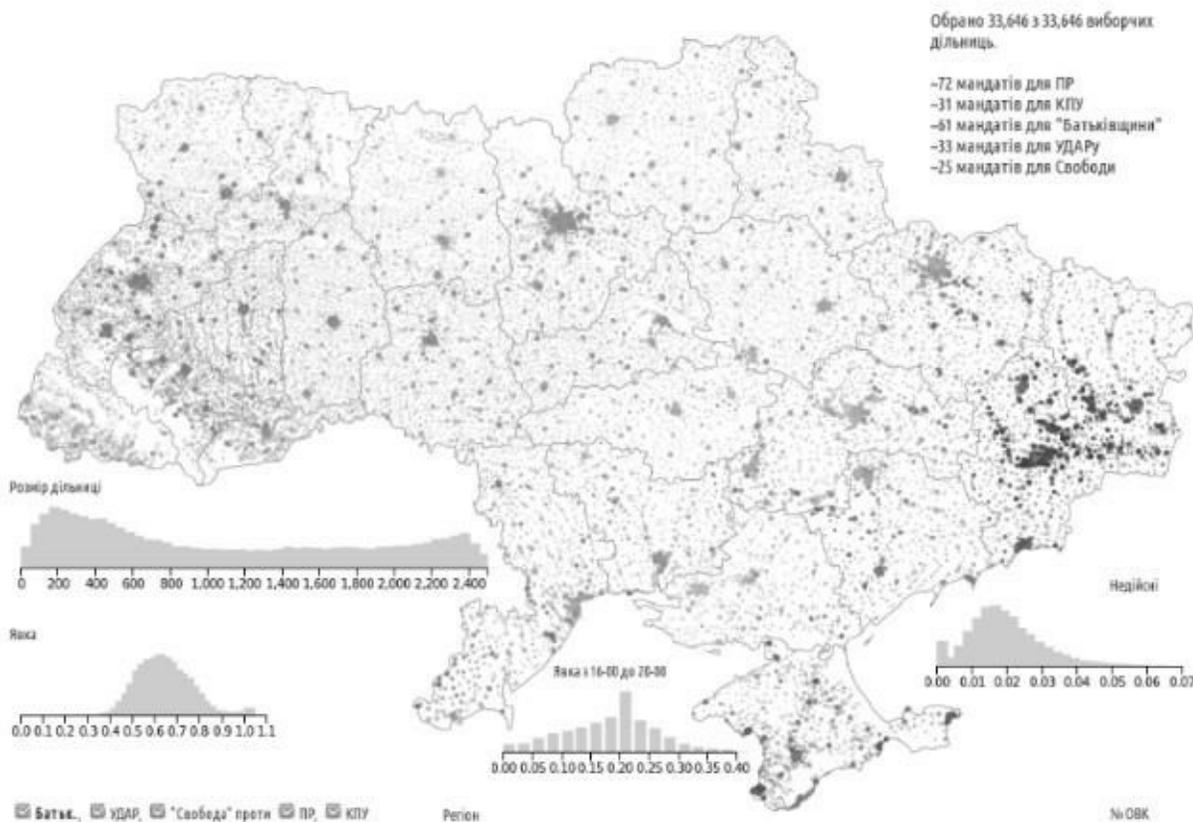
Created with [Datawrapper](#)

Source: International Federation of Health Plans

Además, cuando Vox.com creó los gráficos, cometió un grave error. En uno de los gráficos que hicieron, muestran cuánto cuesta pasar un día en el hospital en Estados Unidos, Nueva Zelanda, Australia, Argentina y España. En el encabezado del gráfico se lee: el costo de un día en el hospital en Estados Unidos US\$4.293 - En Argentina US\$702. ¡Por supuesto! Pero los salarios en Estados Unidos son más altos que en Argentina, así que esos datos no son comparables.

El primer principio de visualización, lo tomé de un científico llamado Richard Feynman y dice: "no debes engañarte a ti mismo y eres la persona más fácil de engañar". Debemos recordar esto, porque no podemos saltarnos a la conclusión, tienen que forzarse a sí mismos a navegar en los datos y en el contexto. Existe un término específico para cuando tratamos de saltarnos a la conclusión al ver datos, se llama patronicidad, es la tendencia de ver patrones significativos en datos sin sentido. Existen dos libros que hablan sobre la patronicidad "The Believing Brain" de Michael Shermer y "Thinking Fast and Slow" de Daniel Kahneman, ambos libros son muy importantes.

Resultados de las elecciones parlamentarias (partidos), 2012



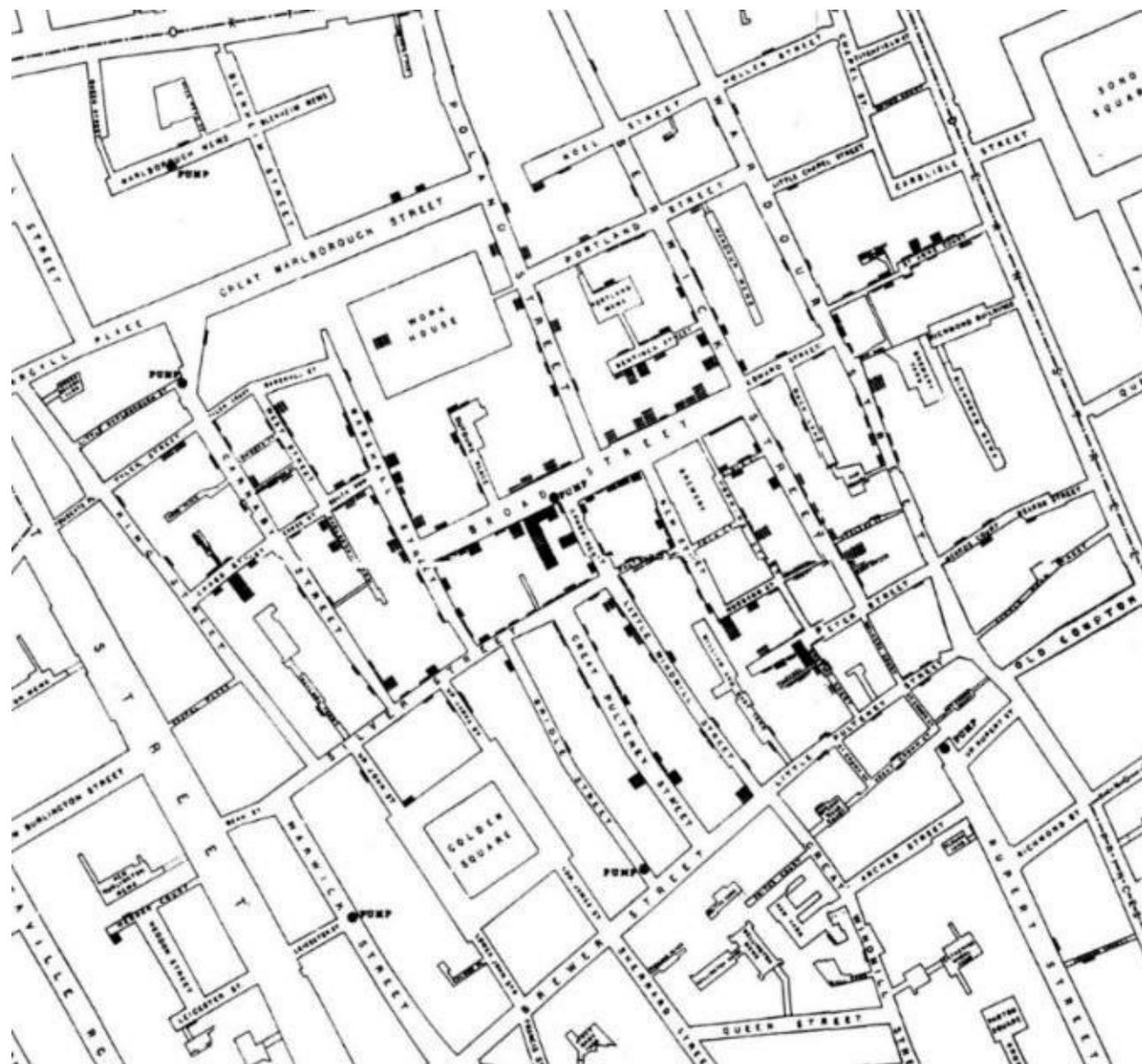
Ahora, este es un ejemplo de patronicidad que me ha pasado. Cada año, voy a Ucrania a impartir un taller sobre infografía y visualización, obviamente todos han leído sobre Ucrania y pues este año no fui por claras razones. La última vez que estuve ahí, esto fue antes que las protestas y la revolución comenzaran, conocí a personas que se dedican a hacer visualizaciones de datos periodísticos para una organización llamada Texty, una organización muy pequeña dedicada a la información, que sólo cuenta con un periodista y un diseñador, ambos muy talentosos. Ellos me mostraron las visualizaciones que habían creado e inmediatamente me di cuenta que eran muy buenas, una de ellas llamó mi atención más que las otras. Hay que recordar que me enseñaron las visualizaciones antes de que las protestas comenzaran. Eran infográficos que mostraban [resultados electorales](#), les mostraré un mapa con dos colores diferentes: anaranjado que representa el partido pro-occidental y azul que representa el partido pro-ruso. Cuando vi el mapa, aunque no hablo ni una palabra en ucraniano, me quedé sorprendido, porque muestra claramente lo que está pasando en Ucrania: la parte occidental del país votó en su mayoría por el partido pro-occidental, mientras que la parte oriental y sur del país votaron por el partido pro-ruso, el partido del ex presidente que había sido expulsado unos meses atrás. Fue una visualización sorprendente... y también un poco confusa, les mostraré porqué. Cuando las protestas comenzaron en Ucrania yo tenía este mapa en mi poder, antes que cualquier organización de noticias de Estados Unidos lo publicara, después empecé a ver los mapas en todos lados. La BBC, por ejemplo, publicó el mapa mostrando en donde empezaron las primeras protestas y si comparas los mapas, si los pones lado a lado hay un traslape, pues verán que las protestas comenzaron justo en la parte pro-occidental del país y cuando vi los dos mapas y los comparé, les envié un e-mail a los chicos de Texty diciéndoles que su mapa explicaba todo, que era muy claro, era demasiado claro... pero la escena política en Ucrania no era tan simple como la presentaba este mapa o como las organizaciones de noticias de los Estados Unidos la presentaba, es mucho más compleja de lo que parece. Los noticieros presentaban los mapas con una clara división en el país, lo que como encabezado es cierto, pero después del encabezado de la nota tienes que mostrar lo complejo.

Los chicos de Texty respondieron mi e-mail mandándome los resultados de muchos sondeos, en donde

se les preguntaba a diferentes personas del país si preferían ser parte de la Unión Europea o si preferían ser parte de la Unión Rusia, y como pueden ver la mayoría de la parte pro-occidental prefirió ser parte de la Unión Europea y si van a la parte oriente y sur en el área más pro-rusa, la imagen no es muy clara, pues sólo el 50% en la parte oriental estaba a favor de unirse a Rusia. Lo que trato de explicar es que en este mapa se está simplificando en exceso, sólo se muestra el encabezado de la nota, necesitamos mostrar una imagen más compleja de las variables y sus relaciones al lector.

Hay cosas que los periodistas hacemos cuando nos saltamos a las conclusiones. Nosotros los periodistas en general, la forma en que creamos historias es que primero tenemos una hipótesis y luego buscamos algo que la confirme, no ponemos atención en lo que potencialmente podríamos comprobar o refutar. Eso no es lo que los científicos hacen, lo que ellos hacen es tener una idea que confirme la hipótesis, que compruebe o refute lo que tienen en mente, así que si alguna vez han hecho esto, tener primero el encabezado de la nota y después la idea, dejen de hacerlo, así no debemos trabajar, olvídense del encabezado por un minuto y exploren primero los datos.

Existe un mapa que fue hecho en el siglo XIX por el médico John Snow, quien investigaba que era lo que causaba el cólera en Londres en ese siglo, ahora sabemos que se debe a beber agua contaminada por una bacteria. En el siglo XIX se pensaba que el cólera se transmitía al inhalar aire contaminado debido a que muchos doctores observaron que los lugares que apestaban era donde se presentaban más casos de cólera, incluso establecieron una relación entre las dos variantes "entre más apeste el lugar, existen más casos de cólera". Sin embargo, John Snow sabía que el cólera no podía ser causado por aire contaminado, él sabía que debía ser por algo más. Su hipótesis era que el cólera era causado por beber agua contaminada, pero ¿cómo comprobarlo en esa época sin los avances de hoy? Lo hizo mediante un mapa.

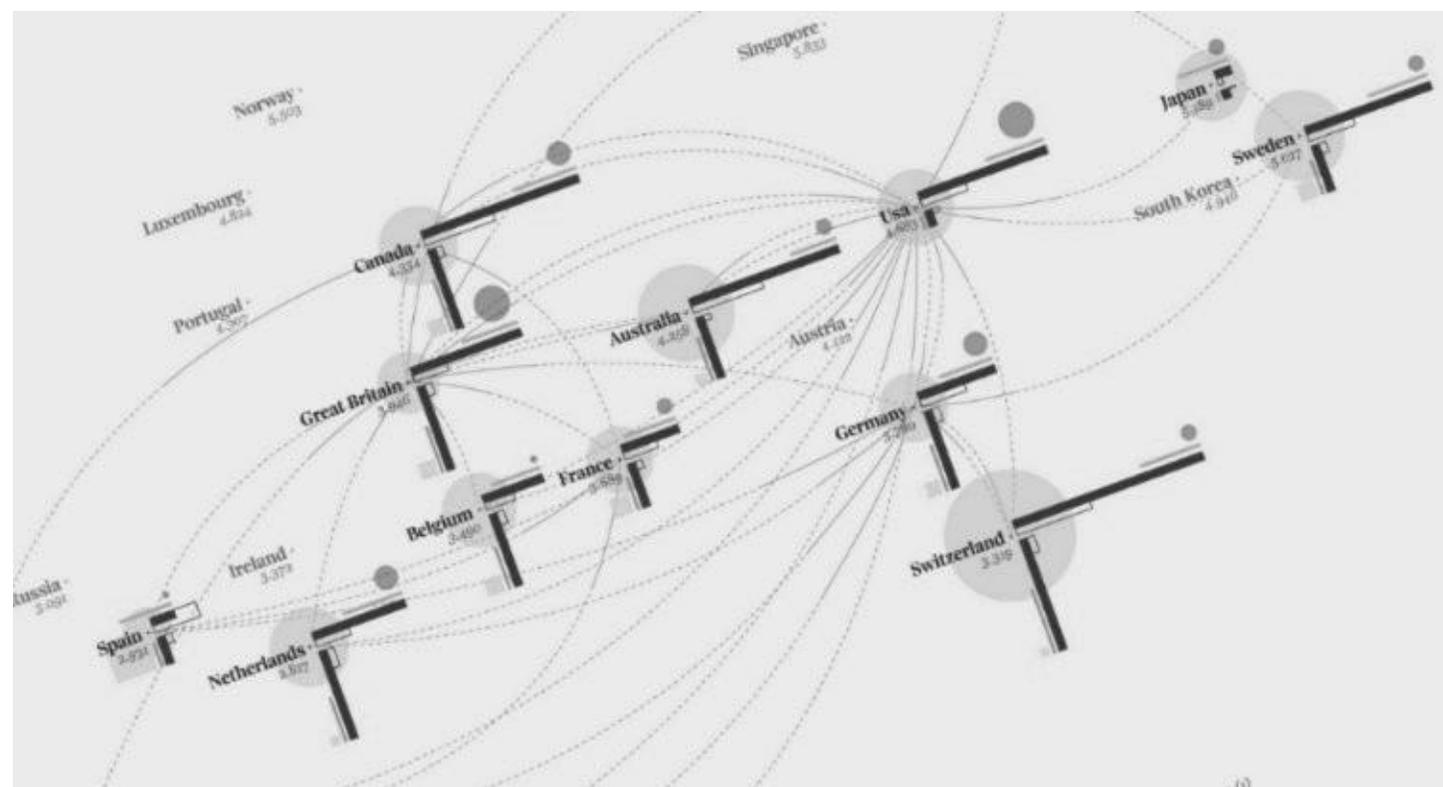


Las rayas más oscuras representan el número de muertes, entre más se aproximen al centro del mapa, más muertes registradas. Cerca del centro hay un pozo de agua y ahí hay más casos de cólera. Es muy claro. Estos patrones de concentración comprueban la hipótesis pero también tenemos aquí ciertas cosas que la podrían refutar, ¿cómo cuáles? También hay gente que murió de cólera, muy lejos del pozo de agua, ¿por qué murieron? Hay muchos edificios alrededor del pozo de agua en donde no se registraron muertes, eso podría refutar potencialmente la hipótesis.

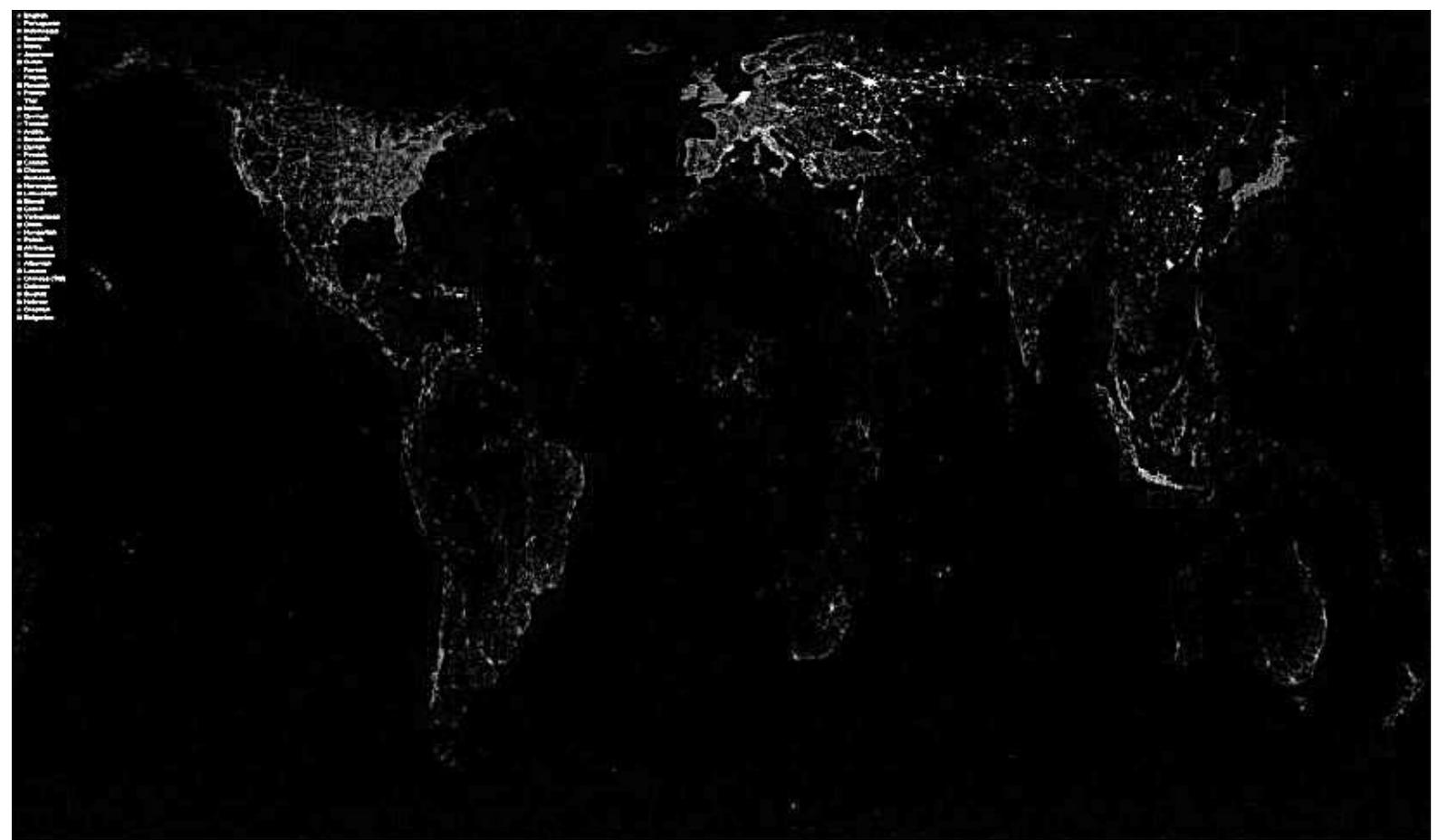
John Snow hizo algo que nosotros los periodistas deberíamos hacer más, puso atención a estas excepciones. Fue a esos lugares y descubrió que personas que murieron de cólera, que vivían lejos del pozo de agua, cada día camino a sus trabajos y a sus escuelas pasaban y bebían agua del pozo, por eso se contagiaron con la enfermedad. También puso atención a los edificios donde no se registraron muertes, fue a estos lugares y descubrió que estas personas tenían un pozo de agua privado dentro de los edificios y era de donde bebían agua, por eso no contrajeron cólera.

Es necesario poner especial atención a las excepciones, al contexto de los datos y a las variables que puedan afectar la historia.

Segunda recomendación. No se confundan. Sean creativos pero también escépticos de la creatividad. No simplifiquen las cosas en exceso pero tampoco las compliquen. Ahora les voy a hablar de las personas o agencias que admiro y les daré algunos ejemplos de esto.



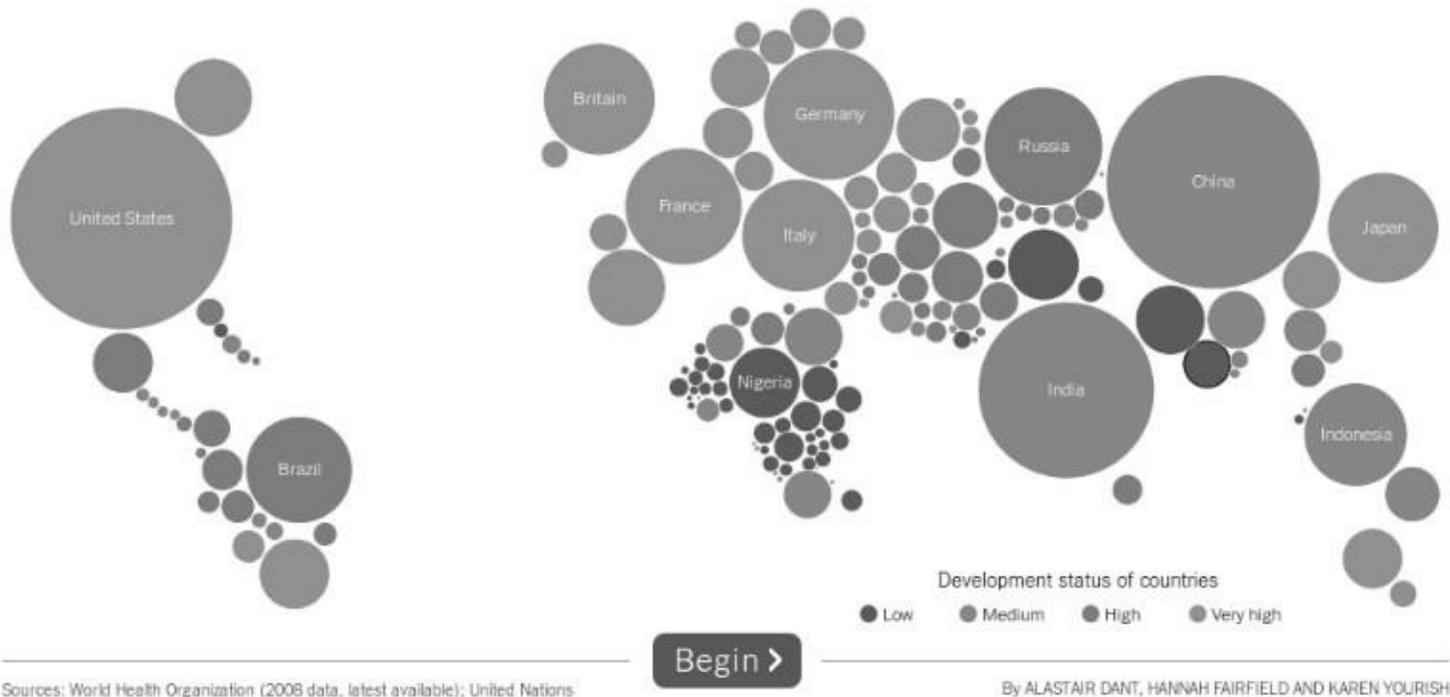
La primera de ellas es [Accurat](#), que producen muchísimos datos con visualizaciones pero algunos de ellos son muy complicados y difíciles de interpretar, algunos de ellos, no todos. Tenemos uno llamado "Brain drain" (Fuga de cerebros) que intenta presentar el porcentaje de investigadores de cada uno de estos países que son extranjeros, comparado con el número de personas que hacen otra cosa, comparado con esto, comparado con aquello, etc. Me perdí. No entendí nada. Si revisan el número de variables, son muchas. Es muy cansador para mí como lector aunque esté acostumbrado a leer visualizaciones. Esto es lo que pasa cuando se quiere ser sumamente creativo, los datos y variables deben de ser un poco más sencillas para un mejor entendimiento. Esto no significa que no debemos crear visualizaciones hermosas, significa que primero deben ser funcionales, debemos pensar en cómo nuestros lectores las van a interpretar.



Una persona que también admiro es Eric Fischer, quien creó un hermoso mapa de idiomas usados en Twitter. Aunque es un mapa muy bonito, básicamente te muestra algo que ya sabemos: las personas en España mandan tweets en español, los franceses en francés, los ingleses en inglés, los alemanes en alemán, etc. Esto representa a la población en general, ¿en qué sentido es esto interesante? ¿cómo podríamos mejorar este gráfico? ¿qué le falta a este hermoso gráfico? Pues, anotaciones, resaltar las excepciones, los detalles, las historias interesantes en los datos, etc. Les voy a mostrar un ejemplo: yo me pregunto, ¿quiénes son esas personas? ¿cuántas personas en Francia no mandan mensajes en francés? En fin, a este gráfico le falta resaltar los datos interesantes y decir algo respecto a las personas, colocar un número y decir algo sobre estas personas, de esa forma se tendría información del contexto.

Where Is Breast Cancer?

Breast cancer is the world's most common cancer in women, with an estimated 1.4 million new cases diagnosed every year.



Sources: World Health Organization (2008 data, latest available); United Nations

By ALASTAIR DANT, HANNAH FAIRFIELD AND KAREN YOURISH

Otro gráfico que está muy bien hecho en mi opinión, es del [New York Times](#). Verán que es un gráfico en escala que muestra la relación que existe entre el número de casos de cáncer de seno detectados en estos países y la tasa de mortalidad detectada en cada uno de los mismos. Entre más casos detectados, menor es la cantidad de muertes y entre menos casos detectados, mayor es el número de muertes. Lo interesante del gráfico es cómo resaltan las historias interesantes de los datos, porque incluyen un botón que al darle clic va a desplegar parte de la historia que te dirá algo sobre estos países, existe una interacción entre las palabras y los visuales y eso es la clave, es narrativa, es contar una historia. Debemos tener presente estas técnicas, este tipo de dispositivos.

Por otro lado, debemos tener en cuenta que en algunas ocasiones es conveniente representar nuestros datos más de una vez y lo que veo que es muy frecuente entre diseñadores visuales es que ellos solamente los representan una vez y ya, miles de gráficos y entiéndalos como puedan. Lo que pueden hacer es presentar sus gráficos de diferentes formas: mapas, diagramas, cuadros, barras, etc. porque cada uno de estos gráficos va a llevar al lector a alguna parte.

Tercera Recomendación. Mi última recomendación es enfocarse en aquello que más importe, lo más relevante. Muchos diseñadores visuales se enfocan en gráficas triviales en lugar de enfocarse en cosas de más importancia, como inequidad, salud, acceso al agua potable, etc. Existen algunas visualizaciones que muestran gráficos muy bien hechos pero triviales, sin mucha importancia.

Por ejemplo, la hermosa visualización de Citi Bike de Nueva York, que de hecho es un video, básicamente muestra cuantos viajes en bicicleta se realizan en esa ciudad, a pesar de que visualmente es muy llamativa, no hay ninguna información, después pensé: ¿qué tan relevante es esta información? Tal vez si vives en Nueva York. Creo que tal vez otro tema sería más interesante.

See HealthCare.gov Rates by State

Enter your age, location and plan type to search for prices in the 36 states on the federally run insurance exchange.

AGE

STATE

PLAN TYPE

COUNTY

SEARCH FOR PLANS

Source: Centers for Medicare and Medicaid Services

Un ejemplo de una visualización simple y sorprendente hecha por el Wall Street Journal te permite explorar las opciones de servicios de salud que hay en Estados Unidos, en donde puedes navegar utilizando tu dirección o tu código postal, por ejemplo yo que vivo en Miami, escribo mi código postal y me mostrará las opciones que tengo de servicios y planes de salud, los costos, etc. Esto es muy útil, importante y fácil de usar.

Por último, recuerden que nuestra primera responsabilidad como periodistas, diseñadores, es con el planeta, la humanidad, los ciudadanos, y después de eso, después de crear herramientas, nuestra responsabilidad es con los clientes. Nuestro principal objetivo debe ser crear dispositivos que hagan del planeta un lugar mejor antes de considerarnos artistas, creo que muchos diseñadores se consideran a sí mismos como artistas antes de considerarse ingenieros de la información y creo que eso somos, somos ingenieros, creamos dispositivos de información de una forma significativa, relevante y útil.

Alberto Cairo (@albertocairo)

Profesor en la Escuela de Comunicaciones de la Universidad de Miami desde 2012, dirige también el Programa de Visualizaciones de la UM's Center for Computational Science. Autor de “Infografía 2.0: Visualización interactiva de información en prensa” y “The Functional Art: an Introduction to Information Graphics and Visualization”, dos de los libros más consultados y citados respecto al tema. A principios del 2000 lideró la creación del Departamento de Infografías Interactivas del diario español El Mundo, ganando diversos reconocimientos internacionales entre 2001 y 2005. Es invitado habitual para exponer en eventos de periodismo de datos y visualizaciones en todo el mundo.

Publica periódicamente en su blog [The Functional Art](#).

Geoperiodismo: relatos que dialogan con el territorio

Cómo los reporteros pueden participar en el proceso de mapeo contemporáneo

Gustavo Faleiros

La relación entre periodismo y ubicación a menudo define lo que es una noticia. Entre las preguntas que deben ser respondidas en una investigación (quién, cómo, dónde, cuándo y por qué), la ubicación está siempre presente. Lo extraordinario en el mundo del periodismo rara vez ocurre en un espacio abstracto: ocurre en la calle, en la ciudad, el campo o incluso fuera de nuestro planeta, tal como lo muestran los telescopios que fijan su mirada hacia las estrellas.

Aunque estos lentes súper poderosos nos permiten mirar más allá de nuestro jardín, pocos periódicos, revistas y canales de televisión tienen, de hecho, un alcance global. Los medios crean su comunidad, su reputación, a través de su presencia local.

Piense en los periódicos o revistas que publican su guía de entretenimiento con películas y nuevos restaurantes. O piense en los resultados del equipo de la ciudad o las secciones que sólo reportean sobre barrios o la falta de servicios públicos. Incluso cuando se trata de asuntos internacionales, un medio prefiere a menudo trabajar con "enviados especiales", que se referirán a los hechos a través del lente de la audiencia local. El reportero, en estos casos, se convierte casi en un diplomático, hablando en representación de los intereses de sus ciudadanos.

Paradójicamente, Internet, que permite una comunicación global, ha empujado inmediatamente las noticias a un *hipernicho* y una ubicación determinadas. Múltiples servicios permiten a los usuarios personalizar la entrega de información respecto a sus intereses específicos y (muy a menudo) relacionados a dónde viven. La verdad es que por largo tiempo las herramientas digitales han filtrado el contenido para ser vistas tomando en consideración el interés y la ubicación.

Como periodista especializado en medio ambiente, puede haber parecido trivial poner más atención a la pregunta acerca de dónde están ocurriendo los hechos. En una mirada más próxima al mundo natural, las noticias no ocurren en un barrio, ciudad o país: ocurren en el Amazonas, el Sahara o en el Ártico, por ejemplo. Pero este ángulo territorial no debería ser considerado distante o nuevo comparado con el trabajo hecho día a día en un noticiario local, siempre enfocado en los problemas de las ciudades.

De hecho, son prácticas extremadamente similares, lo que cambia es la escala. Para dirigirse a asuntos tan grandes como el cambio climático, la deforestación del Amazonas o la quema agrícola en tiempos de sequía, observar el territorio provee mucha información al reporte basado en una sola ubicación.

Uno podría decir que el periodismo y la geografía tienen una función en común: brindar un conocimiento suplementario a otras disciplinas, generar un contexto para entender los problemas sociales y económicos. Como si fueran una capa bajo la historia. La idea de que el periodismo puede trabajar como una capa contextual al mundo físico y socioeconómico ganó una nueva dimensión con la tecnología actual. Hoy, agregar noticias y clasificarlas en grandes bases de datos puede ser tan importante como las bibliotecas de los periódicos, tan consultadas todavía por los investigadores.

Nuestra práctica mediante el proyecto InfoAmazonia.org, que describiré a continuación con más detalle, demuestra que es posible organizar las noticias actuando como cartógrafos modernos.

Lo que hacemos es incorporarle a las noticias información sobre su ubicación, que puede ser algo simple: el nombre de un país, estado o provincia, o con mayor granularidad, como una dirección completa o sus coordenadas geográficas. Lo que creamos son capas de noticias que pueden dialogar con otros datos. A este intento de crear narrativas en diálogo con la geografía lo denominamos geoperiodismo.

Los primeros experimentos de geoperiodismo

En 2008, cuando me convertí en editor del sitio O Eco, una agencia de noticias medioambientales sin fines de lucro con sede en Río de Janeiro, uno de los temas que escogimos como más importantes de cubrir fue la incidencia de los incendios dentro de las áreas protegidas. La decisión se basó en casos desastrosos como Chapada Diamantina, donde enormes incendios alcanzaron y destruyeron gran parte de la vegetación nativa.

Nuestra iniciativa nos condujo a estudiar reportes diarios hechos por el Instituto Nacional del Medio Ambiente (IBAMA), los cuales estaban basados en información satelital. Los documentos nos mostraron cuáles fueron los parques, reservas y estaciones ecológicas que estaban sufriendo por las llamas. Estos documentos, sin embargo, no revelan cuáles fueron las causas y las historias detrás de los siniestros. ¿Fueron intencionales o naturales? ¿Cuáles fueron las mayores pérdidas en biodiversidad?

Para obtener este tipo de información, hemos creado un procedimiento en el cual todos los reporteros deberán hacer turnos en una ronda de llamados de los guardias de las áreas protegidas donde hubo alertas de incendio en los boletines del IBAMA. Creamos un blog donde publicamos los resultados, con fotografías. No lo sabíamos entonces, pero estábamos en medio de un proceso colaborativo y un mapeo casi en tiempo real de los eventos.

Cuando estas historias comenzaron a acumularse, nuestro instinto básico nos llamó a crear un mapa con noticias localizadas acerca de los incidentes. Sin experiencia alguna como cartógrafos, sólo usando Google Maps para crear una capa con los límites de las áreas protegidas y dentro de estas fronteras, incluimos contenido periodístico (fotos, texto, etcétera). Con el paso de los meses, ganamos experiencia y nos asociamos con empresas que combinan los sistemas de información geográfica (GIS) y referencias cruzadas de nuestra base de reportes e información satelital.

Esta primera experiencia nos llevó a pensar en el uso de mapas como una herramienta narrativa. Fuimos inspirados a buscar maneras de mejorar la disponibilidad de información en la web. En 2010, cuando nuevamente hubo una situación de un gran número de 'hotspots' (puntos calientes) en el país, repetimos el experimento recolectando los reportes de los incendios dentro de las áreas protegidas.

Sin embargo, notamos que pudimos expandir la escala de nuestra cobertura utilizando el mapa como un medio de monitoreo colectivo. Así que, paralelamente, con la coordinación de una red de reporteros que continuó investigando las causas y consecuencias mediante entrevistas con los directores de los parques nacionales, comenzamos a pedirle al público que nos envíe información, como fotografías y reportes de incendios en sus ciudades. La decisión no estuvo basada sólo en la creencia de que estaría disponible al público para ayudar. De hecho, la idea vino al momento en que las redes sociales hicieron un llamado público en contra de la situación incendiaria y calamitosa en el país. Usando el marcador (hashtag) #chegadequeimadas, fue posible encontrar una auténtica protesta virtual organizada vía Twitter.

Hemos creado un mapa colectivo, el cual no sólo está abierto para recibir reportes directos del público sino que también utilizamos la habilidad de agregación proveída por la herramienta [Ushahidi/Crowdmap](#) para rastrear la información de redes sociales. Tal como lo haremos con proyectos futuros, agregamos coordenadas geográficas para asegurarnos de que las cuentas funcionarían como una capa en diálogo con otra información dispuesta en el mapa.

Pensando en nuevas formas de distribución

Tal como fue mencionado anteriormente, durante el proceso de creación de proyectos de mapeo, dos factores fueron primordiales siempre. La habilidad de organizar información generando contextos para el periodismo de noticias y generar una sensación de comunidad en los usuarios identificados con

las narrativas mapeadas. En ambos casos, un factor común es la forma en la cual esta información es distribuida. Además de permitirnos tomar ventaja y alcanzar una escala, las herramientas de mapeo digital también nos permiten promover el interés a través de la belleza. Entonces, pensar en las maneras en que la información será distribuida siempre fue crucial en los proyectos que comenzamos. Desde el momento en que comenzamos a hacer las primeras revisiones de datos con información obtenida de imágenes satelitales o puntos que representan incendios en un mapa, nuestra voluntad fue la de crear mapas en el Amazonas. Lo que llamó nuestra atención es que aunque sea tan relegada cuando de políticas públicas se trata, el Amazonas es necesario para hacer investigaciones científicas, y mejor aún, los datos generados por ellos, están frecuentemente accesibles al público.

Así que cuando es contrastado con 20 años de información detallada sobre el proceso de deforestación en Brasil, comenzamos a pensar en cuál sería la mejor forma para representarlo a nuestra audiencia, para que esta se sienta más comprometida con el problema. Más que una representación gráfica, ¿cómo podríamos agregar el valor del periodismo y el análisis y contexto a la tendencia demostrada por los datos de la destrucción?

Con estos asuntos en mente diseñamos el proyecto InfoAmazonia.org, lanzado en junio de 2012 durante la Conferencia de las Naciones Unidas sobre Desarrollo Sustentable.

El periodismo con mapas ¿puede ser el futuro?

Eventos recientes como la Primavera Árabe, ya sea en Egipto o en Siria, demuestran en gran medida que los reportes hechos por ciudadanos pueden reemplazar la importancia del reporte factual. Uno puede discutir el problema de la parcialidad o el alineamiento político de esos reportes, pero el poder del público para discernir estas tendencias no debería ser subestimada. La *hiperlocalidad* de la web también está vinculada con el tema de la reputación, determinada por quién muestra las noticias. Es nuestra comunidad la que decide qué considera verdadero.

Gustavo Faleiros (@gufalei)

Periodista, coordinador de InfoAmazonia.org y administrador de la Red de Periodismo Terrestre (Earth Journalism Network), ex becado de periodismo de la Knight International en conjunto con la ICFJ (Centro Internacional de Periodistas).

Proyecto Barcelona Commercial Footprints

Xavi Giménez

El proyecto [Barcelona Commercial Footprints](#) se desarrolló como participación en el concurso Inova Big Data Challenge, organizado por la entidad bancaria [BBVA](#). Dicho concurso se enmarca dentro del ámbito de la innovación. Se ofrece a los desarrolladores una API mediante la cual pueden ser consultados los datos relativos a todas las transacciones hechas mediante tarjeta, en las ciudades de Barcelona y Madrid, durante un período de 6 meses, por los clientes de BBVA. Con esos datos disponibles, comienza el desarrollo de una nueva idea.

Propósito del proyecto

Los datos ofrecidos por BBVA tenían multitud de vertientes a explorar, ya que la información estaba desglosada por multitud de criterios (edad, género, recurrencia de las compras, importe de las compras, geolocalización de las operaciones, categorización de las mismas, etc.). Por otra parte, se tenía la posibilidad de mezclar dicha información con fuentes de datos externas, de cara a enriquecer posibles “*insights*” surgidos de la visualización.

En mi caso me interesó:

- Visualizar a nivel geográfico donde había más volumen de compras
- Cruzar dichos datos con alguna fuente de datos externa

Con tal propósito, en el proyecto se visualiza en tres dimensiones la ciudad de Barcelona, mostrando aquellas zonas que atraen a usuarios de más lejos y qué zonas atraen a compradores de más proximidad. Los compradores son clasificados en 3 modalidades: a) compradores residentes en Barcelona, b) compradores residentes fuera de la ciudad pero dentro de la provincia de Barcelona y c) compradores residentes fuera de la provincia de Barcelona.

De esta forma se puede ver qué puntos de la ciudad atraen a compradores de la misma ciudad o de fuera. Al mismo tiempo, haciendo clic en cada una de las zonas interactivas de la ciudad se puede ver la relación entre la distancia que recorren los usuarios hasta esa zona y el importe de las compras que realizan.

Tecnologías y recursos utilizados

Todas las tecnologías utilizadas en el proyecto son tecnologías Open Source. Los datos extraídos de fuentes externas también son considerados Open Data.

Python

Se ha utilizado el lenguaje de programación Python para realizar:

- Toda la gestión de petición de datos a la API de BBVA.
- Toda la transformación de datos necesaria para acabar teniendo ficheros de datos con un tamaño aceptable para ser consumidos vía web: hay que tener en cuenta que se realizó un proceso en Python que estuvo durante horas consultando datos de la API para acabar teniendo los datos durante todo el período disponible. Esto conllevó tener un tamaño de datos inmanejable para cualquier aplicación web. Se utilizó Python para generar ficheros JSON con los datos justos y necesarios para la visualización.
- El procesamiento de los ficheros que contenían la información geográfica de la ciudad.

Portal Open Data 'CartoBCN', del Ayuntamiento de Barcelona

Gracias a dicho portal se pudo obtener toda la información cartográfica de la ciudad de Barcelona, y poder así representar la ciudad a un nivel de detalle excelente (a nivel de manzanas de casas).

Software QGIS

Un sistema de información geográfica libre y de código abierto. Dicho software permitió manipular toda la información cartográfica obtenida del portal web CartoBCN y poder transformarlo a ficheros en formato GeoJSON, un standard también opensource.

D3.js

Popular librería JavaScript para manipular y visualizar documentos basados en datos. Permitted realizar todas las visualizaciones y gráficos en 2 dimensiones.

Three.js

Librería JavaScript que permite realizar modelados 3d en el navegador.

El proyecto Barcelona Commercial Footprints se encuentra en la actualidad expuesto dentro de la exposición [Big Bang Data](#), una exposición donde se adentra en el fenómeno de la explosión de datos en el que estamos inmersos.

Xavi Gimenez (@xavigimenez)

Freelance, Ingeniero Multimedia y especialista en Visualización de Datos. Con años de experiencia desarrollando todo tipo de proyectos interactivos. Creador de Barcelona Commercial Footprints.

Transparencia y acceso a la información en Iberoamérica

Falta de transparencia gubernamental en el Perú

Yoni Pacheco

En el año 2003, en el Perú se promulgó la ley N° 27806 - "Ley de Transparencia y Acceso a la Información Pública" (modificada por la Ley N° 27927), la cual dice lo siguiente: “toda la información que posea el Estado se presume pública, salvo las excepciones expresamente previstas por el Artículo 15”.

El artículo 15 dice que las excepciones son asuntos militares, de inteligencia, información correspondiente a la defensa nacional, entre otros. Es por esta razón que el Gobierno peruano desarrolló una plataforma web para que las instituciones estatales puedan compartir su información con fines de transparencia. Estas instituciones comenzaron a incluir un enlace a la plataforma donde los ciudadanos podían encontrar y descargar datos sobre contrataciones de personal, gasto de presupuesto, etc. Si bien esto comenzó con fuerza luego de promulgada la ley, luego de unos años se perdió el interés y ahora es común encontrar instituciones que han dejado de publicar información desde hace unos años.

Muchas instituciones peruanas llevan un control de las personas que visitan físicamente sus instalaciones. Algunos ministerios se dan el trabajo de digitalizar el registro y publicar la información usando un buscador web desarrollado por el Estado peruano. Pero el buscador no es muy amigable y sólo se puede ver la lista de personas que visitaron las instalaciones día por día. No es posible buscar personas usando nombres como palabras clave, o buscar por documentos de identidad.

Usando estos buscadores, algunos periodistas peruanos han encontrado que familiares de políticos ingresan a la institución de contrataciones estatales para hacer diligencias y caen en evidente conflicto de interés. Sin embargo, su labor periodística sería mucho más ágil si los buscadores fueran de mejor calidad.

Con el fin de ayudar al periodismo peruano, en el blog [“el Útero de Marita”](#) hemos desarrollado software que se encarga de descargar todos los días la información nueva que va apareciendo en algunas instituciones que tengan el registro de visitas en línea. Esta información es almacenada en una base de datos local que sirve como soporte para el motor de búsqueda que hemos desarrollado. Si bien nuestro buscador es simple, también es poderoso. Los periodistas pueden ingresar parte del nombre de una persona y automáticamente aparecerán todas las coincidencias almacenadas en la base de datos. Hemos bautizado a nuestro buscador con el nombre de ["Manolo"](#) y puede ser visitado en el sitio.

Si bien algunas instituciones del Gobierno publican su registro de visitas, tenemos al Congreso peruano que se niega a hacerlo. Periodistas y abogados han hecho pedidos oficiales al Congreso, amparándose en la Ley 27.806, para que libere su registro. El Congreso ha contestado múltiples veces que su registro es secreto y no lo va a publicar, lo cual está en evidente infracción de la legislación peruana.

Otro problema frecuente es que las instituciones estatales peruanas liberan su información únicamente mediante la construcción de un buscador web y no hacen posible la descarga masiva de sus datos. Sería ideal si la información fuera publicada de las dos maneras: un buscador para la mayoría de la población y un enlace para descarga masiva para los periodistas de datos. Ya que la información es pública y además está publicada en la internet no deberían poner trabas para que los periodistas de datos puedan hacer una copia local de la información con el fin de hacer un eficiente minado de datos.

En el Útero de Marita hemos desarrollado y liberado software open source para realizar *webscrapping* y descargar masivamente documentos de las bases de datos estatales. En esto nos ha sido muy útil la librería open source llamada [Scrapy](#).

Algunas instituciones como el Poder Judicial llegan al extremo de poner un *captcha* que debe ser ingresado para poder descargar cada documento sobre sentencias de la Corte Suprema. Esto hace que nuestro trabajo de scrapping sea un poco más dificultoso. Pero en el Útero de Marita hemos desarrollado varios scrappers y en junio del 2014 logramos liberar casi 60 mil archivos estatales en nuestro blog.

Una de las instituciones peruanas que posee más datos (y quizá más importantes) es el Instituto Nacional de Estadística e Informática (INEI). Esta institución constantemente recopila información en todo el Perú mediante encuestas y censos. Además genera reportes, boletines y publican los datos crudos en su página web. Sin embargo, estos datos consisten en archivos producidos con software propietario: archivos MS Excel, SPSS y STATA.

Esta situación tiene varios problemas. Primero, que para poder leer los archivos hay que comprar el software y se supone que los contribuyentes ya pagaron para que la información se recopile y no sería justo que tengan que pagar dinero adicional para leer los datos recopilados. El otro problema de esto consiste en que no hay garantía que futuras versiones del software propietario pueda ser capaz de leer estos archivos y ya no se pueda acceder a la información. Sería un gran paso si las instituciones usaran únicamente formatos abiertos al momento de diseminar su información.

Si bien las instituciones estatales peruanas han hecho un esfuerzo inicial para transparentar sus funciones y actividades, pueden hacerlo mejor. Parte de la problemática es que las plataformas y buscadores son difíciles de usar. Otro problema es que las autoridades políticas no siempre tienen un buen entendimiento de la tecnología y liberan datos públicos pero con trabas y restricciones. Ojalá que se mejore con el tiempo.

Yoni Pacheco (@AniversarioPeru)

Biólogo con un PhD en ecología. Como parte de sus hobbies es investigador en usos de las tecnologías de información, periodismo de datos, y colabora escribiendo para el blog "el Útero de Marita"

Acceso a datos públicos en Brasil

Marcelo Träsel

En mayo de 2012, fue sancionada por el Presidente de Brasil la Ley 12.527, o la Ley de Acceso a la Información.

Provee una política pasiva y activa en los niveles federal, estatal y municipal, y en las tres ramas (Legislativo, Judicial y Ejecutivo). Ciudades con menos de diez mil habitantes no tienen obligación para publicar datos sobre los presupuestos del Gobierno en Internet, pero deben responder a las solicitudes de acceso a la información de los ciudadanos.

En general, la ley funciona. El mayor problema es la calidad de los sitios web de transparencia en vivo, que ofrecen información, pero rara vez en formato de datos abierto. En general, la facilidad de uso es muy mala y en otros casos, hay poca información disponible. Pero el balance de estos dos años de la ley parece positivo.

En cuanto a la transparencia pasiva, el cumplimiento de la ley depende en gran medida de la esfera y el tipo de organismo público. El Gobierno federal responde rápidamente y por lo general cumple con las peticiones correctamente, ya que los gobiernos municipales a menudo ignoran las órdenes, responden en forma parcial, lenta o negativa.

Un problema es el Artículo 13 del Decreto Reglamentario de la Ley 12.527, que dispone que las solicitudes de información sólo tienen que ser satisfechas por el Gobierno si los datos ya están disponibles y no es necesario invertir recursos en su producción. Muchas agencias han utilizado este mecanismo para negarse a proporcionar información.

Brasil es signatario de la Iniciativa de Gobierno Abierto y mantiene una política nacional en materia de datos abiertos. Hay un gran incentivo para que los estados y municipios adopten normas internacionales de datos abiertos, sobre todo a través del desarrollo de software libre.

En Rio Grande do Sul, el Gobierno creó una oficina digital, que ha estado trabajando en la mejora de la transparencia pública y servicios digitales a los ciudadanos. La capital, Porto Alegre, ha creado recientemente el primer portal de datos abiertos municipal en Brasil, [DataPOA](#). Hay iniciativas similares se están implementando en todo el país, especialmente en las zonas de mayor desarrollo económico y social. El desafío consiste en llevar estas políticas de apertura a una mayor población.

Leyes de prensa, información y transparencia en Guatemala

Enrique Naveda

Aunque en Guatemala existe un marco normativo que, desde la Constitución hacia abajo garantiza el acceso, no fue hasta la promulgación de la Ley de Acceso a la Información Pública (LAIP) en 2008 que se contó con un instrumento legal que aclarara las dudas que las otras leyes más generales dejaban, y entregara un método para obtener la información. La LAIP, una norma bastante garantista, no es una herramienta para periodistas, sino para cualquier interesado, y aunque los periodistas emplean sus mecanismos con cierta frecuencia (menos de la esperada), el de los abogados es el gremio que más recurre a ellos. El garante del derecho es la Procuraduría de Derechos Humanos (PDH) y la normativa establece consecuencias civiles y penales para quienes la violen, además de plazos de entrega que se encuentran entre los más bajos en este tipo de leyes (diez días hábiles con una prórroga posible de otros diez).

En los seis años que la ley lleva vigente los resultados han sido mixtos, aunque con tendencia a la mejoría: por un lado, el periodista ya sabe, a grandes rasgos, de qué manera puede solicitar la información y qué criterios regulan la entrega, la prórroga, o la negación. Pero por otro, cada institución tiene reglamentos sobre cómo concretar la ley que no han sido homologados para todo el Estado. Esta falta de homologación, que afecta a todas las aristas de la ley (desde cómo se entiende la gratuidad hasta qué sistema de solicitud o qué formularios son válidos), conviven con violaciones flagrantes del texto debido a la creencia de muchos funcionarios de que la información es su patrimonio particular, a su voluntad de ocultar sus fechorías, a la simple ignorancia de los sujetos obligados, o a que, ante la desconfianza de no saber para qué quiere alguien tanta información, prefieren negarla. Todo esto es en buena medida resultado del legado autoritario del país.

Sin embargo, aunque hay manera de moverse en zonas grises y, sin negar de hecho la información, [negarla de facto](#), la ley ha generado formas administrativas y legales (por lo general eficaces aunque desesperantes y costosas) en que los interesados en la información puedan apelar una y otra vez y obligar, si están en lo correcto, a las instituciones a brindar la información, algo que antes era más complicado, incluso imposible.

Hasta ahora, que sepamos, Plaza Pública ha sido el único medio de comunicación y una de las pocas instituciones que ha llevado una violación del derecho de acceso ante la Corte de Constitucionalidad, que en más de un año no lo ha resuelto. Por otro lado, recientemente la PDH informó de los resultados para 2013 y de los [avances en materia de transparencia](#). Su reporte incluye todo tipo de solicitudes y de solicitantes. Aunque aún no podemos ofrecer datos cuantificados porque todavía estamos procesando la experiencia de nuestro proyecto “Clientelismo y Patrimonialismo en el uso de fondos públicos”, que exigía cantidades masivas de información pública, nuestra experiencia, como periodistas, es más agri dulce. Aparentemente, las instituciones son más recelosas de entregar información a periodistas que a otros ciudadanos. Esto puede deberse al hecho de que creen que un medio de comunicación puede dañar más gravemente su imagen o a que la información que pedimos les suele parecer más delicada para sus intereses.

Algunas instituciones están generando grandes cantidades de datos relevantes para comprender el funcionamiento del Estado. Por ejemplo, el Sistema de Contabilidad Integrada o Guatecompras. Son herramientas cuya interpretación no es sencilla. Aunque entenderlas equivale a entender los códigos y lenguajes de la burocracia (porque no están pensadas como herramientas de transparencia hacia el ciudadano, sino como sistemas de compras y contrataciones o de contabilidad) son instrumentos relevantes para la investigación periodística especializada. Aun así, algunos casos adolecen de grandes debilidades, como que la información es imprecisa, está duplicada (de manera que el monto de una

factura se suma dos veces y da apariencia de que algo costó el doble de lo que en realidad costó), o es directamente errónea.

Otros sistemas de publicación de datos abiertos que si han sido pensados con el objetivo de generar transparencia, como el tablero electrónico del Congreso de la Republica, que registra presencias, ausencias, votos, etc. (de ahí nuestra app Congreso en datos) son sin embargo constantemente saboteados para que no desempeñen su función. Por último, un número decreciente pero aún inaceptable de instituciones del Estado violan o tergiversan la norma que regula la información que es pública de oficio, y rara vez la entregan en formatos legibles para el procesamiento por computadora, aunque cuenten con ellos.

En otra línea, algunos de los registros más importantes (Registro de las Personas, Registro Mercantil, Registro de la Propiedad, Catastro, etc.) carecen de la opción de hacer consultas en línea, o están restringidas a la mayoría de la población; y la información en papel resulta cara y puede destruirse con facilidad. Ahora hay un esfuerzo de centralización de datos estadísticos en el INE, pero la información disponible en datos abiertos es muy poca, apenas de 2009 a la fecha. También existe buena cantidad de información macroeconómica de calidad en el portal del Banco de Guatemala, aunque en formatos muy ineficientes que obligan a copiarla y pegarla.

Enrique Naveda (@quiquenaveda)

Es el coordinador general de Plaza Pública, un diario digital guatemalteco de periodismo de profundidad fundado en 2011. Coautor del libro de política local *Bestiario del poder*, y autor de *El consigliere tiene un proyecto, un perfil* y ha editado también *El rector, el coronel y el último decano comunista*.

Acceso a la información pública en Uruguay

Maite Fernández

Desde el 17 de octubre de 2008 Uruguay posee una Ley de Acceso a la Información que regula el derecho de las personas a acceder a información en manos del Estado. Según la [Guía para periodistas](#) elaborada por el Centro de Archivos y Acceso a la Información Pública (CAInfo) y el Banco Mundial, esto ha sido fruto de un gran esfuerzo de la sociedad civil a través del Grupo de Acceso a la Información Pública, así como de parlamentarios uruguayos que vieron la necesidad de modificar las instituciones uruguayas en busca de mejores formas de gobierno.

Esta primera guía para el uso de la herramienta del Acceso a la Información Pública para la investigación periodística, pretende disipar todas las dudas que puedan surgir en el ejercicio del Derecho a la Información. Según el manual “la Ley uruguaya contiene elementos positivos tales como la gratuidad del acceso a la información. Asimismo, contempla el establecimiento de un organismo de implementación de la ley, la [Unidad de Acceso a la Información Pública](#), “la cual tiene, entre sus múltiples competencias, asegurarse el cumplimiento de la Ley de acceso a la información, en relación al deber del Estado de clasificar información, así como de publicar la información que la ley establece como obligatoria.”

La ley ha tenido luces y sombras desde su promulgación. Distintos estudios marcan que queda mucho camino para que sea realmente conocida por la población y ampliamente utilizada más allá de activistas y periodistas. Es muy desigual el nivel de cumplimiento y respuesta de los distintos organismos del Estado.

En 2013 se intentó introducir una reforma regresiva a través de la Ley de rendición de cuentas que desató la reacción de la sociedad civil organizada a través de [Declaración.uy](#), impulsada por [DATA](#), CAInfo e ICD en un principio, sumando diversos apoyos. Finalmente, se logró una nueva redacción con aspectos que mejoran y otros que empeoran la ley original, pero claramente mejor que la modificación propuesta inicialmente.

UAIP

Hay desde la aprobación de la ley críticas a la independencia del órgano de control que es la Unidad de Acceso a la Información Pública. La relación con la sociedad civil es buena y en diversas oportunidades ha fallado obligando a publicar información declarada como reservada.

¿Qué sabes?

Lanzado en 2012, es el primer y hasta ahora único [portal de acceso a la información pública en Uruguay](#). Es una implementación del Alaveteli, software libre de Open Knowledge. Ha dado muy buenos resultados a nivel de uso continuado desde su lanzamiento y a nivel de cifras comparativas con otros portales similares. Al momento de su lanzamiento no estaba claro si los pedidos de acceso por correo electrónico debían ser respondidos según la ley y el portal -con trabajo de DATA y CAInfo- logró una resolución que saldó positivamente el tema.

La ciudad de Montevideo fue pionera en América Latina, publicando [el primer portal de Datos Abiertos de la región en 2009](#). Desde entonces ha mantenido de forma sostenida el portal, aumentando los juegos de datos, la actualización de los mismos y trabajando con diversas organizaciones en la reutilización de los datos, que han resultado en múltiples aplicaciones y herramientas. Hoy preparan junto a DATA los primeros servicios abiertos, llevando más lejos esta propuesta.

A nivel regional [el primer portal de Latinoamérica fue el de Uruguay](#), lanzado en 2012 y basado en CKAN. Desde entonces ha mantenido un crecimiento sostenido de juegos de datos y organismos participantes, acompañado por iniciativas como concursos, Data Bootcamp y apoyo a iniciativas de la

sociedad civil. Mantiene además un catálogo actualizado de [aplicaciones que utilizan los juegos de datos](#).

Periodismo de datos en Ecuador: un camino por iniciar

Paúl Mena Erazo

El periodismo de datos presupone el manejo de grandes volúmenes de datos. Los periodistas aplican diversos procesos con el propósito de verificarlos, sintetizarlos, explicar qué significan, colocarlos en contexto, e indagar cuáles son las historias que se pueden desprender de ellos. Pero, ¿qué ocurre cuando las bases de datos accesibles para los ciudadanos no son muy abundantes o no son ampliamente conocidas por la ciudadanía en general? Y, junto con ello, ¿cómo trabajar en este campo cuando el periodismo se desenvuelve en un ambiente de una permanente pugna del poder político y medios de comunicación, así como de preocupaciones en torno a los niveles de libertad de expresión y acceso a la información pública? ¿Qué hacer cuando no muchos medios apuestan por un periodismo de profundidad y de investigación? ¿Cómo iniciarse en el periodismo de datos en este contexto?

Estas, entre otras, son preguntas que se hacen muchos periodistas en Ecuador cuando conocen del auge del periodismo de datos en diversas latitudes del mundo, y quieren adentrarse en este campo.

En materia de acceso a datos públicos, desde 2004 está vigente en Ecuador la Ley Orgánica de Transparencia y Acceso a la Información Pública (LOTAIP). Esta norma establece una serie de disposiciones para garantizar el derecho de acceso a información pública señalado en la Constitución ecuatoriana. Sin embargo, la Ley tiene aún limitaciones en su aplicación. A partir de la aprobación de la LOTAIP las instituciones públicas han colocado en sus sitios web una serie de datos relativos a su estructura, directorio, remuneraciones, datos de presupuesto, viáticos, entre otros aspectos. Pero más allá del nivel de cumplimiento en la publicación de datos en sitios web por parte de las instituciones públicas, organizaciones de la sociedad civil han efectuado durante los años de vigencia de la Ley varios estudios que concluyen que los pedidos de información pública presentados por ciudadanos, entre ellos periodistas, no siempre son contestados o atendidos de manera correcta.

No todas las entidades públicas de Ecuador entregan a la Defensoría del Pueblo su reporte anual de cumplimiento de la Ley, tal como lo determina dicho cuerpo legal. De otro lado, si bien hay varios periodistas de investigación que han hecho uso de la Ley para obtener información y datos públicos, aún se necesita que un mayor número de reporteros y ciudadanos en general apliquen la Ley con el propósito de impulsar su cumplimiento.

De igual forma, se requiere poner en conocimiento de los ciudadanos un mayor número de bases de datos de información pública. Y aún falta capacitación a grupos ciudadanos y periodistas sobre cómo hacer uso de varias bases de datos ya existentes, si bien entidades como el Instituto Nacional de Estadística y Censos han venido trabajando en presentar a la opinión pública datos sobre diversas temáticas importantes.

Pero al hablar de la labor de los periodistas en Ecuador no se puede dejar de mencionar el contexto de la pugna del Presidente de la República, Rafael Correa, con diversos medios locales y periodistas, la cual se ha desarrollado a la par de la creación de nuevas normas jurídicas alrededor de la comunicación, acciones legales contra determinados periodistas y medios de comunicación, y campañas mediáticas para refutar publicaciones periodísticas.

La polémica Ley de Comunicación, aprobada en junio de 2013, tiene varias implicaciones para el desarrollo de un periodismo investigativo y de datos en Ecuador. La norma legal fija una serie de regulaciones para los medios de comunicación y periodistas. La Ley crea dos nuevas entidades de control: el Consejo de Regulación y Desarrollo de la Información y la Comunicación, y la Superintendencia de la Información y la Comunicación. Este último es el ente encargado de aplicar sanciones previstas y tiene como máxima autoridad a un funcionario electo de entre una terna de

candidatos presentada por el Presidente de la República.

Entre las disposiciones de la Ley de Comunicación que pueden tener incidencia en el periodismo de investigación y de datos está la creación del llamado “linchamiento mediático”, figura legal que no constaba en la normativa ecuatoriana.

Artículo 26.(Linchamiento mediático) Queda prohibida la difusión de información que, de manera directa o a través de terceros, sea producida de forma concertada y publicada reiterativamente a través de uno o más medios de comunicación con el propósito de desprestigiar a una persona natural o jurídica o reducir su credibilidad pública.

¿Quién determinará cuándo un trabajo periodístico cae en un “linchamiento mediático” y cuándo se trata del seguimiento propio del periodismo investigativo? Será la Superintendencia de Información la que aplique las sanciones administrativas cuando dictamine que una información fue lesiva al prestigio de una persona. La Relatoría Especial para la Libertad de Expresión de la Comisión Interamericana de Derechos Humanos ha señalado que con la figura del “linchamiento mediático”, “cualquier denuncia sostenida de corrupción, que pueda conducir a la reducción de la credibilidad pública del funcionario involucrado, podría ser calificada por el órgano administrativo competente como “linchamiento mediático” y ser objeto de las correspondientes sanciones”.

La Ley de Comunicación, que mantiene principios como la reserva de fuente y cláusula de conciencia, abarca además una serie de aspectos que van desde sanciones por incumplimiento de normas deontológicas o por la omisión reiterativa de “temas de interés público” hasta la inclusión en los medios de comunicación de defensores de la audiencia designados mediante concursos organizados por el Estado y la redistribución de frecuencias de radio y televisión.

El Gobierno ecuatoriano ha defendido la Ley de Comunicación colocándola como una herramienta contra “los claros excesos que tiene cierta prensa en Ecuador”.

En paralelo, no son muchos los medios de comunicación ecuatorianos que apuestan por un periodismo de investigación, y aún no se puede hablar de un desarrollo sostenido de trabajos de periodismo de datos con la participación coordinada de periodistas, programadores y diseñadores. Pocos son los medios que cuentan en sus redacciones con unidades de investigación establecidas. Pese a todo, varios periodistas sí han logrado revelar casos que han despertado la atención del país.

¿Cómo empezar?

Contextos similares al ecuatoriano se pueden encontrar, en uno u otro grado, en otros países de la región. En situaciones de carencia de una cultura generalizada de apertura de datos públicos, muchos periodistas se preguntan ¿por dónde empezar en el desarrollo del periodismo de datos? Frente a ello puede ser útil establecer una guía de primeros pasos que tome en cuenta recomendaciones de expertos en periodismo de datos y experiencias en diversos países.

En este punto reproduzco aspectos incluidos en el paper académico: “Periodismo de datos en un ambiente de pugna Gobierno-medios: ¿Por dónde empezar? Caso ecuatoriano”, que presenté en la Global Investigative Journalism Conference, realizada en octubre pasado en Brasil, y que ha sido la base para la realización del presente texto.

Capacitación en periodismo de datos

Los proyectos de periodismo de datos muchas veces inician con una pregunta periodística. Luego viene la búsqueda y manejo de datos para responder a dicha pregunta. Con el fin de introducirse en el manejo de amplios volúmenes de datos y su análisis, los periodistas deben adentrarse en áreas nuevas, en especial en materia de tecnologías, herramientas digitales, normas legales, y aspectos relacionados con matemáticas y estadísticas. De allí que el periodismo de datos requiere de una capacitación

sostenida. De otro lado, es importante que los periodistas se capaciten en el desarrollo de un manejo crítico de los datos, colocándolos en contexto y aplicando procesos de curaduría y verificación. No en pocas ocasiones se emiten en la web datos falsos, manipulados con diversos intereses, incompletos, o presentados de formas muy complejas y confusas, así como hay ocasiones cuando aparecen datos de una entidad oficial que parecen contradecirse con los datos provenientes de otro ente público.

Aplicación de leyes que posibilitan acceder a información pública

Para trabajar en periodismo de datos es importante conocer y aplicar las leyes vigentes que posibiliten acceder a información pública. El periodista que haga investigación y busque obtener datos públicos debe saber cómo poner en ejecución leyes de acceso a la información pública u otras normas que dispongan a entes públicos la divulgación de datos. En muchas ocasiones es conveniente mencionar en el pedido de información pública que se entrega a un ente estatal la ley o leyes que respaldan dicho acceso, pues ello mostrará que el periodista, o ciudadano en general, que hace la solicitud conoce acerca de sus derechos en materia de acceso a la información pública.

Uso de bases de datos públicas y privadas existentes

El acceso y uso de bases de datos es condición básica para el desarrollo del periodismo de datos. Como uno de los pasos iniciales para que periodistas y medios de comunicación de un país emprendan en el periodismo de datos es importante que ellos conozcan qué bases de datos públicos ya están disponibles para su acceso, sean aquellas manejadas por entes estatales, instituciones internacionales, organizaciones de la sociedad civil, universidades, o entidades privadas. En varios casos conjuntos de datos están diseminados de una u otra forma en la web, pero los periodistas no los conocen o no saben cómo adentrarse en documentos llenos de datos. Y también se podría considerar el crowdsourcing, es decir la recopilación de datos a través de la participación de ciudadanos vía aplicaciones móviles, mensajes de texto o formularios web, con la consecuente aplicación de acciones de verificación y curaduría.

Establecimiento de redes de periodismo de investigación y de periodismo de datos

En el periodismo de datos es importante que periodistas, programadores y diseñadores puedan aprender unos de otros, intercambiar experiencias y compartir herramientas técnicas para la obtención, manejo y visualización de datos. Con el incesante avance tecnológico, cada vez aparecen nuevas herramientas que pueden ser de mucha utilidad a la hora de enfrentarse a volúmenes de datos. La conformación de redes de periodismo de investigación y de periodismo de datos puede contribuir a crear espacios físicos y en línea para dicho intercambio, así como para promover mayor capacitación y resaltar en medios de comunicación la importancia de incursionar en este campo.

Impulso al periodismo de datos por parte de organizaciones periodísticas

Las organizaciones periodísticas pueden jugar un papel importante en el impulso al periodismo de datos en un país. Muchos de los eventos de periodismo de datos en diversos países han sido promovidos por entes de periodistas que han convocado a diversos medios de comunicación, así como a expertos en el manejo de datos. Las organizaciones de periodistas además pueden contribuir con la realización de eventos de capacitación sobre periodismo de datos, así como con el desarrollo de iniciativas para dar a conocer a los periodistas y ciudadanos en general acerca de las bases de datos públicas ya existentes en un país.

Adaptación del periodismo a las nuevas tecnologías

La práctica de periodismo de datos requiere de la aplicación al periodismo de herramientas tecnológicas, aplicaciones web y móviles, y software en general que faciliten la obtención, gestión y análisis de conjuntos de datos de interés público. Desde el manejo de Excel hasta de la gestión de datos a través de Google Drive, Google Fusion Tables, la realización de mapas por medio de Google

Maps y Google Earth, conversión de hojas de cálculo en tablas HTML, manejo de PDF, diseño de visualizaciones e infografías, herramientas de periodismo móvil y crowdsourcing, así como de geolocalización y redes sociales, todo ello puede tener un rol que jugar en medio del ejercicio de labores de periodismo de datos. No se trata de que los periodistas en general se conviertan en técnicos programadores especializados, sino justamente en que conozcan las herramientas para poder efectuar una gama de actividades de manejo de datos y para entenderse con programadores y diseñadores que se unan al equipo de periodismo de datos.

Vinculación entre periodistas e informáticos, diseñadores y expertos en otras áreas

El desarrollo del periodismo de datos demanda del manejo de extensos volúmenes de datos, los cuales, para su presentación periodística, requieren de visualizaciones interactivas. En dicha labor es necesaria la vinculación entre periodistas, desarrolladores de software y diseñadores.

En este marco en diversos países se han organizado eventos denominados *Hackatón*, *Datafest*, *Hacks/hackers*, en los cuales equipos integrados por periodistas, diseñadores y programadores se juntan para desarrollar proyectos de trabajos de periodismo de datos, que además de visualización, pueden incluir aspectos de *crowdsourcing*, datos desde teléfonos móviles, geolocalización, redes sociales, validación de información y sobretodo creación de historias periodísticas.

Estos siete pasos descansan sobre el hecho que el trabajo de periodismo de datos demanda sin duda de la aplicación de las bases éticas de todo periodismo. La contextualización, verificación, curaduría, precisión, independencia, rigor periodístico siempre serán fundamentales. Así, una efectiva combinación de los aspectos aquí anotados puede ser de gran utilidad para periodistas y medios de comunicación que decidan ingresar en el campo del periodismo de datos en medio de sociedades donde las bases de datos accesibles al público no son muy abundantes, y en donde se evidencia una constante pugna del poder político con la prensa. Todo con el fin de que los ciudadanos conozcan en qué maneras conjuntos de datos pueden incluir aspectos que tienen un efecto en sus vidas.

Paúl Mena Erazo (@PaulMena)

Corresponsal de prensa internacional desde Ecuador y profesor de Periodismo Multimedia de la Universidad San Francisco de Quito (USFQ), con 18 años de experiencia profesional. Desde 2009 colabora regularmente con BBC Mundo. Ha escrito para diario El País de España, así como para medios de comunicación de Ecuador como diario Hoy, diario Expreso y revista Vistazo. Posee una maestría en periodismo obtenida en Estados Unidos en calidad de becario Fulbright.

Ha sido profesor de periodismo en universidades ecuatorianas como UDLA y PUCE. En los últimos años ha participado como instructor en talleres y conferencias para periodistas, estudiantes de periodismo y diversos medios de comunicación de Ecuador en áreas como periodismo digital, herramientas web para periodismo de investigación, introducción al periodismo de datos, SEO para periodistas, periodismo móvil, reestructuración y rediseño de sitios web de medios, redes sociales, periodismo electoral, y leyes y ética de la comunicación.

Gobierno Abierto en Chile

Juan José Soto

La iniciativa de Gobierno Abierto de Chile nace el año 2011, cuando desde el Ministerio Secretaría General de la Presidencia se inicia el proceso de ingreso a la [Alianza de Gobierno Abierto \(AGA\)](#). En septiembre del año 2011 Chile suscribe a dicha Alianza, y presenta su Primer Plan de Acción para la Alianza de Gobierno Abierto. Los Planes de Acción de AGA implican el tomar compromisos concretos a cumplir en un período determinado, y para Chile implicó diversas iniciativas, como la creación de un Portal de Datos Abiertos.

Hay que entender el Gobierno Abierto como una oportunidad para cambiar las formas de relación entre Gobierno y ciudadanía, implica la creación de procesos participativos en la elaboración de políticas públicas y la co-creación de soluciones a desafíos concretos de las administraciones. La colaboración estado-sociedad civil debe ser fortalecida y con esto se abren las puertas a nuevas formas de trabajo para los funcionarios públicos, desde un enfoque en el cual los gobiernos buscan detectar los problemas, diseñar las soluciones y luego simplemente entregarla a los usuarios, hacia un enfoque donde los ciudadanos son activos partícipes en el diseño de los servicios y políticas públicas que se llevan a cabo, desde la detección de los problemas a la generación de las soluciones.

Lo anterior supone cambiar las maneras de trabajar de los gobiernos frente a una ciudadanía que se encuentra en medio de una reconfiguración de la ostentación del poder político, económico, social o comunicacional en el que gobiernos, empresas o medios de comunicación pierden parte de su poder y este es cedido a los ciudadanos. Es probablemente este cambio de paradigma en la manera en la cual se diseñan las políticas y servicios a los ciudadanos, uno de los mayores desafíos de modernización a los cuales se enfrentarán los gobiernos en los próximos años.

Uno de los proyectos en los cuales Chile está avanzando y se está poniendo especial énfasis, es el Portal de Datos Abiertos, pilar de la Estrategia de Gobierno Abierto. Se fortaleció públicamente a través de un instructivo presidencial emanado el año 2012, que convocaba a las instituciones públicas a publicar 5 datasets, intentando que la misma siguiese la línea de la información más solicitada por pedidos de transparencia. Habiendo sido puesto online sin mucha publicidad en el año 2011, en 2012 el Portal de Datos Abiertos del Gobierno de Chile (datos.gob.cl) contaba con 101 datasets y para el final de 2013 la cantidad había aumentado a más de 1000 publicaciones. Se logró aumentar de 14 instituciones públicas publicando en el año 2012 a 245 en ese mismo período, además de haber contabilizado más de 44.850 visitas al portal durante el 2013.

Si bien ha habido un avance continuo por parte de Chile en la publicación de datos, hoy se avanza en un aspecto esencial para que la misma tenga impacto real, y es la identificación de aquellos datasets de mayor demanda del ciudadano para su publicación.

Para incentivar la reutilización de los datos publicados en el portal, se organizó en noviembre del año 2012 el “Codeando por Chile”, el primer *Hackaton* organizado por el Gobierno. De igual manera se desarrollaron aplicaciones “*in house*” como el *iTransantiago* o el *iFarmacias*, cuyo código fue abierto en Github para que cualquiera pudiera reutilizar.

En febrero del año 2013 se hace pública la Norma Técnica para la Publicación de Datos versión 2.1 en Chile, un trabajo colaborativo y sometido a Consulta Pública antes de su redacción final. Dicha Norma Técnica explica “(...) un aspecto de gran relevancia al abordar la preparación y exposición de datos públicos en formato electrónico en Chile es normalizar dichos procesos, a fin de que cada institución utilice metodologías, procedimiento y prácticas comunes que garanticen la organización lógica, física y de formatos de los datos reutilizados”. Se establecen allí las características mínimas que deben

cumplir los conjuntos de datos para poder ser publicados en el Portal de Datos Abiertos. Al mismo tiempo se presenta la [Guía Rápida de Publicación de Datos](#), para facilitar el acercamiento a esta política de todas las instituciones².

Los desafíos no son muy distintos a los que todas las iniciativas de Gobierno Abierto de los países de la región: el cambio cultural que implica el repensar la relación entre el Estado y el ciudadano para llegar a trabajar de manera colaborativa es un proceso continuo.

En términos prácticos, el fortalecimiento de la Estrategia Nacional de Datos Abiertos girará en torno a mejorar la calidad de los datos, el nivel de conocimiento del portal, trabajando con instituciones clave que al día de hoy no publican datos (o sólo publicaron a pedido del Instructivo Presidencial del año 2012 y luego no continuaron). Por otro lado, se buscará fomentar el acercamiento de éstos datos a la ciudadanía con distintas iniciativas, desde el trabajo participativo para recolectar input de los ciudadanos sobre qué datos publicar, a maratones de visualizaciones para que las historias detrás de los datos sean cada vez más claras.

En este sentido, el trabajo colaborativo con la comunidad de periodismo de datos se torna crucial, dado que es la comunidad donde están quienes pueden encontrar y contar las historias que hay detrás de los datos.

Juan José Soto (@sotojuanjo)

Coordinador Ejecutivo de la Unidad de Modernización y Gobierno Digital, Ministerio de la Secretaría General de la Presidencia de Chile.

Acceso a la información pública en Venezuela

Emilia Díaz-Struck

A diferencia de otros países de la región para la obtención de datos, Venezuela no cuenta con una Ley de Acceso a la Información Pública. Sin embargo, [dentro de la Constitución](#) se garantiza el derecho de acceso a la información y algunos de sus artículos como el 28, 51, 57 y 58 pueden ser citados al momento de hacer una solicitud de acceso a la información pública.

El Artículo 28 hace referencia al derecho de las personas a acceder a datos sobre sí mismas, el 51, al derecho de las personas a dirigir peticiones ante cualquier autoridad sobre asuntos de su competencia y de obtener oportuna respuesta. Por su parte, el 57 garantiza la libertad de expresión y el 58, el derecho a la información oportuna, veraz e imparcial, sin censura.

[La Coalición Proacceso](#) en Venezuela, integrada por una serie de organizaciones que promueven este derecho, tiene en línea un [tutorial](#) sobre cómo hacer una solicitud de información pública en el país. La respuesta a estas solicitudes aún es limitada. Falta generar una cultura que promueva la transparencia y el uso de datos abiertos disponibles vía electrónica.

En diciembre de 2007, un trabajo publicado por el entonces vespertino El Mundo (en la actualidad El Mundo Economía y Negocios), daba cuenta de que casi dos meses le tomaba a un funcionario contestar una carta y que de las solicitudes entregadas, únicamente 37,5% recibieron respuesta (25% por escrito y 12,5% a través de entrevistas). Esto no significó que en todos los casos las respuestas facilitasen la información solicitada. El trabajo se hizo a partir de los registros de las 16 cartas enviadas por la unidad de investigación del diario a instituciones oficiales ese año. En la actualidad, la tendencia se mantiene.

Igual es posible acceder a algunos espacios oficiales en línea, que dan acceso a información de interés público como gacetas oficiales, sentencias judiciales y empresas contratistas del Estado venezolano.

Los formatos no siempre son los ideales, pero los datos que allí se encuentran sirven de punto de partida para el trabajo periodístico. En el caso de la información sobre empresas, se encuentran en línea aquellas que tienen algún contrato con el Estado. Para las demás, es necesario acudir a los Registros Mercantiles y consultar en los archivos la información que se encuentra en papel.

Las posibilidades de acceder a datos para historias vinculadas con Venezuela se multiplican cuando las historias se conectan con otros países y a través del uso de leyes de acceso a la información así como la consulta de espacios digitales con data abierta, se puede tener acceso a información relacionada con personajes, empresas y cifras ligadas a Venezuela.

En el proceso del trabajo con los datos, un elemento clave es la verificación de los resultados obtenidos, una vez superadas las etapas de limpieza y análisis de datos. Todos estos pasos estarán acompañados del reporte necesario para corroborar que la información que se presenta es confiable. En este punto es clave tener en cuenta: cuál es o cuáles son las fuentes de información y cómo se obtuvieron los datos con los que se está trabajando.

Aunque hay limitaciones evidentes en el acceso a datos abiertos y es necesaria una Ley de Acceso a la Información Pública, el abordaje que se ha dado en Venezuela a partir del contexto actual está vinculado con las posibilidades que existen al conectar historias con otros países así como al generar tablas de datos propias a partir de documentos.

Dónde buscar

Para conseguir datos relacionados con Venezuela, la lógica a emplear es la misma que en otros países. La información en línea se puede obtener a través de espacios como registros mercantiles, sentencias

judiciales, organizaciones internacionales, redes sociales, y registros de personas.

Para un taller sobre periodismo de datos, organizado por el [Instituto Prensa y Sociedad de Venezuela \(IPYS Venezuela\)](#), que contó con la participación de los canadienses Phillip Smith y Hugh Stimson como parte del grupo de instructores, distintos periodistas contribuyeron en la elaboración de una lista de recursos útiles en línea sobre [.dónde conseguir datos](#).

Entre los espacios de interés, que contienen datos asociados con Venezuela se encuentran:

- [El Tribunal Supremo de Justicia](#): tiene a disposición las sentencias judiciales emitidas por los tribunales venezolanos.
- [El Consejo Nacional Electoral](#): contiene datos sobre los resultados de los distintos procesos electorales y, en ocasiones, coloca a disposición el registro de electores, que puede ser descargado en formato CSV.
- [El Registro Nacional de Contratistas](#): ofrece información sobre las empresas que han sido contratistas del Estado venezolano. Incluye datos como accionistas y capital de la empresa. Para el resto de las compañías, es necesario ir a los Registros Mercantiles donde la información puede ser consultada en papel.
- [El Banco Central de Venezuela](#): tiene indicadores económicos sobre Venezuela. Están disponibles datos como tipos de cambio, tasas de interés, índice nacional de precios al consumidor, agregados macroeconómicos y balanza de pagos, entre otros.
- [El Instituto Nacional de Estadística](#): ofrece datos estadísticos sobre Venezuela que incluyen indicadores sobre población, regiones y económicos.

La lista se hace más extensa en función del tema a abordar y la lógica que se emplea para conseguir los datos, que abarca tanto páginas venezolanas como páginas en otros países. La clave es la creatividad y el desarrollo de búsquedas organizadas en las que se sepa con claridad qué se busca y cómo intentar encontrarlo. También intentar obtener la información en los formatos disponibles y no pensar que únicamente se encontrarán en los formatos clásicos de bases de datos. Muchas veces se encuentran en PDF y es necesario transformarlos para luego comenzar el proceso de limpieza y análisis.

Emilia Díaz-Struck

Periodista independiente. Publica historias en el espacio venezolano especializado en periodismo de investigación Armando.info, del cual es cofundadora. Colabora con el Washington Post, el Instituto Prensa y Sociedad de Venezuela (IPYS Venezuela) y el Consorcio Internacional de Periodistas de Investigación (ICIJ), organización de la que es miembro. Es profesora del departamento de periodismo de la Universidad Central de Venezuela.

Calidad de leyes de transparencia en México

Jesús Ibarra

La escasa penetración del periodismo de datos en México viene acompañada por la mala calidad de las leyes de transparencia estatales y trabas para obtener acceso a información pública. Según el Centro de Análisis e Investigación Fundar hay una enorme mediocridad en estos instrumentos a lo largo del país, como lo refleja el Índice del Derecho de Acceso a la Información en México (IDAIM Fundar, 2014).

El IDAIM mide la calidad de las leyes de transparencia en México con relación a las mejores prácticas nacionales e internacionales en la materia. Se compone de tres variables principales: diseño normativo, diseño institucional y procedimiento de acceso a la información pública y obligaciones de transparencia. Cada variable se compone de siete, cinco y nueve indicadores respectivamente, los cuales se alimentan de 196 criterios.

Dichos criterios representan los elementos que debe contener una ley de transparencia para que garantice y proteja el ejercicio del derecho de acceso a la información y favorezca la transparencia.

Fundar determinó con el IDAIM que las leyes de transparencia y el derecho al acceso a la información, entre las 32 entidades federativas generan una calificación promedio reprobatoria de 5.2 en la escala del 1 al 10. El IDAIM destaca el caso de la Ley del Distrito Federal, con una calificación superior al resto de las entidades, al obtener un puntaje mayor a 8. A la par, Fundar señala un punto de conflicto que impacta en la calidad de estas leyes con respecto al proceso de designación para renovar a los institutos de transparencia estatales, ya que de las 32 entidades federativas sólo cuatro incluyen la participación social y el resto tienen mecanismos que favorecen la elección de sus integrantes por medio de la injerencia de partidos políticos y el Gobierno.

Aunque las reformas en materia de transparencia fueron aprobadas por el Poder Legislativo, los congresos locales han aprobado reglamentaciones en materia de transparencia que se encuentran aún por debajo de los estándares nacionales e internacionales en cuanto al acceso a la información se refiere. No existen mecanismos suficientes para que los integrantes de los órganos garantes respondan a la autonomía con que cuentan los institutos de transparencia y no hay un esquema de rendición de cuentas. Es decir, a diferencia de la elección de los comisionados del IFAI, donde el Senado de México creó un mecanismo¹ para plantear de alguna manera que haya transparencia en el proceso de selección, en la mayoría de las entidades federativas no existen mecanismos claros para la elección de estos integrantes.

El criterio que menos leyes contemplan es el que considera al derecho de acceso a la información como un derecho humano. “Muchas leyes establecen que la ley garantiza el acceso a la información pública sin considerarlo parte de los derechos de las personas como se establece en el Artículo 19 de la Declaración Universal de los Derechos Humanos 11 y en el Artículo 13 de la Convención Americana de Derechos Humanos 12. Además, en el marco normativo nacional, su reconocimiento se establece en el Artículo Sexto Constitucional” (IDAIM, 2014).

De acuerdo con el análisis elaborado por Fundar, cuatro leyes incluyen actualmente a los sindicatos como sujetos obligados a entregar información (Distrito Federal, Durango, Michoacán y Oaxaca). Sin embargo, con las recientes reformas constitucionales en materia de transparencia, todos los estados deberán incluir no sólo a los sindicatos sino también a los partidos políticos como sujetos obligados. Por otra parte, la organización encontró que siete leyes no consideran aún como definitivas las resoluciones de sus órganos de transparencia: Campeche, Chiapas, Chihuahua, Guerrero, Hidalgo, Michoacán y Nuevo León.

En cuanto al mecanismo de elección de los integrantes de los órganos de transparencia, ninguna ley prevé la publicación del currículum vitae de los candidatos a consejero o comisionado. A pesar de que las leyes son heterogéneas, en 26 casos otorgan autonomía constitucional a los órganos de transparencia. De acuerdo con el diagnóstico de Fundar, los institutos no logran convertirse en un contrapeso ante los poderes que se niegan a entregar información a la persona que así lo solicita.

Leyes que favorecen (o no) el desarrollo del periodismo de datos en España

Jesús Flores Vivar

Ley de Transparencia, Acceso a la Información Pública y Buen Gobierno

Según resume la periodista Yolanda Marin, la Ley de Transparencia, Acceso a la información pública y Buen Gobierno (Ley 19/2013, de 9 de diciembre) establece una serie de principios éticos generales y también de obligaciones concretas para los miembros del Gobierno, los altos cargos de la Administración General del Estado y de las entidades del sector público estatal. Es decir esta norma obliga a los políticos a informar en qué gastan el dinero público y permite a los ciudadanos consultar a través de una web las subvenciones, los contratos o los sueldos de los cargos públicos, así como solicitar más información. Tras las aportaciones ciudadanas, se amplió la información económico-presupuestaria que debe hacerse pública, entre la que se incluyó lo relativo a los contratos públicos. Además, se estableció la obligatoriedad de hacer públicos los informes de auditoría y fiscalización, así como más datos sobre los contratos, entre los que deberán incluirse los contratos menores.

Sin embargo, desde antes de su promulgación, organizaciones como la Fundación CIVIO, en un post publicado en su sitio web en agosto de 2013, ya hacía referencia a que la norma es controvertida puesto no reconoce el derecho fundamental de acceso a la información, y además de no incluir informes, borradores y material de trabajo de las administraciones, lo más llamativo es que el presidente del Consejo por la transparencia será nombrado por el Gobierno. Es decir, un burócrata que actuaría bajo las directrices del partido político de turno en el Gobierno.

Leyes de prensa

En España, actualmente, rige la Ley Orgánica 2/1997, de 19 de junio, reguladora de la cláusula de conciencia de los profesionales de la información. Posterior a esta ley, se trabaja en el Proyecto de Ley del Estatuto del Periodista Profesional. Esto significa el desarrollo de un nuevo Estatuto del Periodista Profesional en España que defina la figura del/de la profesional del periodismo abandonando la mera regulación al acceso aprobada por el Decreto 744/1967 conforme lo previsto por la Ley de Prensa de 1966. Este proyecto de Ley comenzó en 1994 y aún continúa debatiéndose. Uno de los principales problemas consiste en que los propios profesionales no se ponen de acuerdo sobre qué o quién es periodista, así, la única ley aprobada al respecto durante la Democracia, la Ley de Cláusula de Conciencia, elude precisar a quién o quienes corresponde ese derecho de conciencia, cuando toda ley debe delimitar exactamente su ámbito de aplicación.

Ley Orgánica de Protección de Datos

La Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (LOPD) es una Ley Orgánica española que tiene por objeto garantizar y proteger, en lo que concierne al tratamiento de los datos personales, las libertades públicas y los derechos fundamentales de las personas físicas, y especialmente de su honor, intimidad y privacidad personal y familiar. Aprobada en 1999, esta ley se desarrolla fundamentándose en el Artículo 18 de la Constitución Española de 1978, sobre el derecho a la intimidad familiar y personal y el secreto de las comunicaciones. Su objetivo principal es regular el tratamiento de los datos y ficheros, de carácter personal, independientemente del soporte en el cual sean tratados, los derechos de los ciudadanos sobre ellos y las obligaciones de aquellos que los crean o tratan.

Práctica común en los problemas de acceso a la información

Obviamente, si por un lado se trata de salvaguardar los intereses en primer lugar de consumidores, por otro lado, algunas articulaciones de todas estas leyes acarrearán cierto peligro para el fácil desenvolvimiento del periodismo en general y del periodismo de datos en particular. Sólo la palabra

dato, guarda relación con la LOPD, que considera una serie de informaciones de carácter sensible y que no se puede utilizar para su publicación.

Respecto a la última ley publicada de acceso a la información pública, según Victoria Anderica, investigadora y coordinadora de Access Info Europe, "La ley no está a la altura. El Ejecutivo, desde sus inicios en la elaboración de la ley, se ha centrado en el buen gobierno y se ha olvidado de la transparencia. La ley debería recoger el derecho de toda persona a solicitar cualquier información a cualquier administración en cualquier formato y no lo hace."

Jesús Flores Vivar

Jesús Flores es Fundador y co-director del Observatorio de Periodismo en Internet y Fundador de la Asociación de Periodistas Bloggers. Como periodista ha trabajado en prensa, radio y prensa digital y como consultor tecnológico en PwC y Andersen Co. (UTE de Nakua Technologies). Ha sido Ponente en el International Journal of Arts and Sciences Conference en la Universidad de Harvard y es conferenciante invitado en congresos nacionales e internacionales. Actualmente, es profesor de Tecnología en la Universidad Complutense y profesor de Ciberperiodismo en el periódico ABC.

Es Doctor en Ciencias de la Información/Periodismo por la Universidad Complutense y Master en Información y Documentación por la U. Carlos III de Madrid.

Open Data y transparencia en Portugal

Dora Santos Silva

La libertad de expresión y de pensamiento por la prensa y por otros medios de comunicación social ha sido devuelta a los portugueses, después de la dictadura y de un período de censura que llevó casi 50 años, por el decreto Ley nº85-C/75 del 26 de febrero (1975), donde fueron registrados los principios básicos de la actual Ley de Imprenta. En ella está prevista la libertad de expresión, el derecho a la información, la libertad de publicación y difusión, la transparencia de propiedad periodística y el acceso a las fuentes de información por la administración pública, por las empresas públicas y por las empresas con participación mayoritaria de personas colectivas de derecho público.

Respecto a la política de datos abiertos, en el ámbito de la estrategia de transparencia promovida por la Unión Europea y de la iniciativa portuguesa “Governo Aberto”, se verifica la existencia de algunas iniciativas por parte del Gobierno de disponibilizar datos producidos por la administración pública, agregados en el portal dados.gov. Sin embargo, el movimiento open data se encuentra mucho más dinamizado por la sociedad civil y por algunas fundaciones en particular. Se destacan, por ejemplo, el proyecto [PORDATA](#), una base de datos sobre Portugal contemporánea, promovida por la Fundação Francisco Manuel dos Santos, y [demo.cratica](#), un repositorio de información sobre el parlamento, promovido por tres entidades independientes.

Acceso a la información en Colombia: entre la nueva ley y la cultura del secreto

Miriam Forero y Óscar Parra

Uno de los métodos más recurrentes entre los periodistas de investigación para obtener los datos que analizan, son las solicitudes formales a entidades públicas. Éstas han estado amparadas por la ley desde 1985, pero sólo desde marzo de 2014 el país cuenta con una [Ley de Transparencia y Acceso a la Información Pública](#), que juntó, organizó y actualizó la normatividad dispersa y confusa que existía.

Aunque es muy reciente para medir su efectividad, esta ley contiene varias novedades que le [servirán al periodismo](#), especialmente en el trabajo con bases de datos. Para empezar, incluye a las personas o instituciones particulares que cumplen funciones públicas, a los prestadores de servicios públicos y a los partidos políticos entre los obligados a entregar información.

Además, exige que las entidades del Gobierno publiquen en sus sitios web los directorios de sus trabajadores con las escalas salariales, su plan de compras, un inventario de todos los tipos de información que administran y datos abiertos en formatos reutilizables y estandarizados; da prioridad a las respuestas por vía electrónica y establece un procedimiento especial para solicitar información con identidad reservada, para los casos en los que esté en riesgo la seguridad del solicitante.

El impacto de la nueva ley se verá en la medida en que se vayan superando los dos principales obstáculos que tradicionalmente ha tenido el acceso a la información pública.

Por un lado, está el bajo uso que hacen los periodistas de este derecho; [una encuesta](#) del Proyecto Antonio Nariño (PAN) arrojó, en 2013, que para acceder a información pública, los periodistas recurren sólo en un 26% de los casos a solicitudes escritas (derecho de petición) y en un 24% al archivo público de entidades oficiales, en contraste con la frecuencia de uso de redes sociales para este propósito, que es del 61%. Sin embargo, se han dado experiencias exitosas, como el caso de Zoom Online que ha sido construida gracias a más de 90 solicitudes de acceso a la información.

Por otra parte, el Gobierno no siempre está dispuesto a liberar datos públicos como lo exige la ley, sobre todo en las regiones, donde es muy fuerte la llamada “cultura del secreto”, es decir, la tendencia de los funcionarios públicos a creer que son dueños de la información del Estado y que deben negarla cuando un ciudadano la requiere. De ahí que las principales dificultades de acceso que identificó el PAN son las trabas administrativas, la exigencia de revelar el propósito de la solicitud de información y la entrega de datos condicionada a no develar la fuente.

Para sortear estas dificultades, los periodistas en Colombia cuentan con herramientas legales que les permiten insistir en caso de respuestas negativas; pero además existe la ayuda institucional, encarnada en la [Fundación para la Libertad de Prensa](#), que ofrece múltiples posibilidades en cuanto al acceso a la información: un [manual actualizado](#), capacitaciones, asesoría y acompañamiento legal.

Apertura de datos

En las instituciones oficiales a nivel nacional se observa una mayor disponibilidad para liberar datos. Colombia se unió en septiembre de 2011 a la [Alianza de Gobierno Abierto](#) y por ella surgió el [Portal de Datos Abiertos del Estado](#), en el que se encuentran desde bases de datos útiles como la de número de víctimas por departamento, hasta datasets irrelevantes y con errores de descarga, como una lista de cuatro videos relacionados con el post consumo de llantas y baterías.

De manera similar, en los sitios web de diversos entes públicos se pueden hallar casos muy positivos de apertura de datos (consulta de antecedentes fiscales y penales, listados de personas sancionadas, reportes de financiación electoral, contratación pública, etc.), así como otros que, más que apertura, representan obstaculización, pues los datos están en formatos ilegibles y difícilmente reutilizables.

En Colombia se están dando los primeros pasos de la apertura de datos, de la mano con los primeros avances del periodismo de datos. Aún hay un largo camino por recorrer, y la principal necesidad es la conformación de equipos colaborativos que, con toda seguridad, ayudarán a darle un nivel más alto a esta práctica en el país y lograr, así, retos como trabajar con grandes datos, elaborar visualizaciones complejas y construir bases de datos robustas al interior de cada medio.

Acceso a la información pública en Costa Rica

Ronny Rojas

En Costa Rica existe suficiente normativa que garantiza el acceso a la información pública, la transparencia y la libertad de prensa y expresión. También hay jurisprudencia que garantiza a los ciudadanos el acceso a las bases de datos públicas, siempre que no tengan información de otros ciudadanos considerada confidencial: secreto bancario, expediente de salud, declaraciones de renta, números de teléfono, dirección de la casa de habitación, fotografías o lugar de trabajo, en el caso de que lo hagan en la empresa privada.

La Constitución Política de Costa Rica garantiza el derecho de petición en sus artículos 27 y 30:

Artículo 27.- Se garantiza la libertad de petición, en forma individual o colectiva, ante cualquier funcionario público o entidad oficial, y el derecho a obtener pronta resolución.

Artículo 30.- Se garantiza el libre acceso a los departamentos administrativos con propósitos de información sobre asuntos de interés público. Quedan a salvo los secretos de Estado.

La normativa establece un plazo legal de 10 día hábiles para que las instituciones públicas respondan a las peticiones de información de los ciudadanos.

El acceso a la información pública, especialmente a las bases de datos, tema que interesa, quedó claramente establecido por la Sala Constitucional de la Corte Suprema de Justicia en su resolución 2003-02120, del 14 de marzo del 2003. Dicha resolución la emitieron los magistrados en respuesta a un recurso de amparo interpuesto contra la Caja Costarricense de Seguro Social por una periodista del diario La Nación, a quién le había sido negada una base de datos con la información de todos los beneficiarios de subsidios del Estado por pensiones no contributivas.

En esa resolución, la Sala IV señaló:

"El contenido del derecho de acceso a la información administrativa es verdaderamente amplio y se compone de un haz de facultades en cabeza de la persona que lo ejerce tales como las siguientes:

- a) acceso a los departamentos, dependencias, oficinas y edificios públicos;
- b) acceso a los archivos, registros, expedientes y documentos físicos o automatizados -bases de datos, ficheros
- e) derecho de conocer el contenido de los documentos y expedientes físicos o virtuales

"consecuentemente, una hermenéutica finalista o axiológica de la norma constitucional, debe conducir a concluir que los administrados o las personas pueden acceder cualquier información en poder de los respectivos entes y órganos públicos, independientemente, de su soporte, sea documental -expedientes, registros, archivos, ficheros-, electrónico o informático -bases de datos, expedientes electrónicos, ficheros automatizados, disquetes, discos compactos-, audiovisual, magnetofónico, etc." (Jinesta Lobo, 2003).

Desde diciembre del 2011, Costa Rica forma parte de la [Alianza de de Gobierno Abierto](#), una coalición internacional de países y organizaciones de la sociedad civil comprometida a aumentar la apertura y rendición de cuentas del Gobierno. La Declaración de Gobierno Abierto emitida por la OGP en el 2011 y respaldada por el Gobierno costarricense desde entonces, señala:

Los gobiernos recogen y almacenan la información en nombre de las personas, y los ciudadanos tienen derecho a solicitar información sobre las actividades gubernamentales.

"Nos comprometemos a proporcionar activamente información de alto valor, incluidos los datos primarios, de manera oportuna, en formatos que el público pueda encontrar, comprender y utilizar fácilmente, y en formatos que faciliten su reutilización".

Reconocemos la importancia de los estándares abiertos para promover el acceso de la sociedad civil a los datos públicos, así como para facilitar la interoperabilidad de los sistemas de información del Gobierno.

Por otro lado, en el 2006, el Tribunal Supremo Electoral emitió una resolución que garantiza la naturaleza pública de toda la información de los partidos políticos:

“Toda aquella información relativa al patrimonio de los partidos políticos, independientemente de su origen privado o público, así como la cantidad de cuentas corrientes, sus movimientos y los balances que los partidos políticos que administran los Bancos Comerciales del Estado, bancos privados y cualquier entidad financiera son de interés público y, por consiguiente, pueden ser accedidos por cualquier persona. Dichas entidades no pueden negar o limitar el acceso a la información financiera de los partidos políticos, que se encuentre respaldada mediante soportes materiales, virtuales o digitales. El derecho al acceso de la información financiera y bancaria de los partidos políticos es una herramienta indispensable, como otras tantas, para la vigencia plena de los principios de transparencia y publicidad recogidos en el párrafo tercero del artículo 96 de la Constitución Política”.

Hay algunos casos, como el de las declaraciones de bienes de funcionarios públicos, en los que no se ha logrado conseguir acceso a la información. El Artículo 21 de la Ley contra el enriquecimiento ilícito obliga a dichos funcionarios a presentar estas declaraciones. Pero, esa misma ley, en su Artículo 24, establece la confidencialidad de los datos ahí contenidos y sólo le da acceso a las comisiones especiales de investigación de la Asamblea Legislativa, la Contraloría General de la República, el Ministerio Público o los tribunales de la República, para investigar y determinar la comisión de posibles infracciones y delitos.

En términos generales, las leyes de acceso a la información pública en Costa Rica funcionan bien y poco a poco han ido abriendo las puertas de las instituciones públicas. En mayor o menor medida, prácticamente todo el sistema institucional costarricense entrega información cuando se le solicita. Sin embargo, todavía es común encontrar cierta oposición para entregar determinada información, tal es el caso de la Contraloría General de la República, que el año pasado se opuso inicialmente a entregarle al diario La Nación la base de datos del Sistema Integrado de la Actividad Contractual (SIAC), que contiene la información de las compras del Estado. La Contraloría alegó que dicha base de datos podía contener errores que constantemente son corregidos por sus técnicos y argumentó que buscaba evitar que se informase de situaciones inexactas o irreales. Sin embargo, luego de que La Nación presentase algunos de los argumentos legales que anteriormente se detallan en este documento, el ente contralor accedió a entregar la base de datos completa.

Usualmente, cuando una institución se niega a entregar información pública, los ciudadanos o los medios de comunicación acuden a la Sala Constitucional, mediante un recurso de amparo. Los magistrados siempre han fallado a favor de los ciudadanos, cuando la información solicitada tiene carácter público legalmente.

Ronny Rojas (@ronnyrojas)

Periodista de la Unidad de Inteligencia de Datos del diario La Nación de Costa Rica. Especializado en periodismo de datos. Ha realizado investigaciones con Transparencia Internacional e IPYS (Buenos Aires, Argentina, 2010). Becado del Curso Avanzado de Periodismo de Investigación de IPYS (Rio de Janeiro, Brasil, 2013).

Exigiendo información al Estado

Verificando el discurso en un país sin ley de acceso a la información: el caso de Chequeado

Olivia Sohr y Laura Zommer

Dedicarse a verificar y contrastar con datos afirmaciones de líderes públicos en un país sin ley nacional de acceso a la información pública -ni política de datos abiertos- y en el cual las estadísticas oficiales están muy cuestionadas, es un gran desafío. Pero, contrariamente a lo que muchos escépticos pensarían, no es un reto imposible de superar.

Trabajar con paciencia y persistencia con la información existente -que no siempre significa disponible ni a la mano-, constituye un primer paso (la tecnología juega acá un gran papel). Acceder y utilizar los datos que muchos organismos públicos producen y, a veces, publican, es lo siguiente. Echar mano a las herramientas legales que tenemos disponibles para exigir información al Estado, cuando éste la niega u oculta sin razón válida, es otro paso necesario. Y, analizar luego los datos obtenidos con la colaboración de expertos de universidades, centros de estudios académicos y consultoras privadas, nos permite desde 2010 poner en circulación más datos e información verificada para contribuir a mejorar la calidad del debate público. En Chequeado hemos establecido un [método](#) para asegurarnos de la calidad de los datos que usamos.

Tener más información y datos y menos sesgos y prejuicios presentes en las discusiones y debates con el objetivo de que los ciudadanos tengan mayor información para participar en los asuntos públicos y reducir la impunidad intelectual de lo que se dice, se calla o se oculta es lo que guía nuestro trabajo.

En la Argentina no existe una ley nacional de acceso a la información que asegure el derecho de cualquier persona de exigir información y transparencia al Estado. Aunque la Constitución asegura el derecho de libertad de expresión y los tratados en ella incorporados en 1994 reconocen en forma expresa el derecho a la información -que incluye el acceso a la información-, no hay una ley nacional específica que lo reglamente.

Lo que sí hay es un decreto (el N° 1172/03), que garantiza que en el ámbito del Poder Ejecutivo Nacional se debe responder a los pedidos que se formulen. Esto incluye a todos los organismos, entidades, empresas prestatarias de servicios públicos que funcionen bajo esa órbita, así como también a las organizaciones que reciben subsidios del Poder Ejecutivo Nacional.

Es una herramienta limitada, no sólo porque deja fuera de alcance al Poder Judicial y al Congreso, sino porque al ser un decreto muchas veces es incumplido alegando la primacía de una ley sobre él (la de Protección de Datos Personales, o la de Procedimientos Administrativos, entre otras). Además, los numerosos rechazos del Ejecutivo a responder, combinados con largos tiempos de trámite en la Justicia para dirimir estos conflictos, implica que la información, en caso de que el Gobierno no la quiera brindar “por las buenas”, llegará muy lejos de los plazos periodísticos.

Un caso reciente y paradigmático es un pedido de acceso a la información formulado por el Centro de Implementación de Políticas Públicas para la Equidad y el Crecimiento (CIPPEC) para conocer la nómina de beneficiarios de planes sociales en 2008, que se demoró seis años en resolverse. La resolución vino de un fallo de la Corte Suprema contra el Ministerio de Desarrollo Social, a cargo de Alicia Kirchner, hermana del ex Presidente, que además exigió al Congreso a que [legisle al respecto](#).

La interpretación del decreto es muchas veces restrictiva. Chequeado, por ejemplo, solicitó en 2014 información sobre los términos bajo los cuales se firmó un acuerdo comercial entre YPF, empresa que desde abril de 2012 es mayoritariamente estatal, y Chevron, para la explotación de reservas de gas y petróleo no convencionales en Vaca Muerta, en el sur del país. La respuesta fue que YPF no es una empresa pública, por lo tanto, no está sujeta a los requerimientos del decreto. Otro intento fallido

tuvimos en Chequeado con Aerolíneas Argentinas, también manejada por el Estado, que respondió con un [argumento similar](#).

Esto ocurre a nivel nacional. Pero al tratarse de un país federal, hay mucha información que pasa directamente por las manos de los gobiernos provinciales. Y en este caso, la situación es desigual. Mientras algunas provincias legislaron sobre el tema y tienen mecanismo relativamente eficientes y aceitados para permitir el acceso a información pública, en otras no existe ningún tipo de regulación y dependemos por completo de la buena -o mala- voluntad de los funcionarios. Hoy 15 de las 24 provincias [tienen algún tipo de legislación](#).

El marco legal descrito sin duda nos limita. Pero no todo está perdido. Estamos obligados a utilizar mecanismos algo informales para acceder a la información y recurrir de manera mucho más frecuente a fuentes alternativas.

Cuando empezamos a trabajar en Chequeado temíamos que buena parte de nuestro sitio terminase destinado a la denuncia de la información faltante y a un reclamo indignado por mayor transparencia por parte de los poderes públicos. Y, hay que decirlo, muchas veces tenemos que recurrir a la denuncia.

Es el caso de los datos sobre delitos, por ejemplo. Cada vez que hablamos de homicidios (por tomar sólo alguno de los más graves) tenemos que adivinar tendencias y tasas en base a información parcial de algunas provincias y algunos registros judiciales, porque el Ministerio de Seguridad incumple no sólo el decreto 1172/3 sino también la ley que lo obliga a publicar [los datos criminales anualmente](#). Es también lo que nos pasa cuando queremos analizar la evolución de la pobreza: las estadísticas oficiales fueron manipuladas a partir de 2007 y, aunque intentamos suplirlo con mediciones alternativas, no se pueden reemplazar del todo los datos oficiales.

Quizás el peor caso que nos tocó enfrentar fue el cambio que hizo el Gobierno en la presentación de la información para dificultar su comprensión. Se trata de la información sobre la cantidad de viviendas construidas por algunos planes estatales. La primera vez que consultamos la información publicada online (para chequear si el actual Gobierno es el que más viviendas construyó, como aseguró la Presidenta, Cristina Fernández de Kirchner) aparecían en columnas diferentes la cantidad de viviendas construidas y las soluciones habitacionales (arreglos que mejoran una casa ya existente). Luego de la publicación de [nuestra nota](#), que desmintió la afirmación de la Presidenta, la presentación cambió: se combinaron las columnas y desde entonces es mucho más difícil saber exactamente cuántas casas se construyeron en la Argentina.

Pero, para entusiasmar a los optimistas, descubrimos que esos ejemplos no son la norma. Hay mucha más información disponible de la que esperábamos. El Ministerio de Trabajo de la Nación, por ejemplo, publica series de empleo de los últimos 30 años, el de Agricultura permite saber el uso de las tierras desde 1970 y el Ministerio de Educación nos da los números exactos de matrículas privadas y públicas por jurisdicción y tipo de educación. Parecen cosas básicas, pero son datos que solían no circular y que nos permiten entender mejor la realidad nacional, salir de los debates subjetivos y saber concretamente qué es lo que pasa en diferentes ámbitos.

Y nuestra arma de última instancia, el pedido formal de acceso a la información, funciona algunas veces. Cuando quisimos conocer el detalle de las comunicaciones entre el Gobierno Nacional y el Gobierno de la Provincia de Córdoba, durante la noche en la que la Policía cordobesa se declaró en huelga y hubo saqueos y disturbios en toda la provincia, se nos entregaron copias de los fax que se habían intercambiado esa noche, [a los pocos días de haberlos pedido](#).

El panorama del acceso a la información pública y la apertura de datos en la Argentina está muy lejos de ser ideal. Nos falta mucha información, pero en general logramos, por una vía o por otra, conseguir

Los datos que necesitamos para tener un debate más informado. Y cuando no los hay, consideramos que es nuestro trabajo denunciarlo, exponer la falta de información y transparencia para alertar a la ciudadanía y poner presión sobre las autoridades que tienen que proveerla.

Esperamos que así, en el futuro, tenga un mayor costo ocultar datos y, como consecuencia, nuestro trabajo y el de cualquier que quiera verificar el discurso se vuelva más simple.

Laura Zommer (@lauzommer)

Lic. en Ciencias de la Comunicación y Abogada, Universidad de Buenos Aires (UBA). Posgrado en Comunicación política, Universidad Austral. Colaboradora permanente del suplemento dominical Enfoques, del diario La Nación. Se desempeñó como directora de Comunicación de CIPPEC de 2004 a 2012 y fue jefa de Gabinete de la Secretaría de Seguridad Interior del Ministerio de Justicia y Derechos Humanos de la Nación en 2003 y 2004. Por su trabajo como periodista, fue becaria del diario El País de Madrid; recibió el Premio en la categoría Abogacía Argentina, otorgado por la Asociación de Entidades Periodísticas Argentinas (Adepa), 2005; el Premio Italia de Periodismo Joven, 2002; el Premio en la categoría Periodismo en Profundidad, otorgado por la Sociedad Interamericana de Prensa (SIP), Houston, 1999; y el Premio en la categoría de Bien Público, otorgado por Adepa, 1998.

Olivia Sohr (@olisohr)

Licenciada en Sociología, Universidad de Paris 8 Vincennes Saint-Denis. Maestría en Sociología, especializada en estudios de medios de comunicación. Ecole d'Hautes Etudes en Sciences Sociales (EHESS). París, Francia. Trabajó como pasante en la editorial de "Shanghai Business Review" en Shanghai, China (2009) y como investigadora en ARPP en París, Francia (2008). Fue editora del sitio web Scitizen.com, París, Francia (2008); pasante de comunicación en Jooce.com, París, Francia (2008); escritora freelance en Radio Universidad de Chile, Santiago, Chile (2007).

Abriendo el Gobierno de Puerto Rico: el proyecto Abre PR

Edgar Ríos

[ABRE Puerto Rico](#) es una plataforma dedicada a fomentar la apertura y transparencia gubernamental mediante la compilación y publicación de datos gubernamentales en un portal web (www.abrepr.org). Nuestra organización es sin fines de lucro y se estableció formalmente hace poco más de un año atrás pero lanzamos la plataforma públicamente el 5 de marzo de 2014.

El proyecto surge por la falta de acceso de datos gubernamentales en distintos renglones del país. Ante este escenario, un grupo de estudiantes graduados de política pública, preocupados por esta carencia, decidieron atender este problema. Auscultaron distintas maneras de atacar este problema hasta que coincidieron que hacía falta un esfuerzo local de datos abiertos que centralizara los datos gubernamentales sin impedir el acceso. Esto dio pie a una exploración de herramientas virtuales que serían de utilidad tanto para profesionales como a personas que sencillamente poseían curiosidad de conocer a su país más a fondo sin conocimiento en estadísticas.

Desarrollaron características que sirvieron de guías e incluían el acceso a archivos y bases de datos que se pudieran descargar, proveer funcionalidades básicas de organización de los datos, crear gráficas con los datos y ubicarlos en un mapa. Sin embargo, la característica más importante fue proveer acceso libre a esa información, sin registros ni cuotas. Luego de una búsqueda intensiva, identificaron una plataforma del Gobierno de Nueva York, la cual optaron por emular.

A modo de complementar los datos y facilitar su consumo, el equipo decidió crear visualizaciones de la data liberada. Dichas visualizaciones, o infografías, demuestran patrones destacables de los datos recopilados, ejemplifican cómo usar los datos y proveen un resumen visual de la información. Las visualizaciones han sido bien recibidas ya que han mostrado ser efectivas para comunicar datos complejos y patrones cuya relación no es aparente a audiencias que no estén familiarizadas con el manejo de datos estadísticos. Por tal motivo las infografías se han convertido en una parte esencial de la plataforma y se producen regularmente.

La programación de la plataforma ocurrió en paralelo a los primeros esfuerzos de solicitar y compilar un acervo de datos diversos. Sin embargo, como era de esperarse, dichos esfuerzos resultaron ser onerosos. Se confrontó resistencia de funcionarios, quienes desconocían el derecho del ciudadano a revisar cualquier información producida o albergada por dependencias gubernamentales. Existía aprensión por el uso que se le pudiera dar a la información suministrada, o sencillamente no existían procedimientos en las agencias para atender estas peticiones. Esto redundó en un recorrido laberíntico de oficinas dentro de estas entidades con el fin de obtener la información. Claro está, había un puñado de agencias más receptivas, pero definitivamente eran la minoría. Algunas de aquellas entidades receptivas se convirtieron en aliados que nos apoyaron con información y referidos. Por ejemplo, el Instituto de Estadísticas y la Oficina del Contralor de Puerto Rico se han convertido en los aliados más allegados en el Gobierno.

En el mes de diciembre se decidió lanzar la plataforma el 5 de marzo, a pesar de las limitaciones de datos y recursos, con el fin de agilizar el acceso a los datos ya recopilados. Se realizó un evento de lanzamiento ese día y desde entonces, hemos estado realizando esfuerzos para dar a conocer nuestra plataforma mientras continuamos ampliando y diversificando nuestro catálogo de datos.

A partir de nuestro lanzamiento hemos recibido muchas muestras de apoyo de individuos alrededor de la isla. Igualmente, han surgido oportunidades de colaboración entre otras entidades y ABRE PR. Algunas de estas oportunidades han surgido por acercamientos que han hecho estas entidades a la organización, pero también han surgido por los esfuerzos del equipo de ABRE PR.

De hecho, una de las colaboraciones más destacables desde el lanzamiento fue con El Nuevo Día, el periódico de mayor circulación en la isla tanto la versión impresa como su portal web. Ese trabajo se centró en la compilación de todos los presupuestos del Estado Libre Asociado de Puerto Rico desde el 1970 hasta el presente. Dichos datos se suministraron al periódico y también se publicaron en el portal. La colaboración resultó en una serie investigativa por parte del periódico donde destacaban los tipos de gastos en los que ha incurrido el Gobierno, una comparación de las plataformas de Gobierno versus sus prioridades fiscales según los gastos en los que incurrían, y los efectos de la asignación de fondos en la crisis fiscal aguda que actualmente experimenta la isla. Esta colaboración destaca la importancia de la existencia de un depósito de datos gubernamentales accesibles para la prensa.

Otra colaboración que se debe destacar es con la Fundación por Puerto Rico, una organización sin fines de lucro cuyo propósito es desarrollar una estrategia sustentable e identificar oportunidades de crecimiento socio económico para la isla. Esa organización conocía de ABRE PR antes del lanzamiento público y ha buscado la manera de apoyar sus esfuerzos. Han provisto a la organización de estudiantes para que realicen sus prácticas y para apoyar con las tareas de ampliar el catálogo de datos, además, estarán donando un espacio para albergar el personal de ABRE PR. Claro está, dicha fundación reconoce el valor de los datos abiertos para con el sector empresarial, ya que se torna una herramienta útil al momento de tomar decisiones.

Actualmente ABRE PR está buscando establecer una colaboración con una plataforma académica de estadísticas que dejó de existir en el 2008, pero aún está disponible por la web. Dicha plataforma fue de gran utilidad para la academia y otros sectores ya que se utilizaba para realizar investigaciones.

En efecto, los últimos cuatro meses desde nuestro establecimiento han estado llenos de oportunidades que denotan el impacto positivo en los distintos sectores de la isla. El interés que han demostrado los ciudadanos en la organización refleja el valor que encuentran en el quehacer de ABRE PR. Sin embargo, no basta con eso. Los próximos pasos para ABRE PR será ejecutar su plan de impacto comunitario para asegurar que esta información llegue a una mayor audiencia donde se destaque el uso y la importancia de la accesibilidad a datos gubernamentales.

Aparte del plan de impacto comunitario, se está desarrollando una funcionalidad de intercambio de información en la página donde cualquier persona pueda realizar una petición de datos gubernamentales a agencias públicas. Una vez se realice una petición, automáticamente se enviará a un representante designado en la agencia gubernamental correspondiente con copia al personal de ABRE PR. De este modo, ABRE PR podrá monitorear y divulgar las respuestas de estas entidades. Igualmente, los datos que se suponen provean las agencias serán enviados al peticionario con copia a ABRE para su colocación en la plataforma y asegurar que estén disponibles de manera permanentes sin la necesidad de solicitarlas nuevamente a las agencias. De esta manera se amplía automáticamente el catálogo de datos y se asegura que dicho catálogo responda a las necesidades de los ciudadanos.

La plataforma aún está en pleno desarrollo. Existen varias vertientes de cómo se puede fomentar la transparencia gubernamental mediante el acceso a datos por lo que el terreno es fértil para este esfuerzo en Puerto Rico. Actualmente, ABRE PR está buscando fuentes que aseguren su sustentabilidad. Por tal razón, se buscan maneras de monetizar algunos servicios y cumplir con los objetivos de mayor transparencia y apertura gubernamental. Ante el apoyo que hemos recibido y la efervescencia del movimiento de datos abiertos, podemos decir que el futuro es prometedor.

Edgar Ríos

Edgar Ríos es el Oficial de Proyecto de ABRE Puerto Rico. Es responsable por las operaciones y

esfuerzos diarios de la organización. Posee más de diez años trabajando con organizaciones sin fines de lucro en entidades educativas. Ha colaborado en entidades en Puerto Rico, México y distintas ciudades en Estados Unidos. También se ha desempeñado como asociado de investigación en una firma de investigación en política pública. Posee una maestría en Administración Pública con concentración en el manejo de organizaciones sin fines de lucro. Posee otra maestría en Ciencias Sociales con especialidad en métodos de investigación.

Paraguay, un país de “papel manteca”

Maricarmen Sequera

En la era digital, en la transiciones de nuestras sociedades del mundo analógico a la era de la sociedades de la información, el acceso y manejo de datos por parte de la ciudadanía comienza a tornarse clave para la transparencia de los gobiernos y al mismo tiempo se traduce como un derecho humano. Este proceso en Paraguay es bastante lento e incipiente, la tendencia de datos abiertos se está instalando de a poco y esto se puede observar en el plan de acción de Gobierno Abierto de 2011 con la publicación de datos del Ministerio de Hacienda y Dirección de Contrataciones Públicas, y con un mayor compromiso por parte del Gobierno paraguayo en el plan de acción 2014-2016 para Gobierno Abierto en lo que se refiere datos abierto exclusivamente.

Asimismo, la sociedad civil paraguaya se va empoderando de la tecnología en favor de la transparencia y monitoreos de las rendiciones de cuentas acerca de las actividades del Gobierno. Hasta la fecha existen algunas iniciativas interesantes acerca de uso de datos a gran escala, como punto de partida para análisis, investigación y hasta triangulación convirtiéndose en inteligencia útil para mayor conciencia e incidencia de la ciudadanía paraguaya.

El periódico ABC Color publica “El buscador del pueblo”, la base de datos de funcionarios públicos de Paraguay proveída por el Gobierno y ciudadanos/as que acercaron información “omitida y/u olvidada” por el estado paraguayo. Sobre este último punto, según Mabel Rehnfeld, periodista de ABC Color, tuvieron 4000 correos electrónicos aproximadamente provenientes de la sociedad civil. Otra interesante experiencia sobre periodismo de dato es de un twitterero paraguayo César Sánchez, quién geolocalizó todos los locales comerciales de la ciudad de Asunción que prohibían el acceso a congresistas que votaron en contra del [desafuero de senadores en noviembre de 2013](#).

Otro importante acontecimiento es el nacimiento del capítulo [Hacks Hackers Asunción](#) en mayo 2013, que busca crear una comunidad de periodismo de datos e insertar el concepto en los departamentos de investigación de los periódicos paraguayos y usuarios de internet con talleres de capacitaciones y formaciones continuas. Algunos de los prototipos creados en la comunidad son: [“Compras hechas por vía de excepción del Estado Paraguayo 2003-2013”](#) y [“Plataforma de visualización del Presupuesto General de la Nación: ¿A dónde va mi plata?”](#)

En cuanto a normas nacionales e internacionales que protegen la actividad periodística, está la Constitución Nacional que declara como un derecho la libertad de informar y ser informado en sus artículos 26 “De la Libertad de expresión y prensa”, Art. 27 “Del empleo de los medios masivos de comunicación social”, Art. 28 “Derecho a Informarse” y Art. 29 “De la libertad de ejercicio del Periodismo”. Además es reconocido por los principales tratados internacionales de derechos humanos suscritos por el país y por lo tanto este se obliga internacionalmente a respetarlo: el Pacto Internacional de Derechos civiles y Políticos y la convención Americana sobre Derechos Humanos y el Derecho Internacional Humanitario (DIH). Hasta el momento Paraguay no cuenta con una ley de medios en comparación a los países de la región.

Asimismo, desde hace 9 años aproximadamente desde la sociedad civil organizada (Grupo Impulsor de Acceso a la Información -GIAI) se está realizando asesoría y debate en el Congreso Nacional para la sanción de la Ley de Acceso a la Información. Si bien, “la liberación de datos (e información) responde a los principios de transparencia, promoción de la innovación y participación ciudadana”, según [Open Knowledge Foundation](#). La coalición del GIAI se enfrenta a conservadores congresistas y al monopolio de medios de comunicación que acceden a la información pública por sus medios, que siempre tienen la exclusiva y por tanto [no están de acuerdo con esta iniciativa](#). Existe una alta probabilidad que este 2014 entre en vigor la ley luego de la presión por parte de la sociedad civil con

el apoyo de Organizaciones Internacionales y de Cooperación.

A pesar de que existan estos indicios de cambios, Paraguay “lleva una larga transición hacia la democracia (1989-2014) tan larga como su dictadura (Alfredo Stroessner 1954-1989)” lo describe Yeny Villaba, Abogada especialista en DDHH en su artículo sobre “Libertad de Expresión en Internet.”: “Un país con ciudadanos, hijos de la dictadura, con pensamientos conservadores como un papel manteca; existe, pero no se puede reescribir ni borrar.” Leyes que no responden a las necesidades de un país con 40% de la población rural y 35% en situación de pobreza, [con muerte de periodistas en el interior del país por casos de narcotráfico](#).

No obstante, los datos abiertos y el periodismo de datos facilitarán a los ciudadanos el acceso a la información y al conocimiento. No hay seguridad de que esto llevará a reflexionar, incidir e impactar para una mejor política pública. Desde la comunidad Hacks/Hackers Asunción se ha replanteado esta interrogante, sin embargo el compromiso y la colaboración de jóvenes universitarios amantes de la tecnología es uno de los primeros manifiestos de la participación ciudadana digital en cuestiones de transparencia del Gobierno. Este tímido y pequeño paso debe ser orgánico.

Maricarmen Sequera (@marsebu)

Abogada especialista en Propiedad Intelectual. Directora Ejecutiva de TEDIC.org. Líder pública de Creative Commons en Paraguay. Miembro de Hacks Hackers Asunción.

Cómo pedir información pública en Argentina

Romina Colman

A nivel normativo, el derecho de acceso a la información pública se encuentra contemplado en la [Constitución Nacional](#) en sus artículos 1, 14, 38, 41, 42, 43 y 75 inciso 22. En este último caso, se brinda jerarquía constitucional a una serie de tratados que reconocen entre otras cosas, el derecho de las personas a buscar, recibir, pedir y difundir informaciones e ideas.

Sin embargo, en la Argentina aún no contamos con una ley nacional en la materia pero sí un decreto, el 1172/03 que en su anexo VII regula el mecanismo de acceso a la información en el ámbito del Poder Ejecutivo Nacional y a las “organizaciones privadas a las que se hayan otorgado subsidios o aportes provenientes del sector público nacional, así como a las instituciones o fondos cuya administración, guarda o conservación esté a cargo del Estado Nacional a través de sus jurisdicciones o entidades y a las empresas privadas a quienes se les hayan otorgado mediante permiso, licencia, concesión o cualquier otra forma contractual, la prestación de un servicio público o la explotación de un bien del dominio público”.

También es importante destacar que en los municipios y provincias existen leyes u ordenanzas para garantizar el acceso a la información. (Ver mapa elaborado por el equipo de [LA NACION DATA](#) junto a la Asociación por los Derechos Civiles, el Foro de Periodismo Argentino y la Fundación Mujeres en la Igualdad).

Libertad de prensa:

Artículo 14, 32 y 43, además de su reconocimiento en los tratados internacionales con jerarquía constitucional.

Otras normativas sobre transparencia gubernamental:

Declaraciones juradas: la [Ley 25.188](#) (1999) sobre la ética en el ejercicio de la función pública establece qué funcionarios están obligados a presentar declaración jurada patrimonial, para rendir cuenta de sus bienes. Además, en su Artículo 10, especifica que “toda persona podrá consultar y obtener copia de las declaraciones juradas presentadas con la debida intervención del organismo que las haya registrado y depositado, previa presentación de una solicitud escrita”. También, prevé que la declaración jurada “deberá contener una nómina detallada de todos los bienes, propios del declarante, propios de su cónyuge, los que integren la sociedad conyugal, los del conviviente, los que integren en su caso la sociedad de hecho y los de sus hijos menores, en el país o en el extranjero.”

Ahora bien, en el 2013 se sanciona la [Ley 26.857](#) sobre el carácter público de las declaraciones juradas patrimoniales de los funcionarios públicos. Esta normativa tiene como punto positivo que ordena la publicación en Internet de las declaraciones juradas. Sin embargo, “se envía” al anexo reservado la información patrimonial del cónyuge, conviviente e hijos menores no emancipados, lo cual implica un gran retroceso en relación a la cantidad/calidad de datos a la cual se podía acceder con anterioridad. Además, con la reglamentación de la nueva ley cambió el formulario en el que se detalla la información patrimonial y así se “recortaron” datos que eran de vital importancia. En particular, se agrupó más la información y se perdieron niveles de desagregación. Es decir, con la nueva ley se facilita el mecanismo de acceso, pero disminuye en forma sustancial la información a la que pueden acceder los ciudadanos.

Dos informes elaborados en 2013, uno de la Asociación Civil por la Igualdad y la Justicia (ACIJ) y otro de la Asociación por los Derechos Civiles (ADC), muestran que las respuestas a las solicitudes de acceso a la información son dispares.

Sobre este punto, Ramiro Álvarez Ugarte, director del área de acceso a la información de ADC al

hablar sobre los dos informes en el blog de LA NACION DATA, afirmó: “De las 44 solicitudes presentadas por la [ACIJ](#) [año 2013], 14 no fueron respondidas y 14 fueron contestadas fuera del plazo de los diez días. En el caso de la [ADC](#), sobre un total de 49 pedidos, el 47% fue respondido satisfactoriamente, el 53% de forma incompleta, de manera negativa, o directamente no fueron contestados. También se registraron retrasos en el 55% de los pedidos respondidos (...) las conclusiones a las que hemos llegado por diferentes caminos evidencian respuestas incompletas a los pedidos realizados sin justificación de las razones por las cuales no se contesta a lo solicitado y demoras injustificadas”.

En relación al decreto 1172/03, el primer desafío con el que uno se encuentra a la hora de preguntar es que los pedidos deben realizarse por escrito y en papel. En la actualidad, no existe un sistema electrónico que permita generar, dar seguimiento y recibir la información de manera online.

Para enfrentar los obstáculos, en nuestro blog redactamos [diez consejos](#) útiles para facilitar el ejercicio de este derecho:

El pedido: debe entregarse en la [mesa de entradas](#) de la entidad obligada a proveer esa información. Es importante llevar una copia. Allí colocarán la fecha y sello de recepción. También es fundamental exigir el número de expediente de la solicitud.

¿Otra opción? Enviarlo por correo postal, preferiblemente con acuse de recibo. En ambos casos, estos documentos serán la constancia para reclamar a la [Oficina Anticorrupción \(OA\)](#) ante el silencio del organismo, respuesta parcial o si lo entregado no corresponde a lo consultado. En 2008 la Subsecretaría para la Reforma Institucional y Fortalecimiento de la Democracia (SPRIyFD) creó un [formulario voluntario de información pública](#). Quien desee usarlo, sólo está obligado a completar la primera sección. La segunda es para datos estadísticos.

Información de contacto: el decreto 1172/03 solicita la “identificación del requiriente”. La normativa no aclara los datos a los que se refiere este punto. No puede faltar el nombre y apellido, DNI, domicilio completo (dirección, ciudad y código postal), firma y teléfono.

Horario de atención al público de las mesas de entradas: por lo general es de 9 a 17. En el sitio web de la Jefatura de Gabinete de Ministros (JGM) se detallan las direcciones y teléfonos de las entidades obligadas por la normativa. Antes de ir, es preferible llamar a la dependencia para verificar el horario.

Simple es mejor: no hay que dar vueltas con la pregunta. La clave está en “ir al grano”. Contar con conocimientos mínimos sobre el tema es una ventaja. Quien está informado pregunta mejor. En caso de que la petición no sea clara, la dependencia podrá contactarse con la persona dentro de los 10 días hábiles de la presentación para pedir aclaración. Esta situación no suspende el cómputo del plazo para recibir respuesta.

Preguntas frecuentes

¿Es necesario ser ciudadano argentino para presentar un pedido de acceso a la información pública en nuestro país? No. “Toda persona física o jurídica, pública o privada, tiene derecho a solicitar, acceder y recibir información”.

¿Debe realizarse la presentación a través de un abogado? No. La letra del decreto establece que no se requiere de patrocinio letrado para pedir información pública.

¿Cuál es el tiempo previsto para recibir respuesta? Los organismos obligados por el decreto 1172 tienen hasta 10 días para responder a la solicitud. Pueden pedir una ampliación de este plazo por 10 días más, y por única vez, en caso de “mediar circunstancias que hagan difícil reunir la información solicitada”. Si esto sucede, deberán remitir a quien pide la información una notificación por escrito fundando el uso de días adicionales. Es decir, explicar por qué se necesitan más días para responder al

pedido.

¿Se debe explicar el motivo de la presentación? No. “No puede exigirse la manifestación del propósito de la requisitoria”.

¿Cuál es la autoridad de aplicación del Reglamento de Acceso a la Información Pública? La SPRIyFD, dependiente de la JGM. Franco Vitali Amado es quien está actualmente a cargo de este organismo.

¿Qué hacer si la entidad, aun estando obligada a recibir el pedido, se niega a hacerlo? En este caso, se tiene que presentar la denuncia, junto con la solicitud, en la [Oficina Anticorrupción](#) (OA). Dentro de los dos días esta institución la enviará a la SPRIyFD, que remitirá el pedido al Responsable de Acceso a la Información de la jurisdicción denunciada. Una vez recibida por este sujeto, comenzará a correr el plazo de diez días hábiles.

¿Dónde presento la denuncia ante la ausencia de respuesta o contestación parcial? En la OA. La denuncia tiene que formularse por escrito, con la identificación de quien la realiza.

Si no se recibió respuesta: Acompañar copia del pedido con el sello de la entidad, fecha y número de expediente.

Si la respuesta entregada por la institución no es satisfactoria: Adjuntar copia del pedido y de la información brindada por el organismo.

¿Cómo sigue el proceso? La OA informa a la jurisdicción denunciada y a la SPRIyFD sobre la presentación. A continuación, a través del Responsable de Acceso a la información, se pide el descargo al organismo que tendrá que entregarlo dentro de los 20 días hábiles y por escrito. Luego, la OA remitirá al denunciante la copia de este documento o notificará de su falta de presentación. A partir de ese momento, el sujeto contará con 15 días hábiles para realizar observaciones al respecto, en caso de querer hacerlo.

Según la Resolución Conjunta 1/2008, dentro de los 20 días hábiles de cumplido este plazo, la OA enviará un informe preliminar con la actuación administrativa a la autoridad de aplicación, la que deberá resolver el caso. ¿Cuál es la debilidad de este proceso? La SPRIyFD únicamente puede emitir recomendaciones y no puede obligar a los organismos a cumplir con ellas. Esto representa un enorme problema para el funcionamiento del sistema en general.

Es cierto que puede recurrirse a la Justicia, pero sus tiempos pocas veces coinciden con los de persona que quiere acceder a la información y una de las características de este derecho es la posibilidad de ejercerlo con la mayor celeridad posible.

A través de la [resolución 538/2013](#) se creó el Sistema Nacional de Datos Públicos (SINDAP). Entre los principales objetivos del programa SINDAP, según el texto de esta normativa, se encuentran administrar el Portal, facilitar la publicación de datos y alentar el desarrollo de aplicaciones.

Entre las formalidades que deben cumplir los datos que se suban al portal, la resolución exigen el empleo de un formato abierto y la utilidad pública de los datos.

Además del [Portal Nacional de Datos Públicos](#), existen otros sitios de datos públicos en otras jurisdicciones: [Ciudad Autónoma de Buenos Aires](#), (Buenos Aires Data), [Bahía Blanca](#) y [Misiones](#).

Más allá de estos sitios, también se evidencian avances que no necesariamente cumplen con los requisitos de formatos abiertos, pero muestran un aumento en el compromiso para mostrar cómo se gastan los fondos públicos. Un ejemplo de ellos es la publicación de los sueldos de los funcionarios de la ciudad de Rosario en el [sitio oficial del municipio](#).

El ingreso de la Argentina a la Alianza para el Gobierno Abierto brinda esperanzas para avanzar en materia de transparencia gubernamental. En la actualidad hay esfuerzos locales e incluso nacionales

pero es vital la voluntad política para generar cambios en la cultura de la administración pública y fomentar la conciencia ciudadana al respecto. Para esto es clave la sanción de una ley nacional de acceso a la información pública.

Hay dos puntos clave que se deberían trabajar más fuertemente en materia de portales de datos públicos. El primero de ellos, vinculado a la necesidad de subir datasets de alta relevancia para las personas. Por otro lado, el tema del trabajo de actualización de los datos existentes.

Romina Colman (@colmanromi)

Estudiante de Comunicación de la Universidad de Buenos Aires, activista por el acceso a la información pública en Latinoamérica y delegada argentina en la primera conferencia de Alaveteli. Es parte del equipo de La Nacion DATA.

Refinando datos para desvestir autoridades

"Verita": para que los candidatos y candidatas digan #sololaverdad

Ernesto Cabral

Este año 2014 se llevarán a cabo las elecciones municipales y regionales en el Perú. Estos procesos siempre son una oportunidad para los periodistas, pero también están llenos de dificultades. ¿Cómo fiscalizar a más de cien mil candidatos?, fue la pregunta que nos hicimos los miembros del blog [El Útero de Marita](#), dirigido por el periodista Marco Sifuentes.

Existía una posibilidad. Cada candidato tiene la obligación de presentar una hoja de vida ante el Jurado Nacional de Elecciones (JNE). En ella deben indicar desde sus nombres completos, el cargo al que postulan y la agrupación a la que pertenecen; hasta sus antecedentes penales y declarar los bienes que poseen, entre otras cosas ¿Por qué no aprovechar toda esta data?

Es así que, en alianza con la Asociación Civil Transparencia, nace "Verita" (@verita_pe). Se trata de un proyecto digital que procesa y analiza los datos de las 116.252 hojas de vida que el JNE puso a nuestra disposición.

Como encargado de procesar la data, seguí los siguientes pasos:

- El JNE proporcionó un archivo Excel que contenía siete hojas distintas. Con la ayuda del hacker Aniversario Perú, separamos cada una individualmente y las convertimos a archivos .csv
- Estos archivos los subí a Open Refine para limpiar cada hoja y normalizar los datos. Por ejemplo, los candidatos escribieron un mismo delito de diferentes formas ('homicidio', 'omisidio').
- Los antecedentes penales y civiles estaban en dos hojas separadas. Sin embargo, Transparencia encontró que los candidatos no habían respetado esta división. Por eso, decidí juntarlas en una sola tabla utilizando una hoja de cálculo en Google Drive.
- Una vez obtenida, en Open Refine eliminé las filas donde el candidato indicaba que no había sido condenado o la información que proporcionaba no permitía asegurarlo.
- Finalmente, agrupé las acusaciones según los delitos o faltas más generales. Por ejemplo, la categoría "Alimentos" recoge datos como "pensión alimentaria", "alimentaria", entre otros.

La base de datos resultante ha sido el eje sobre el cual hemos obtenido información relevante, que es publicada en <http://utero.pe>. Hasta ahora se ha descubierto lo siguiente:

- Aproximadamente mil cuatrocientos candidatos tienen sentencias penales y civiles. La información fue presentada en un gráfico elaborado en [Jsfiddle](#).
- Trece candidatos han sido condenados por el delito de tráfico ilícito de drogas. Con la ayuda de Aniversario Perú -que confrontó esta lista con las resoluciones ministeriales que el Ministerio de Justicia aloja en su página web- encontramos que dos candidatos habían recibido una conmutación de pena por dos ex presidentes. Uno de los ex mandatarios es acusado de cometer una infracción constitucional al conmutar las penas de narcotraficantes.
- Informamos acerca de los partidos políticos que tienen en sus filas a la mayor cantidad de candidatos con sentencias. Aquí tuvimos otro problema: la información sobre las agrupaciones políticas se encontraba en una hoja diferente a nuestro compilado de antecedentes penales y civiles.

Como solución, el hacker Aniversario Perú elaboró un script que compara los DNI de los candidatos en los dos archivos diferentes. Si los DNI son idénticos, extrae la organización política de una hoja y

la pasa a la otra. [El script utilizado puede ser revisado y descargado](#). La información también fue presentada en un gráfico elaborado en [Jsfiddle](#).

Estos hallazgos son revisados y confirmados por Transparencia, que realiza a la par una limpieza y filtrado de las hojas de vida. La asociación civil también se encarga de realizar reportes de nuestros hallazgos, que son remitidos al Jurado Nacional de Elecciones. Además, actualiza la cuenta de Twitter de “Verita”, donde también se publican los resultados.

Como se puede ver hasta ahora, no fueron pocos los problemas que encontramos. Una de las principales causas de esta situación es que el formato de las hojas de vida no reduce lo suficiente la posibilidad de error a la hora de llenarlas. Aunque también hay un número importante de candidatos que omiten información, sobre todo de antecedentes penales y civiles.

Con el transcurso de los meses, esperamos consolidar el trabajo de “Verita”. A corto plazo, esperamos seguir procesando la información. Por ejemplo, queremos elaborar mapas de calor (utilizando Open Refine, el API de Google Maps y CartoDB) que nos indiquen en qué regiones del Perú se concentran la mayor cantidad de candidatos con sentencias.

A mediano plazo, queremos elaborar un buscador de disposición pública parecido a [“Manolo”](#), otro proyecto elaborado por nuestro equipo. El objetivo es que los ciudadanos puedan ejercer su derecho al acceso de información pública, en este caso sobre sus candidatos. Y que estos últimos se acostumbren a decir #sololaverdad.

Ernesto Cabral (@ErnestoCabralM)

Es periodista, escribe regularmente en el blog [El Útero de Marita](#).

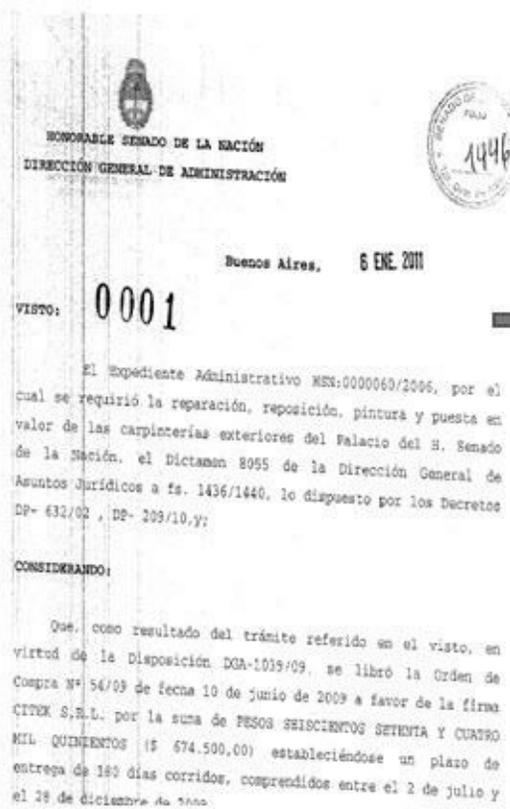
La Nación: Gastos en el Senado en la Argentina 2004-2014

Un proyecto de periodismo de datos, colaboración y apertura

Angélica Peralta Ramos

A fines de 2010, en la Argentina aún no existían portales de datos abiertos ni Ley de Acceso a la Información Pública y los casos de corrupción vinculados a la administración pública eran un problema creciente. Sin embargo, LA NACION decidió comenzar un proyecto propio de “periodismo con bases de datos”. Para lograr este objetivo, fue necesario comenzar desde cero a construir las bases para luego abrirlas en la publicación de sus artículos periodísticos.

La investigación sobre los “Gastos en el Senado” comenzó en 2012 y tuvo su origen en un informante anónimo que hizo llegar a nuestra redacción, vía correo electrónico, una fotografía del despacho de Amado Boudou en la que se veía una mesa importada de Italia. Luego se detectaron en el sitio oficial del Senado tres datasets con archivos en formato .pdf de contrataciones desde 2004, que descargó el equipo de LA NACION Data. Como estos documentos eran imágenes escaneadas en .pdf, se transformaron a texto aplicando distintas técnicas de reconocimiento de caracteres (OCR) y luego fueron llevados a formato .xls (Excel), estructurando los registros mediante la extracción de entidades y fechas, y utilizando expresiones regulares (Regex).



Durante 2012 y 2013, gracias a este dataset y a numerosas investigaciones periodísticas, se publicaron artículos en la tapa de LA NACION, respuestas por parte del vicepresidente en ejercicio Amado Boudou y de quien anteriormente ocupara su cargo (a su vez presidentes del Senado) y una investigación judicial con relación a esos gastos que involucraba al mismo Boudou. Además, el trabajo del equipo permitió descubrir que algunos de los gastos correspondientes a viajes oficiales fueron presentados con fechas que se superponían entre sí o, incluso, que algunos de esos viajes directamente no se habían realizado.

El tema tuvo una amplia repercusión en diferentes programas de televisión y radio, y también en otros periódicos competidores de LA NACION. Y en 2013, la investigación fue premiada con el Data

Journalism Awards en la categoría Best Data Driven Investigation.

En qué consistió la investigación

El trabajo estuvo dividido en varias fases, cada una con requerimientos diferentes en cuanto a la utilización de distintas herramientas:

Primera fase. Descarga de los archivos .pdf, que no eran otra cosa que imágenes escaneadas de los documentos en papel. Desarrollamos una aplicación, que se conectaba con el sitio del Senado y buscaba en secciones diferentes los .pdf ([Decretos del Senado](#), [Departamento Administrativo del Senado](#) y [Departamento Contable del Senado](#)).

Segunda fase. Remoción de la protección de los .pdf contra copia e impresión.

Tercera fase. Conversión de los .pdf a archivos rastreables con Omnipage 18 (Batch Processing).

Cuarta fase. Análisis de los datos. La misma aplicación mencionada en la primera fase abre los archivos .txt, realiza búsquedas de nombres de senadores, compañías, los montos de los gastos (en pesos, dólares, euros y libras esterlinas), fechas, palabras claves (como “compra”, “compra directa”, “agentes de seguridad”, “viajes”, “muebles”, “transporte aéreo o terrestre”, etc.) y además permite insertar los textos completos y cada una de estas entidades en diferentes filas, asignando una fila a cada uno de los 33 mil archivos .txt.

Quinta fase. Las 33 mil filas obtenidas en una hoja de cálculo pudieron ser utilizadas para hacer investigaciones, simplemente con la aplicación de un filtro de Excel.

Sexta fase. La base de datos mencionada en la quinta fase fue importada desde un software de administración de proyectos para generar un gráfico de Gantt que mostraba, en una línea de tiempo, la distribución de los viajes y las superposiciones existentes.

Las Historias

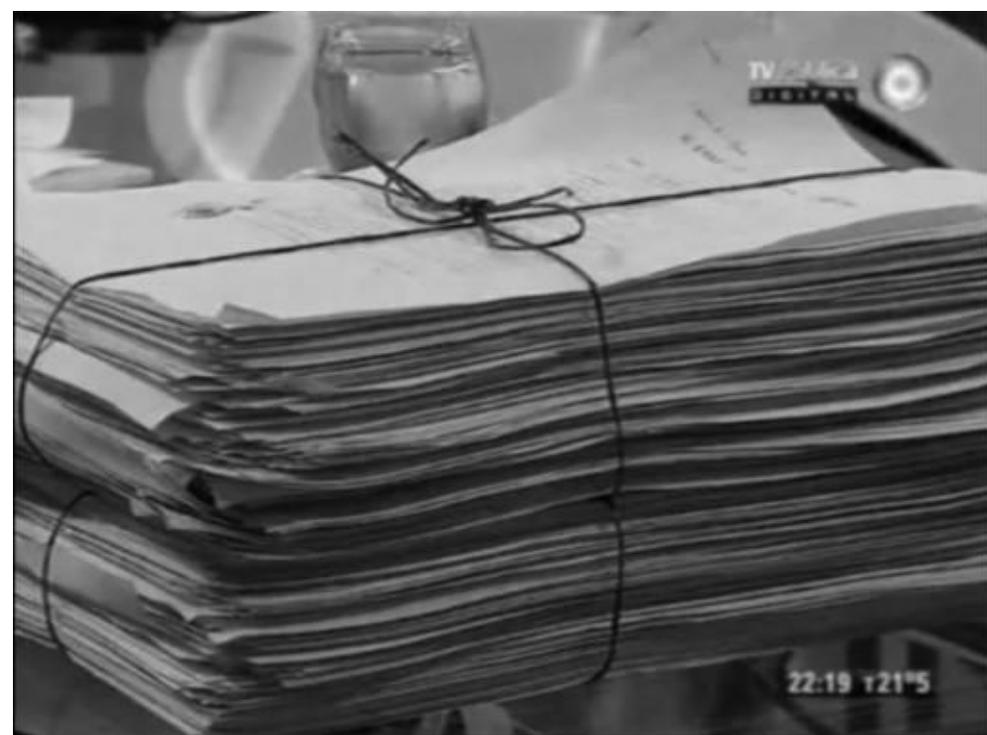
Surgieron numerosas historias de la misma base de datos. A continuación, algunas de ellas:

[Millones gastados en viajes](#)

Laura Serra, periodista de la sección Política, escribió el primero de una serie de artículos que provocaron muchísima repercusión en la opinión pública. Por ejemplo, para una conferencia de un día en Suiza el vicepresidente viajó por 6 días con 4 custodios y 7 asistentes, a un costo de US\$ 100 mil. Asimismo, los custodios solicitaron US\$ 10 mil adicionales “para imprevistos” y gastaron US\$ 10.820.

Para esta misma historia, publicada el 10 de febrero de 2013, se ordenaron las cifras en un gráfico interactivo desarrollado con la plataforma Tableau, presentadas según la cantidad de dinero gastada -y no por fechas-, análisis que se dejaría para un trabajo posterior.

¿Cuál fue el impacto de la publicación? Esa misma noche, el vicepresidente Amado Boudou apareció en un programa de la TV Pública ([ver video](#)) para responder a la publicación, y se excusó mostrando en cámara pilas de papeles (sin enfocar detalles) e indicando que en las mismas se encontraban los decretos a los que LA NACION no podía acceder en la web.



Al día siguiente, Laura Serra, fue invitada al programa del periodista [Ernesto Tenenbaum](#) para explicar en persona el detalle de su documentada investigación. Esta historia fue dividida en artículos diferentes.

[“Boudou gasta fondos de emergencia para comprar muebles de lujo”](#) y [“El vicepresidente Boudou omite informar al juez Oyarbide sobre la compra de estos muebles”](#)

En la misma semana, el 15 de febrero de 2013, otra periodista de la sección Política de lanacion.com, Maia Jastreblansky, encontró en la misma base los datos de la compra en forma directa de los muebles de lujo por el doble del monto permitido, sin licitación previa. La adquisición fue realizada por un procedimiento que se reserva a situaciones de emergencia. También se descubrió que estos gastos no fueron expuestos al juez que investigaba el caso.

Otra vez, el vicepresidente alegó que había recibido la oficina en mal estado, dichos que luego fueron desmentidos en forma pública por su antecesor en el cargo, Julio Cobos.

¿Cuál fue el impacto de la publicación? Luego de publicada, un juez anunció la [reapertura de la causa](#) por los gastos excesivos del vicepresidente.

[“Los viajes sospechosos de Boudou”,](#) y además, [“El Senado pagó viáticos por viajes no realizados”](#).

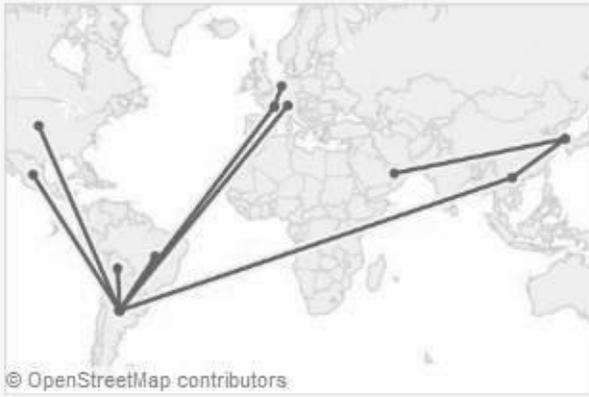
El 3 de abril de 2013, LA NACION reveló viáticos pagados para fechas que se superponían mediante un gráfico Gantt volcado en una visualización interactiva. Esta pieza también muestra la rendición de gastos por viajes que finalmente fueron cancelados. Los gastos fueron extraídos de los mismos .pdf e incluidos como documentos originales para documentar y darle sustento al artículo.



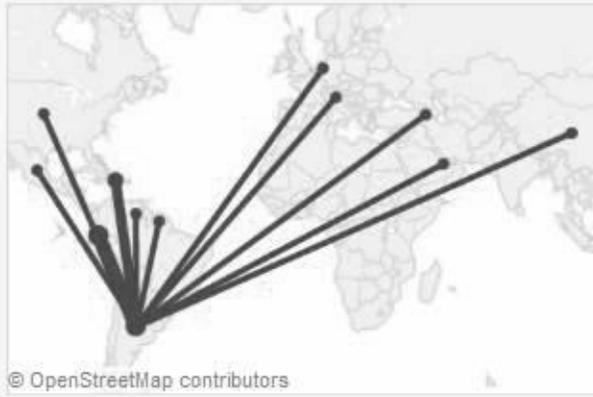
LAS MISIONES INTERNACIONALES DE BOUDOU

En lo que va del año ya realizó el doble de viajes al exterior que en 2012

7 misiones en 2012



14 misiones en 2013 (hasta agosto)



Los viajes de Boudou en 2013: ya realizó el doble de misiones al exterior

El vicepresidente lleva 14 salidas internacionales, muchas veces en representación de Cristina Kirchner. [Los ampulosos viáticos de su comitiva con la caja del Senado.](#)

Boudou lideró la lista de viajeros y se destacó por su amplia comitiva

Fue acompañado por asesores y custodios que podían realizar [gastos extras en dólares o euros.](#)

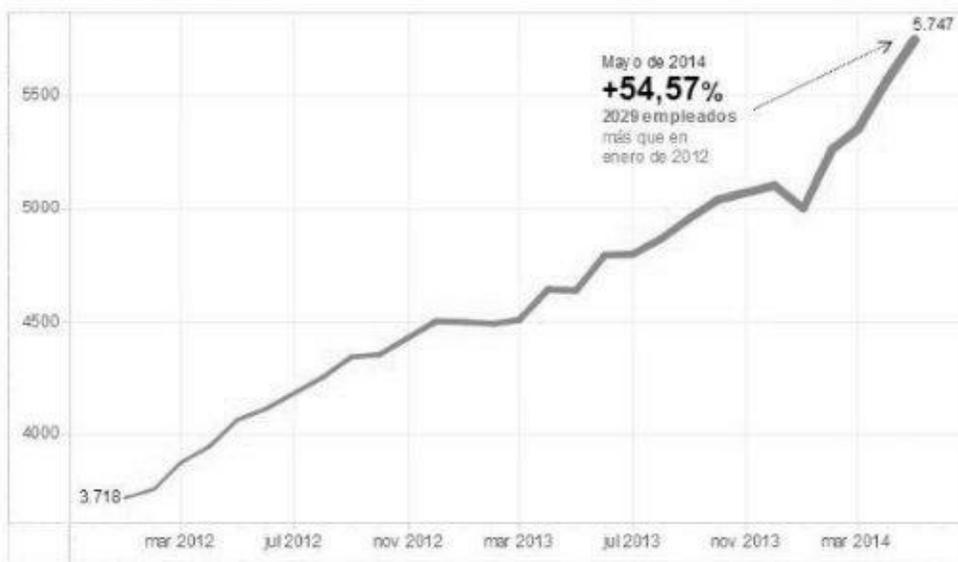
El Senado de Amado Boudou: su gestión sumó 2000 empleados más

La planta creció un 55% desde su llegada; este año se incorporaron, en promedio, más de dos personas por día.

Bajando mes a mes los listados de empleados del Senado, ubicados en páginas que de otra manera se pierden porque se pisan en el mismo URL, el equipo pudo analizar la evolución de los mismos, y los periodistas Iván Ruiz y Maia Jastreblansky fundamentaron con datos los rumores de crecimiento en la [cantidad de empleados y contratados en este organismo.](#)

CRECIMIENTO DEL PERSONAL DEL SENADO EN LA GESTIÓN DE BOUDOU

Periodo: Enero 2012 - Mayo 2014



VozData y las Maratones Cívicas

Si bien el trabajo de conversión de los OCR sirvió para detectar las primeras historias, incluso las que generaron mayor impacto, era evidente que esos datasets estaban perdiendo información ya que muchos archivos .pdf estaban mal escaneados, desalineados o grisados. Fue necesario entonces pedir

ayuda para procesar este material, limpiarlo y clasificarlo. Es por esta razón que se decidió desarrollar una plataforma que permitiera estructurar y abrir bases de datos de manera colaborativa, bautizada como “VozData”.

VozData es una plataforma de colaboración abierta creada para transformar documentos públicos en información útil. Se trata también de una iniciativa para amplificar la voz de los datos que de otra manera permanecen distantes del control o la participación ciudadana. La aplicación se inspiró en los proyectos “Free the Files” de Propublica.org y “MP’s Expenses” de The Guardian y fue desarrollada por Opennews Fellows en LA NACION, con la participación de algunos integrantes del equipo de LN Data.



El primer proyecto “Gastos del Senado 2010-2012” consistió en estructurar y clasificar 6700 archivos .pdf, y se completó en dos meses gracias a la ayuda de 500 voluntarios, y con dos acciones llamadas “maratones cívicas”.



Las maratones consistieron en jornadas presenciales realizadas en la sede de LA NACION, de las que participaron usuarios de la plataforma y varias de las ONG y Universidades que colaboraron con [VozData](#) desde su lanzamiento. En simultáneo -y vía Hangouts- se conectaban desde las provincias dos universidades en Córdoba y Entre Ríos que también participaron del encuentro. El código de la plataforma VozData está liberado con el nombre de “Crowdata”, y los datos están abiertos en formatos Open Data.

Esta iniciativa de LA NACION demostró cómo hasta en contextos adversos, los medios de comunicación pueden ser proactivos, ir a fondo en el periodismo de datos, abrirse a la colaboración y

generar espacios de participación ciudadana para fomentar la transparencia, descubrir nuevas historias y acelerar los procesos de innovación.

Angélica Peralta Ramos (@momiperalta)

Gerente de Desarrollo Multimedia en el diario La Nación de Argentina. Licenciada en Sistemas y MBA. Co-fundadora de lanacion.com en 1995. Desde 2006 ha estado a cargo de La Nación Multimedia, un área de investigación, desarrollo y formación que ha sido clave en facilitar la actualización del diario en nuevas tecnologías de comunicación y periodismo digital. El sitio lanacion.com ha sido reconocido internacionalmente con 4 premios Eppy y dos premios ONA. Desde 2011 también ha liderado el proyecto de periodismo de datos LA NACION Data. El proyecto Gastos del Senado en Argentina recibió en 2013 el Data Journalism Awards, entregado por Google y Knight Foundation.

Un año y medio mapeando el poder: Poderopedia

Mónica Ventura

Fortalecer el derecho de acceso a la información pública y ubicar en perspectiva las relaciones entre los poderes político, económico y social, forman parte de los objetivos que ha impulsado Poderopedia desde su lanzamiento.

La plataforma de periodismo de datos, fundada en Chile en diciembre de 2012 por Miguel Paz, ICFJ Knight International Journalism Fellow, gracias al apoyo de la Fundación Knight y Start Up Chile, ya cuenta con tres capítulos en Latinoamérica:

- [Poderopedia Chile](#), con el apoyo de la John S. and James L. Knight Foundation a través de un fondo del News Challenge.
- [Poderopedia Venezuela](#), en conjunto con IPYS Venezuela y con el apoyo de Transparencia Venezuela y Coalición ProAcceso.
- [Poderopedia Colombia](#) en sociedad con Consejo de Redacción y con el apoyo de Open Society Institute Foundations y la Facultad de Comunicación y Lenguaje de la Pontificia Universidad Javeriana.

Esta expansión es el primer paso de Poderopedia en sus esfuerzos por participar en redes de cooperación entre organizaciones de periodismo y transparencia en Latinoamérica con el fin de fortalecer las democracias locales y promover la innovación.

Logros 2013-2014

Los conflictos de interés, las alianzas políticas, los empresarios que se esconden tras grandes grupos económicos y los dueños del poder, han sido el foco de atención de Poderopedia durante estos dos años desde su lanzamiento, logrando de esta forma revelar casos que han impactado a los medios y a la ciudadanía.

Desde entonces, el equipo de Poderopedia Chile - conformado por Mónica Ventura (investigadora en jefe), Juan Eduardo Hernández (principal desarrollador) y Miguel Paz (CEO y fundador)-, produjeron más de 10 casos de impacto sobre conflictos de intereses, informaron sobre negocios irregulares y posibles malas conductas políticas, además de haber aportado información al debate ciudadano y a las investigaciones periodísticas.

Hoy en día, la plataforma que se ha convertido en un lugar rico en información sobre los poderosos en Chile, reconocida internacionalmente como finalista de los Data Journalism Award 2013 y nominada a Mejor Innovación en los Bobs Awards de la Deutsche Welle 2014, ya cuenta con una base de datos de 3.292 personas, 1.485 empresas y 946 organizaciones y seis medios chilenos ya están republicando sus contenidos. Además ya tiene más de 1.500 perfiles completos con sus respectivas biografías, conexiones, mapas de relaciones y fuentes. Los usuarios registrados ascienden a 3.590 personas y cuenta con ocho voluntarios y colaboradores probono.

La plataforma además ha despertado el interés de profesionales quienes han aportado con sus conocimientos al proyecto en forma desinteresada y por iniciativa propia, tal es el caso del periodista Marcelo Ortiz, quien creó una [extensión de Google Chrome](#) que permite buscar entidades (nombre de persona, empresa u organización) directamente desde cualquier sitio.

Internacionalización

Desde un inicio el equipo de Poderopedia supo que convertir el proyecto en una plataforma de

distribución internacional, operada por periodistas y organizaciones locales, tendría que ser una meta primordial. Ahora ese sueño ya ha comenzado a ver la luz, pues Poderopedia lanzó el 3 de mayo de 2014, el Día Mundial de la Libertad de Prensa, un nuevo capítulo nacional en Venezuela, gracias a una alianza estratégica firmada con el Instituto Prensa y Sociedad (IPYS) de Venezuela. Además el 15 de junio, con ocasión de la segunda vuelta presidencial en Colombia, lanzamos el capítulo en Colombia, en alianza con Consejo de Redacción (CdR). Ambas organizaciones son muy respetadas y reconocidas en Latinoamérica por su labor de promoción y práctica del periodismo de investigación.

Casos de impacto

Poderopedia funciona gracias a un software especialmente diseñado que organiza toda la información que se extrae de plataformas públicas, tales como bases de datos gubernamentales y de empresas, estados financieros o actas de directorios, para crear los mapas de conexiones. Lo anterior, junto al trabajo investigativo de un grupo de periodistas y colaboradores, han revelado en Chile más de 10 casos de impacto sobre conflictos de intereses, información sobre negocios irregulares y posibles malas conductas políticas.

Ejemplos:

- Medios, política y negocios: [Lo que debes saber de la venta de “La Nación”](#).
- [“Conoce a dos parlamentarios que están en las comisiones de las áreas donde tienen negocios”](#).
- ["El extraño episodio de un video porno que involucra al Gobierno Regional de Los Ríos"](#).
- ["La relaciones de poder de los Larraín Hurtado que salvarían a su hijo de ir a la cárcel por atropellar a una persona con resultado de muerte"](#).
- [“Grupo Cueto: Controladores de LAN, amigos y ex socios de Sebastián Piñera”](#).
- [“John O’Reilly, historia y redes del cura favorito de la elite chilena”](#).
- [“Corpesca, la pesquera que colabora con el Gobierno en disputa con Perú en La Haya”](#).
- Tras el debate organizado por la Asociación Nacional de Prensa en el marco de las elecciones presidenciales de Chile 2013, que reunió a ocho de los nueve candidatos presidenciales, Publimetro juntó los perfiles de Poderopedia y publicó [un artículo de cada uno de ellos](#).
- Investigación sobre el lado desconocido de [Marcel Claude](#), el candidato presidencial 2013 de la coalición Izquierda Unida.

Otros hitos y eventos importantes

1.- *Fact checking* en vivo durante el debate presidencial 2013 transmitido por televisión. Poderopedia chequeó y transmitió vía Twitter [las opiniones y declaraciones de los candidatos presidenciales](#).

2.- Nuevo gabinete de Michelle Bachelet. Poderopedia mostró quién es quién en el nuevo equipo de la Mandataria que fue elegida nuevamente Presidenta de la República a fines de 2013, proporcionando de esta forma [información valiosa a las salas de redacción](#).

Este evento fue publicado y promocionado en [W5](#).

3.- Monitoreo de las nuevas autoridades gubernamentales. El equipo Poderopedia realizó un seguimiento investigativo de los nombramientos de las nuevas autoridades del Gobierno de Michelle Bachelet. Este trabajo reveló que muchas de las nuevas autoridades designadas tenían problemas legales. La información fue publicada y recogida por distintos sitios web de noticias:

- Hugo Lara, el futuro subsecretario de Agricultura de Bachelet que enfrenta [querrela por estafa](#).
- Otro subsecretario designado en aprietos: Fue demandado por [estafa y apropiación indebida](#).
- Las causas que originaron que [cayeran designaciones del Gobierno](#).

4.- Monitoreo de las autoridades electas: Durante todo el período, el equipo de Poderopedia investigó a las nuevas autoridades electas, especialmente a los miembros y ex miembros del Parlamento. Como ejemplo, uno de los trabajos más relevantes fue el caso del senador Baldo Prokurica, en el que gracias a la investigación de Poderopedia se dio a conocer un [conflicto de interés del parlamentario en temas de energía y minería](#). Este caso fue tomado por varios medios en el país.

5.- Monitoreo de los conflictos de interés de empresarios y políticos en torno al ex Presidente Sebastián Piñera y su Gobierno. En enero de 2013 un monitoreo realizado por el equipo de Poderopedia reveló un conflicto de Interés de Andrés Chadwick, Ministro del Interior y primo del ex presidente Piñera. Durante ese mes un aluvión en la cordillera de Los Andes dejó a más de 2 millones de chilenos sin agua durante 2 días. La empresa sanitaria “Aguas Andinas” y el Gobierno fueron criticados por los chilenos, ya que éste tardó demasiado tiempo para restablecer los servicios básicos de agua potable. Poderopedia mostró que Andrés Chadwick tenía \$ 800.000 dólares en acciones en Aguas Andinas. Esta revelación llevó a que fuera criticado por los ciudadanos, las ONG y los medios de comunicación.

También se inició un debate público sobre la obligación de las autoridades de actualizar regularmente sus Declaraciones de Intereses. La investigación obligó a las autoridades del Gobierno hacer frente a los conflictos de intereses del ministro Chadwick. El también primo hermano del presidente fue criticado además cuando tres semanas después, el servicio de agua potable dejó de funcionar de nuevo durante un día. Esto, luego de que Poderopedia alertara que el funcionario de Gobierno había sido director suplente en Aguas Andinas.

El diario electrónico El Mostrador publicó un artículo con la revelación de Poderopedia, la que tuvo 2.012 retweet veces y fue compartida en Facebook 7.755 veces en menos de 48 horas.

6.- La revelación de la red de empresas del director de Impuestos Internos de Chile, Julio Pereira. En 2012 Pereira decidió condonarle a la empresa de retail “Johnson” una deuda de 119 millones de dólares en impuestos, dos meses antes Cencosud (otro gigante minorista) compró Johnson en 64 millones de dólares. Johnson recibió asesoramiento en materia tributaria por parte de Price Waterhouse Chile, compañía donde Pereira fue socio en el Departamento de Asesoría Legal y Tributaria hasta marzo de 2010. Tras la cobertura mediática, Poderopedia mostró que [Pereira](#) tenía una red de empresas, conocidas como “cascada” y tenía millonarias acciones en “Ripley” (empresa de retail controlada por los dueños de Johnson), y en el Banco de Chile (el banco que prestó a Cencosud el dinero para comprar Johnson). Pereira y su familia además arrendaron una propiedad a la gigante Cencosud. Este descubrimiento derivó en un escándalo mediático y Pereira estuvo a punto de renunciar a su cargo, pero el Presidente Piñera lo respaldó.

Poderopedia contó la historia de la ministra [Carolina Schmidt](#) y alertó que la personera del Gobierno de Sebastián Piñera tenía acciones de 7 empresas, entre ellas La Polar (compañía de retail acusada de lavado de dinero y fraude) y además reveló que había sido gerente general de una empresa de alimentos y bebidas del Grupo Luksic.

Otras historias de interés

- Muerte de Guillermo Luksic. El empresario que formaba parte del grupo económico más rico de Chile, falleció en marzo de 2013 de cáncer. Poderopedia fue el primer sitio que proporcionó a las salas de redacción información relevante de su vida y sus conexiones. Poderopedia también entregó toda la [información de Guillermo Luksic](#) bajo licencia Creative Commons 3.0, por lo que cualquier sala de prensa podía publicar el material investigado. Muchos lo hicieron, otros citaron la información y utilizaron Poderopedia como una fuente confiable de recopilación de noticias y

comprobación de datos.

- La crisis de Alsacia y los conflictos de interés de su presidente, [Juan Antonio Guzmán](#).
- [Universidad San Sebastián](#). Poderopedia mostró las conexiones y la red de sociedades espejo que utiliza la casa de estudios superiores para sacar dinero de la institución sin fines de lucro.
- Revista Dinero de México utiliza Poderopedia para publicar los [perfiles de los candidatos presidenciales](#).
- Reutilización del contenido de Poderopedia. En mayo de 2013, Poderopedia lanzó la opción de "Republicar" que permite a los medios de comunicación, blogs y cualquier sitio web para reutilizar su contenido. Desde entonces, medios como [El Dínamo](#), [Radio Bio Bio](#), [El Mostrador](#), [El Mostrador Mercados](#), [Publimetro](#) y otros están usando nuestra información.

Mónica Ventura (@monikventura)

Periodista. Con vasta experiencia en el campo del periodismo escrito y multimedia, específicamente en edición de medios online, gestión, usabilidad y arquitectura de información, actualmente es Jefa de Investigaciones de Poderopedia.org, plataforma colaborativa de periodismo de datos que mapea quién es quién en los negocios y la política en Chile. Anteriormente se desempeñó como editora de investigaciones para el portal web de la Dirección de Bibliotecas Archivos y Museos (Dibam). Durante 17 años trabajó en la Dirección de Medios Digitales del Consorcio Periodístico de Chile S.A. (Grupo Copesa) como redactora de contenidos web.

Datos que potencian la investigación: el caso de Venezuela

Emilia Díaz-Struck

El periodismo de datos ofrece una vía para salir de la creciente ola de rumores que reina en Venezuela. Al contribuir con la posibilidad de ofrecer historias sustentadas en datos documentados y verificados, se realiza un aporte para tener menos trabajos periodísticos centrados en rumores y opiniones, producto de la polarización reinante en el contexto político actual del país.

Esta disciplina no es algo totalmente nuevo en Venezuela. Aunque bajo ese nombre ha cobrado mayor popularidad y se han realizado trabajos periodísticos que involucran un gran volumen de datos, ya en los años 90, se desarrollaban trabajos periodísticos en el país que involucraban el uso y creación de bases de datos, cuando entonces se hablaba de periodismo de precisión.

Limitaciones y posibilidades en el acceso

En términos de desarrollo del periodismo de datos en Venezuela y la exploración de sus posibilidades, aún queda un largo camino por recorrer. Hay casos concretos de visualizaciones realizadas por departamentos de diagramación e infografías en los medios de comunicación, así como de investigaciones periodísticas en las que el trabajo con datos ha tenido un gran valor. Falta desarrollar más la integración de los equipos para multiplicar las posibilidades y que, por ejemplo, las visualizaciones no sean sólo el resultado de historias, sino que sirvan para ayudar con el análisis de datos durante la reportería.

La evolución de esta práctica podría estar acompañada por espacios digitales que han surgido a partir de la crisis de los medios y el cambio de propiedad de los mismos. Lo que, por sí mismo, representa retos dentro de la realidad actual.

Cuando se hablaba de periodismo de precisión

En los años 90 hubo una buena oleada de trabajos periodísticos que involucraron el uso de bases de datos. El periodista venezolano, [Carlos Subero](#), se convirtió entonces en uno de los pioneros del periodismo de precisión en la región.

Sus trabajos desarrollados entonces en la sección Expediente del diario El Universal abarcaron la cobertura de temas electorales con data proveniente del Consejo Nacional Electoral. Contó, por ejemplo, historias vinculadas con el financiamiento de campañas electorales. También junto a otros colegas del diario fueron pioneros en la región en la realización de censos parlamentarios. Buena parte de estos trabajos fueron realizados con el uso de programas como Excel y SPSS.

Algunas de las historias entonces publicadas revelaron que la campaña electoral de 20 gobernadores costó Bs. 850 millones en 1995 y cuáles fueron los 10 mecenas más importantes de un gobernador, entre otros.

En la actualidad, buena parte la data que le permitió desarrollar estas historias ya no se encuentra disponible para los periodistas. Subero ha realizado en los últimos años varias solicitudes de acceso a la información para conocer el origen de los fondos de las campañas electorales en los procesos electorales que se han realizado, sin obtener una respuesta favorable que le permita dar continuidad a este tipo de historias.

Nuevas experiencias

Las experiencias venezolanas en periodismo de datos se dividen en trabajos que han abarcado visualizaciones y reportajes de investigación periodística. Aunque aún quedan posibilidades por explorar y mucho por desarrollar, hay trabajos que muestran cómo ha sido la aproximación venezolana a este campo.

Uso de redes sociales

[Tweetómetro](#): para las elecciones presidenciales de octubre de 2012 en las que se enfrentaron Hugo Chávez y Henrique Capriles Radonski, el diario El Nacional bajo el liderazgo de Javier Pereira, creó un espacio que denominó el Tweetómetro, donde se hizo seguimiento a las tendencias e impacto de los candidatos a través de Twitter. A través de esta plataforma se reportaron tendencias virtuales asociadas con la campaña electoral presidencial en Venezuela. El trabajo fue [finalista en la edición del año 2013 de los Data Journalism Awards](#).

La plataforma fue acompañada por una serie de reportajes e historias que el diario realizó durante la cobertura de la campaña electoral. Para ello, en la Unidad de Investigación, los periodistas David González y Cristina González generaron una tabla de datos que sirvió de insumo para los reporteros del periódico y que permitió hacer seguimiento a los lugares en los que los candidatos fueron haciendo campaña con cifras específicas. Mostraron con data independiente que el candidato Capriles buscaba los votos en comunidades en las que Chávez tenía el control, viajó a distintas ciudades y prefería el contacto con la gente; mientras que el entonces presidente Chávez transmitió más mensajes desde Caracas e iba en vehículo a las actividades en las que tuvo contacto con el público.

Visualización de datos

[El audio que sacudió la pantalla](#): en 2013 el diputado opositor, Ismael García, difundió un audio de [Mario Silva](#), quien fue conductor del programa La Hojilla transmitido en el canal del Estado Venezolana de Televisión. El audio revelaba conexiones entre funcionarios del Gobierno analizadas por el personaje defensor del mismo. Últimas Noticias sacó una infografía interactiva realizada por Nathalie Naranjo y Claudia Hernández, en la que era posible hacer seguimiento y entender las informaciones dispersas dentro del audio.

En el mismo diario, el equipo de infografía junto a una reportera de la unidad de investigación (Claudia Hernández, Airam Fernández y Cristhian Rodríguez) realizó un trabajo en 2013 que muestra la cantidad de veces que en 14 años el Gobierno había hablado de magnicidio presidencial. El trabajo estuvo acompañado con una [infografía interactiva](#) y una tabla de datos⁷, que muestra los detalles de los episodios.

Reportajes

[Leche importada por el Gobierno se va de contrabando a Colombia](#): el trabajo de la periodista Lisseth Boon, publicado en 2012, logró demostrar la triangulación de la leche importada desde Nueva Zelanda a Venezuela. Mediante el uso de bases de datos en Venezuela y Nueva Zelanda logró comprobar la diferencia reportada por el Instituto Nacional de Estadística, las bases de datos de movimientos portuarios de Puerto Cabello, Importgenius y Statistics New Zealand. Eso sumado a la reportería realizada en Venezuela y Colombia, logró poner en evidencia la situación de contrabando.

El trabajo obtuvo el segundo lugar del IV Concurso Nacional de Reportajes de Investigación Periodística, organizado por el [Instituto Prensa y Sociedad de Venezuela \(IPYS Venezuela\)](#).

[En el mundo de los nombres y apellidos](#): el trabajo del periodista Carlos Subero publicado en Notiminuto, hizo uso del Registro Electoral Venezolano de 2013, para mostrar cuáles son los nombres y apellidos más comunes en Venezuela.

“Entre los 19 millones de electores incluidos en el RE, unos 530 mil llevan el González como primer apellido. Hay tantos González en toda Venezuela como para igualar a la cantidad de pobladores del Estado Nueva Esparta. En el Registro Electoral Permanente (REP) del año 2000 se contaron 312 mil González. La población de los González en el registro ascendió en 70% durante ese lapso de 13 años mientras que la cantidad total de electores del RE fue de solo 61%. Esto implica que cada vez es más probable que el día de la votación cualquiera se tope con una persona que se apellide así”, cuenta el

reportaje, que estuvo acompañado con una serie de tablas y gráficos que muestran los hallazgos obtenidos.

Alianzas

[Acuerdo crudo por derivados entre Ecuador y Venezuela benefició a los intermediarios](#): una alianza entre Venezuela y Ecuador permitió demostrar que el acuerdo de crudo por derivados entre ambas naciones no logró su objetivo: evitar a las empresas intermediarias de encarecer costos. La investigación, publicada en 2012, demostró que PDVSA contrató a ocho empresas intermediarias para poder entregar los derivados a Petroecuador y asumió en ocasiones el costo del flete. Llegó a traer barcos desde Arabia Saudita para poder cumplir con el acuerdo.

Entre el 2008 y 2011, 26% del volumen de los derivados entregados por PDVSA a Petroecuador se compraron a los intermediarios, un negocio que ascendió a \$ 953 millones, de un total de \$ 4.293 millones. Entre el 2009 y 2011, Glencore y Trafigura se llevaron el 79% de ese pastel.

El trabajo que contó con la participación de [Armando.info](#) y [El Universal de Venezuela](#), [El Universo de Ecuador](#) y [Reuters](#), se hizo a partir de la creación de una tabla de datos en la que se vació la información de las facturas de PDVSA, así como los registros portuarios y certificados de origen de los productos que llegaron a Ecuador.

[Daka se comió el manjar de Cadivi](#): después del anuncio gubernamental que denunciaba el sobreprecio en la venta de electrodomésticos importados con dólares obtenidos a tasa oficial a Bs. 6,30 otorgados por la Comisión de Administración de Divisas (Cadivi) y la fiscalización y orden de reducción de precios; el periodista César Batiz revisó la asignación de divisas Cadivi, correspondiente al año 2012. La data se encontraba en un PDF de 153 páginas, que indicaba el total de divisas aprobadas por el organismo a empresas entre 2004 y 2012. Para trabajar con la misma en formato Excel, el periodista estableció una alianza con la profesora María Esther Vidal de la Universidad Simón Bolívar.

Una vez con los datos en un formato manejable, el periodista pudo revisarlos y comprobar que la empresa Tiendas Daka, una de las intervenidas por el Gobierno por sobreprecio, había sido la tienda de electrodomésticos que mayor número de dólares a tasa oficial había recibido. También mostró cómo había sido la asignación de dólares a otras empresas. El trabajo estuvo acompañado por una infografía que mostró los hallazgos.

[Offshore Venezuela](#): como parte del proyecto Offshore Leaks del Consorcio Internacional de Periodistas de Investigación (ICIJ por sus siglas en inglés), Armando.info hizo seguimiento a las historias de venezolanos con empresas en paraísos fiscales, publicadas en 2013. La investigación tuvo como punto de partida una filtración realizada al reportero australiano, Gerard Ryle, actualmente director de ICIJ; que contenía 2,5 millones de documentos de 10 paraísos fiscales en diferentes formatos. Abarcaban archivos de texto, PDF, bases de datos, imágenes y archivos web.

Para revisar los documentos, ICIJ trabajó junto a 86 periodistas en 46 países en una primera etapa, e involucró al final a más de 115 periodistas en casi 60 países. En el caso de Venezuela, se identificaron al menos 192 personas y 114 empresas que figuraban como depositantes o representantes de empresas en las Islas Vírgenes Británicas. A partir de estos datos, se procedió a verificar documentos y revisar los nombres para constatar que se trataba de venezolanos. Se utilizaron diferentes recursos digitales como: un software que permitía hacer búsquedas en los 2.5 millones de documentos en apenas unos segundos, bases de datos, así como registros públicos en general.

Los datos filtrados por sí solos no contaban historias, eran sólo una lista de nombres y empresas. Fue necesario encontrar las historias e investigar los casos seleccionados. La reportería abarcó tanto técnicas clásicas como novedosas, relacionadas con el periodismo de datos, a través de la consulta de documentos, registros públicos y archivos, así como la colaboración de periodistas en al menos cinco

países.

Formación

Las experiencias en formación en periodismo de datos han estado de la mano de organizaciones independientes, que promueven buenas prácticas periodísticas como el Instituto Prensa y Sociedad de Venezuela (IPYS Venezuela). A través de talleres con periodistas nacionales e internacionales como facilitadores, la organización ha capacitado a periodistas en este campo.

Por un lado, la formación ha abarcado aspectos como el manejo de herramientas y programas, que ayudan a potenciar las distintas etapas del proceso de trabajo con datos: Excel, Tabula, Google Drive, Google Fusion Tables, Tableau, entre otros. Sin embargo, más allá de mostrar programas y su uso, la capacitación ha tenido un componente principal sobre la lógica detrás del periodismo de datos: cómo obtener los datos, cómo pueden ser útiles para encontrar historias, cómo combinarlos con el trabajo de reporte, retos que se plantean, así como las posibilidades del trabajo en alianza con otras disciplinas, principalmente con los programadores.

Un proyecto interesante que emergió de los talleres fue Cadivi Abierta. Muchos periodistas estaban interesados en tener acceso y manejar mejor los datos de la asignación de divisas por parte de la Comisión de Administración de Divisas (Cadivi). Como el conjunto de datos se encontraban en un PDF se hizo la conversión a Excel con Tabula, luego se hizo la limpieza de datos y se colocó a disposición en una plataforma en línea, bajo el nombre [Cadivi Abierta](#), que facilita el manejo de los datos y que combina esta información con la del Registro Nacional de Contratistas y del Tribunal Supremo de Justicia. Como resultado, es posible saber si una empresa recibió dólares a tasa oficial en 2012, si la misma ha tenido contratos con el Estado y si hay sentencias en tribunales venezolanos en alguna causa en la que esté involucrada.

Además, quedó un [registro en línea](#) sobre cómo fue el proceso de elaboración de esta plataforma.

Pensar en historias

Más allá de las limitaciones en el acceso a la información pública, es interesante explorar su potencialidad. En la medida en que el periodismo de datos en Venezuela se desarrolle pensando en las posibilidades de acceder a datos a nivel local y a nivel internacional y combinar esta búsqueda con la creación de bases de datos propias, este campo puede ser un aliado para producir historias de interés público y una herramienta poderosa para el periodismo de investigación. Trascender las fronteras puede contribuir a multiplicar las posibilidades de encontrar hallazgos de interés.

Es importante tener en cuenta el proceso de verificación y el reporte asociado con el trabajo del periodismo de datos. En la medida en que se piense qué historias pueden contar los datos, cuál es su contexto y su valor, se puede superar el enamoramiento causado por los datos y explorar más a fondo su potencialidad. Será posible ver el todo y las particularidades: el bosque, su universo, sus dimensiones; pero también las historias particulares, los árboles que llaman la atención dentro del bosque, del todo.

Pensar más allá de la visualización y considerar cómo el periodismo de datos permite potenciar distintas fases del trabajo periodístico resulta de gran valor: como generador de ideas, herramienta para la recolección de datos, facilitador para el análisis y comprensión de las historias y recurso para contarlas.

Ante el entusiasmo que existe frente al periodismo de datos, hay que tener claro que éste complementa otras formas de periodismo y técnicas desarrolladas como parte del reporte tradicional. No se trata de abandonar una cosa por otra, sino de sumar recursos y herramientas que fortalezcan el trabajo periodístico.

En el proceso de recolección de información para esta nota fueron consultados los periodistas Carlos Subero, César Batiz, David González, Cristina González y Joseph Poliszuk sobre sus experiencias en el tema del periodismo de datos en Venezuela.

El proyecto Plaza Pública: fe en los datos

Enrique Naveda

Plaza Pública nació en 2011 con una idea muy clara: evitaríamos tanto como fuera posible el periodismo declarativo y las opiniones de los expertos que interpretaban cada hecho que acontecía en el país. Esta sola decisión afectaba, naturalmente, el enfoque y las técnicas que el medio adoptaría y lo colocaba en tres sendas que a menudo se han visto como más excluyentes entre sí de lo que en realidad son: el periodismo de investigación o de profundidad, el periodismo narrativo y el periodismo de datos o de precisión.

Simplificando mucho, el primero tiene que ver sobre todo con los objetivos y el alcance, el segundo con la técnica del relato, y el tercero con la técnica de producción de la información.

Aunque después de tres años contamos con una sección de [Mapas y Datos](#), un blog sobre seguridad y violencia, [Panóptico](#), que tiene un marcado componente científico, y hemos logrado incorporar habitualmente visualizaciones a muchos de los reportajes que publicamos, profundizar en el área de datos es lo que más empeño y fe nos ha exigido.

Fe, por una sencilla razón: en general, los trabajos que responden a esta técnica obtienen en nuestro sitio web un número de visitas bastante menor que un reportaje, una crónica o una entrevista de otra índole. Insistimos en ello con cierto interés didáctico: en un país en el que las políticas públicas se suelen discutir sin evidencia científica y que muchos lectores, con aversión hacia los números o los métodos relativamente sofisticados de análisis, se distancian de ella, buscamos generar una familiaridad hacia todo ello que logre que de aquí a algunos años los ciudadanos comiencen a exigirle a medios de comunicación y a políticos información que permita sopesar los beneficios y los costos de sus decisiones. Estamos convencidos de que merece la pena. De ahí la fe.

Mantener el empeño, dada esta recepción, ha sido un compromiso constante y un logro todavía muy incipiente, porque ha implicado y sigue implicando una de las mayores inversiones de recursos: aprender técnicas de análisis de las ciencias sociales, de minería de datos y de data scraping; transformar los métodos de reporte y cambiar los lugares de búsqueda de información; asumir nuevos plazos; encontrar un punto aceptable entre la sofisticación y la sencillez que nos permitiera demostrar o decir cosas que no habían sido demostradas o siquiera conocidas sin espantar a los lectores; definir el tono; desarrollar la tecnología necesaria para gestionar y procesar las bases de datos; conocer el software libre.

También, adquirir el equipo con la potencia necesaria para procesar, ordenar y convertir en visualizaciones grandes cantidades de datos. Y combatir la [reticencia de muchos funcionarios a darnos información factual](#) y consistente, así como sobrellevar que el Estado carezca de ella.

Pero no quiero negar que esta forma de aproximarnos al periodismo haya dado resultados. Al contrario, algunos de los hallazgos más relevantes de estos tres años han sido posibles gracias a historias que hemos trabajado a partir de grandes cantidades de datos y que procesamos mediante técnicas muy rigurosas que en algunos casos hasta hace poco desconocíamos.

Nos ha permitido, por ejemplo, mostrar cómo [el Gobierno infla en un 400% los datos de creación de empleo](#), [refutar discursos racistas con mucho arraigo en la sociedad](#), revelar que el [programa de fertilizantes para campesinos en realidad les causa pérdidas](#), describir el absurdo de los [cobros por la luz eléctrica](#), o demostrar el peso gigante que tiene en la forma en que votan los diputados su [cercanía o lejanía con la cámara de empresarios](#), [su pertenencia a un partido o a un distrito](#), o el [hecho de ser mujer](#).

Al margen de estas investigaciones, hemos hecho pequeñas incursiones en el mundo de las

aplicaciones de noticias, que nosotros llamamos herramientas interactivas. La primera, [El Congreso en datos](#), es una herramienta que informa acerca del comportamiento de los diputados en el hemiciclo: muestra variables como su número de ausencias, cuándo vota a favor y cuándo en contra de un proyecto, a qué otros diputados se parece más en su manera de votar y de cuáles se diferencia más, qué bancada es la que más falta a las sesiones, etcétera. En breve tendremos que afinar esta herramienta, dado que cuando el Congreso fue consciente de algunas de las cosas que podíamos aprender con ella, desconfiguró primero el sistema de provisión de datos y después dejó de emplearlo.

La segunda es una [Calculadora de impuestos](#). La publicamos tras una reciente reforma fiscal, y le permite al lector calcular lo que debe pagar de impuesto sobre la renta y también en el impuesto de circulación de vehículos.

Técnicas usadas para la obtención de datos:

- Solicitudes vía Ley de Acceso a la Información Pública
- Scraping

Limpeza, ordenamiento, procesamiento y análisis avanzado:

- Wondershare PDF editor para pasar de PDF a texto
- [Pajek, para análisis de redes](#)
- [Philcarto para análisis geográfico](#)
- [Excel, SPSS y R para estadística](#)
- [NVIVO para análisis del discurso](#)

Visualización interactiva:

- [Apps propias](#)
- [Tableau Public](#)
- Carto DB
- [Google tools](#)
- [Infogram](#)
- [Dipity](#)
- [JavaScript](#)

Casos, tutoriales y herramientas

Lilia Saúl

Periodismo de datos en el mundo

Cómo se gastan tus impuestos

El diario The Guardian ha sido uno de los medios reconocidos por impulsar el periodismo de datos. Gracias al equipo que encabezó Simon Rogers, se lanzó un Datablog en el que se han publicado decenas de historias utilizando bases de datos, visualizaciones e infografías interactivas. Un ejemplo de ello es el análisis sobre el gasto público o [en qué se gastan los impuestos en Reino Unido](#).

La historia de cada víctima

[The Homicide Report](#) narra a través de visualización de datos, mapas e historias de vida el hallazgo de cada una de las personas que en Los Ángeles son encontradas sin vida. Las personas murieron asesinadas, en riñas o baleadas, según los reportes entregados al diario. Con los datos georreferenciados se puede saber hora, día y zona de esta localidad que cuenta con el mayor número de homicidios hasta ahora.

Los paraísos fiscales

The International Consortium of Investigative Journalists en asociación con La Nación de Costa Rica elaboraron una [aplicación Web](#) que permite hacer búsquedas de nombre y país de empresas dedicadas a invertir en paraísos fiscales.

España en Llamas

En España, la organización Civio, que entonces encabezaba Mar Cabra junto con otros periodistas, obtuvieron diversas bases de datos hasta conseguir crear un [mapa de todos los incendios](#) que padecieron en aquél país y los recursos que se invirtieron para evitar que ocurrieran los siniestros⁴.

Lo que los datos nos dicen: Menos visitas al Museo en Costa Rica

La Nación de Costa Rica es de las redacciones de América Latina que cuenta con una sección dedicada al Periodismo de Datos. La Nación Data expone los reportajes de investigación y muestra, en este caso, cómo es que los datos exponen una realidad: [El Museo Nacional de Costa Rica cerró con un 63% menos de visitas extranjeras en 12 años](#)⁵.

Abusos en pacientes

En California, 5 instituciones cuentan con mil 800 pacientes que cuentan con alguna deficiencia mental. Sin embargo, estos pacientes son los más susceptibles a ser abusados, según determinó el mismo departamento de salud. Por ello California Watch hizo [un reportaje](#) utilizando las bases de datos existentes para documentar quiénes y cuántas veces habían padecido de algún abuso o alguna herida durante su atención médica.

Otros ejemplos de periodismo de datos

Turismo en México

Son destinos turísticos en México pero [poco visitados por mexicanos](#). Así son las playas de Cancún, Los Cabos y La Riviera Maya.

En 2011 se destruyeron -50.959 empresas en España

Documentando la historia de destrucción de empresas en España, debido a la crisis económica. [Cerrado por Reforma](#) es un Mapa Colaborativo que muestra cómo en 2012 más de 53 mil empresas

cerraron en una década.

Vidas contadas: Visualizando microhistorias de represaliados en la España franquista

Pese a la existencia de muchos libros, documentales y páginas web, [el proyecto](#) –creado durante un hackathón— muestra que aún existe un trecho para avanzar en la apertura de datos para visualizar la información existente de los desaparecidos durante la época del franquismo.

Gastos del Senado 2004-2013 en Argentina

Con este proyecto La Nación de Argentina recibió el máximo galardón en la categoría “Investigación originada en Base de Datos” de los Data Journalism Awards 2013. El equipo de Datos de La Nación supo que el Senado había publicado en bruto y en gran cantidad sus gastos desde 2004. [Ordenaron, transformaron y normalizaron datos para darle coherencia a la información.](#)

#Porlosdesaparecidos: Mapa de desaparecidos en México

[#Porlosdesaparecidos](#) es un trabajo elaborado por varias organizaciones que crean un mapa colaborativo con personas desaparecidas. Incluyen varias categorías: desaparición forzada, periodistas, mujeres, niñas y niños.

[Los niños de la frontera: el drama de la inmigración infantil a EE.UU.](#)

Adopciones internacionales

Extranjeros encuentran en México la oportunidad de tener una familia. Sin embargo, de acuerdo a los datos, [no siempre se certifica el proceso.](#)

El Sistema de Procuración de Justicia en México

La organización [México Infórmate](#) investigó el Sistema de Procuración de Justicia en el país a través de los jueces, el número de policías y las necesidades de cada entidad.

[Portal de datos abiertos del Gobierno de México](#)

Herramientas para Periodistas

Guías en Línea:

- [Data Journalism Handbook](#)
- [Recursos de TCIJ](#)
- [Herramientas de Knight Center](#)
- [Material de ICFJ](#)
- [Recursos para visualizaciones](#)
- [Sitio de Sandra Crucianelli](#)

Tutoriales y herramientas:

- [Journalism in the age of data](#)
- [Presentación de Mar Cabra](#)
- [Caja de herramientas](#)
- [Data Driven Journalism](#)
- [Cursos School of Data](#)

Cómo convertir datos y no morir en el intento:

- [PDF a Excel](#)
- [CometDocs](#)
- [Zamzar](#)
- [Tabula](#)
- [AbbyFineReader](#)
- [NitroPDF](#)
- [Google Docs](#)

Sitios de Periodistas que debes conocer

- [Investigative Dashboard](#)
- [WordTree](#)
- [GeoCommons](#)
- [Detective.io](#)
- [Mapas Colectivos](#)
- [Grano Project](#)

Lilia Saúl (@liliasaul)

Periodista de la Universidad Nacional Autónoma de México (UNAM), especializada en temas de transparencia y acceso a la información, enfocada al periodismo de investigación y datos abiertos. La mayor parte de su experiencia la ha desempeñado en El Universal, donde colabora desde 2001. Actualmente es Editora de la sección Periodismo de Datos de El Universal (@Univ_Data). Becaria de la Fundación Nuevo Periodismo Iberoamericano (FNPI), de la Fundación Prensa y Democracia (PRENDE) e integrante de la iniciativa para el Periodismo de Investigación en las Américas de International Center for Journalist (ICFJ-Connectas).

Table of Contents

[Introducción](#)

[Periodismo y nuevas tecnologías](#)

[Periodismo de Datos: historia y momento actual](#)

[Manual de reducción de riesgo digital y móvil para periodistas y bloggers](#)

[Miles de registros y sólo una historia](#)

[No le tengas miedo al “mostrito”](#)

[Qué es el scraping y cómo hacerlo bien](#)

[DocumentoMedia: reportajes a partir de documentos desclasificados](#)

[Entrevista con una base de datos](#)

[Periodismo de datos en las salas de redacción](#)

[Periodismo investigativo a partir de la inteligencia de datos](#)

[Equipo de datos y buenas prácticas](#)

[Especial de La Tercera: 40 Años del 11 de Septiembre de 1973](#)

[Historia del periodismo de datos en Brasil](#)

[Transparencia en Puerto Rico: anatomía de una ilusión](#)

[En la búsqueda: periodismo de datos en Ecuador](#)

[Periodismo de datos en Los Tiempos](#)

[Periodismo de datos: una práctica incipiente en México](#)

[Periodismo económico y datos: la historia que cuentan los números](#)

[Experiencias con periodismo de datos en Iberoamérica](#)

[Casos de periodismo de datos en Uruguay](#)

[El proyecto “Una Vida es una Vida”](#)

[Juntando periodistas y programadores, las experiencias de Data’n’Press](#)

[Trabajar con datos en Perú: Comparamuni y Emergencias.pe.](#)

[Los esfuerzos individuales impulsan el periodismo de datos en Colombia](#)

[Periodismo de datos transnacional para solucionar los retos de periodismo de datos en países en desarrollo](#)

[Academia y periodismo de datos en Portugal: una relación feliz](#)

[Los proyectos “Alertas Universitarias” y “Cuánto ganan los rectores”](#)

[Universidad Nacional de Rosario y Calles Perdidas: el avance del narcotráfico en la ciudad de Rosario](#)

[Periodismo de datos en Puerto Rico](#)

[SocialTIC y la Escuela de Datos](#)

[Mapas y dibujos que cuentan historias](#)

[Infografías y visualizaciones que simplifican en exceso o confunden datos](#)

[Geoperiodismo: relatos que dialogan con el territorio](#)

[Proyecto Barcelona Commercial Footprints](#)

[Transparencia y acceso a la información en Iberoamérica](#)

[Falta de transparencia gubernamental en el Perú](#)

[Acceso a datos públicos en Brasil](#)

[Leyes de prensa, información y transparencia en Guatemala](#)

[Acceso a la información pública en Uruguay](#)

[Periodismo de datos en Ecuador: un camino por iniciar](#)

[Gobierno Abierto en Chile](#)

[Acceso a la información pública en Venezuela](#)

[Calidad de leyes de transparencia en México](#)

[Leyes que favorecen \(o no\) el desarrollo del periodismo de datos en España](#)

[Open Data y transparencia en Portugal](#)

[Acceso a la información en Colombia: entre la nueva ley y la cultura del secreto](#)

[Acceso a la información pública en Costa Rica](#)

[Exigiendo información al Estado](#)

[Verificando el discurso en un país sin ley de acceso a la información: el caso de Chequeado](#)

[Abriendo el Gobierno de Puerto Rico: el proyecto Abre PR](#)

[Paraguay, un país de “papel manteca”](#)

[Cómo pedir información pública en Argentina](#)

[Refinando datos para desvestir autoridades](#)

["Verita": para que los candidatos y candidatas digan #sololaverdad](#)

[La Nación: Gastos en el Senado en la Argentina 2004-2014](#)

[Un año y medio mapeando el poder: Poderopedia](#)

[Datos que potencian la investigación: el caso de Venezuela](#)

[El proyecto Plaza Pública: fe en los datos](#)

[Apéndice](#)

[Casos, tutoriales y herramientas](#)